



A Systematic Review of the Psychometric Properties of Composite LGBT Prejudice and Discrimination Scales

Melanie A. Morrison, CJ Bishop & Todd G. Morrison

To cite this article: Melanie A. Morrison, CJ Bishop & Todd G. Morrison (2018): A Systematic Review of the Psychometric Properties of Composite LGBT Prejudice and Discrimination Scales, Journal of Homosexuality, DOI: [10.1080/00918369.2017.1422935](https://doi.org/10.1080/00918369.2017.1422935)

To link to this article: <https://doi.org/10.1080/00918369.2017.1422935>



Accepted author version posted online: 08 Jan 2018.
Published online: 02 Feb 2018.



[Submit your article to this journal](#)



Article views: 288



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)



A Systematic Review of the Psychometric Properties of Composite LGBT Prejudice and Discrimination Scales

Melanie A. Morrison, PhD, CJ Bishop, and Todd G. Morrison, PhD

Department of Psychology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

ABSTRACT

Prejudice and discrimination against LGBT individuals is widespread and has been shown to have negative consequences for sexual and gender minority persons' physical and psychological wellbeing. A recent and problematic trend in the literature is to compositely measure prejudice toward and discrimination against LGBT persons. As such, a review of the psychometric properties of scales assessing, in a combinatory fashion, negative attitudes and/or behaviors toward LGBT persons is warranted. In the current study, 32 scales were identified, and their psychometric properties were evaluated. Most of the scales reviewed did not provide sufficient information regarding item development and refinement, scale dimensionality, scale score reliability, or validity. Properties of the reviewed scales are summarized, and recommendations for better measurement practice are articulated.

KEYWORDS

Gay; lesbian; bisexual; transgender; prejudice; discrimination; psychometric; homophobia

Previous research has suggested that prejudice (i.e., negative affective responses toward outgroup members who are perceived as “different”; Allport, 1979) and discrimination (i.e., unfair treatment by outgroup members experienced due to an individual's minority group membership; Grollman, 2012)^{1,2,3} against lesbian, gay, bisexual, and transgender (LGBT) individuals are commonplace. As such, considerable attention has been spent documenting the negative attitudes and behaviors directed toward LGBT individuals and the deleterious consequences that ensue from such negativity (e.g., Blosnich, Foynes, & Shipherd, 2013; Boehmer, Miao, Linkletter, & Clark, 2012; Bradford, Reisner, Honnold, & Xavier, 2013; Kosciw, Greytak, Giga, Villenas, & Danischewski, 2016; Marshal et al., 2011; Rosenkrantz, Black, Abreu, Aleshire, & Fallin-Bennett, 2017). Findings from much of this research typically map onto Meyer's (1995) minority stress theory (MST), which posits that sexual minority persons may internalize the experiences of negativity they encounter in their daily lives. Hendricks and Testa (2012) subsequently expanded this theory to include gender minority individuals.

CONTACT Melanie A. Morrison  Melanie.Morrison@usask.ca  Department of Psychology, University of Saskatchewan, Saskatoon SK S7N5A5, Canada.

© 2018 Taylor & Francis

Within the context of MST including sexual and gender minority persons, such internalization is believed to be the catalyst for the negative health and wellness outcomes and risky behaviors commonly observed among LGBT individuals. For example, Bradford et al. (2013) examined transnegativity-based discrimination among a large sample of gender minority individuals ($N = 350$). The authors found that over one quarter (27%) of participants reported having experienced violence and forced or unwanted sex during adolescence or adulthood, while 38% described being physically attacked since the age of 13. Thirty-seven percent of respondents indicated suffering negative experiences in high school that included antagonism from peers and/or staff, and 31% reported that their families were not supportive of their transgender identity.

Results from the Gay, Lesbian, and Straight Education Network's (GLSEN) National School Climate Survey conducted by Kosciw et al. (2016) indicated that 57.6% of participants ($N = 10,528$ self-identified LGBTQ students aged 13–21) reported feeling unsafe at their school because of their sexual orientation, while 43.3% reported feeling unsafe because of how others perceived their gender expression. Further, 81.6% of participants reported that students at their school had experienced some form of LGBT-related discrimination (e.g., were required to use the restroom that pertained to their birth sex; were admonished for public displays of affection; were prevented from promoting and, in some cases, forming gay-straight alliances; and were informed that LGBT topics were “off limits” for class assignments or discussions). The proportion of students that reported being the victims of discrimination was 66.2%. Finally, on the basis of their sexual orientation, 70.8% of students reported experiencing verbal harassment (e.g., being threatened or called names); 27% reported physical harassment (e.g., being pushed or shoved); and 13% reported physical assault (e.g., being punched or hit). The proportions that reported verbal harassment, physical harassment, and physical assault on the basis of gender expression were 54.5%, 20.3%, and 9.4%, respectively.

A meta-analysis of 19 studies with adolescent samples indicated that, in comparison to their heterosexual counterparts, LGB youth were three times more likely to have reported suicide attempts and four times more likely to have reported attempts that required medical intervention (Marshall et al., 2011). In another meta-analysis, which included data from 25 studies in different cultural contexts, the authors concluded that lesbian and bisexual women were twice as likely to have reported lifetime suicide attempts compared to heterosexual women and that gay and bisexual men were four times more likely to have attempted suicide compared to heterosexual men (King et al., 2008). While a meta-analysis on transgender persons and suicidality has not been conducted to date, results from the National Transgender Discrimination Survey (Grant et al., 2012) indicated that 41%

of respondents ($N = 6,450$ self-identified transgender and gender non-conforming individuals) had attempted suicide at some point in their lives, which the authors contrasted with a rate of 1.6% in the general population. Furthermore, the rate of suicide attempts increased for those who reported experiencing greater amounts of specific forms of prejudice and discrimination (i.e., being bullied/harassed at school [51%], experiencing job loss [55%], being physically assaulted [61%], or sexually assaulted [64%]). These findings coincide with previous research documenting that gender minority persons' experiences of transnegativity are associated with an elevated risk of attempting suicide (e.g., Clements-Nolle, Marx, & Katz, 2006).

Overall, these studies demonstrate that LGBT individuals from youth to adulthood both experience instances of prejudice and discrimination and engage in health-compromising behaviors at much higher rates than their heterosexual and cisgender counterparts. Researching the impact of homo-, bi-, and transnegativity on the health and wellbeing of sexual and gender minority persons is of paramount importance. However, the value of this type of inquiry is contingent on the psychometric soundness of the measures used to document prejudice and discrimination toward these groups.

In relation to this point, we have observed a growing trend for researchers to employ composite measures of prejudice and discrimination toward LGBT persons (e.g., Chi, Hawk, Winter, & Meeus, 2015; Hancock & Haskin, 2015; Passani & Debicki, 2016); a trend that we believe is problematic because, as noted by Worthen (2013), each of these target groups—that is, gay men, lesbian women, bisexual men, bisexual women, and female-to-male (FtM) and male-to-female (MtF) transgender individuals—likely experience unique forms of discrimination that may or may not correlate completely with the experiences of other groups. To illustrate: lesbian women are typically rated more favorably than gay men on indicants of homonegativity (e.g., Kiebel, McFadden, & Herbstrith, 2017); lesbian women are typically sexualized to a greater degree than are gay men (e.g., Kosciw et al., 2016); sentiments of binegativity have been observed in both heterosexual individuals and gay men and lesbian women (e.g., Weiss, 2004); and conflicting evidence has emerged over whether transgender women experience greater transnegativity in comparison to transgender men (e.g., Winter, Webster, & Cheung, 2008) or vice versa (e.g., Gazzola & Morrison, 2014).

The purpose of the current study is to review all instruments that measure prejudice and discrimination against LGBT individuals, with the latter treated as a single entity. Thus measures that focus on only one of these groups (e.g., gay men) were not targeted for inclusion in our review. To guide our assessment, we focused on each measure's: (1) content validity; (2) factor structure (i.e., dimensionality); (3) scale score reliability; and (4) construct validity, which may be divided into convergent and discriminant forms (Carmines & Zeller, 1979). The best practice guidelines that we used when

evaluating these four components are summarized, followed by our systematic review of LGBT prejudice/discrimination measures.

Content validity. This psychometric property concerns the relevance and representativeness of a scale's response format, items, and instructions in relation to the construct of interest (Haynes, Richard, & Kubany, 1995). In this context, *relevance* means that a measure's items address the targeted construct rather than other related constructs, and *representativeness* refers to the items tapping into the entire domain of content as opposed to more restricted elements of the construct (Vogt, King, & King, 2004). Content validity is derived from three distinct sources: (1) empirical literature; (2) relevant stakeholders; and (3) experts (Yaghmaie, 2009). First, an exhaustive review of research pertaining to the construct should occur. Second, input from stakeholders or representatives from relevant populations should be solicited as these persons can provide unique insights about the construct in question. Finally, content experts should be recruited as they can expand on and provide additional details about the construct that are not captured by the literature review or by consultation with stakeholders. Although input from experts denotes a crucial strand of support for a measure's content validity, it is not akin to, nor is it a suitable replacement for, consultation with members from the target population (i.e., input from LGBT persons). Content validity should be regarded as a critical first step in the validation process (i.e., the remaining forms of validation depend on a measure's items being content valid). To assess content validity, it is recommended that representatives of the target population and experts evaluate the items (both the initial and refined pools) using formalized rating scales (Haynes et al., 1995). Content validity indices (see Polit & Beck, 2006) can be calculated on aspects such as item clarity, straightforwardness, relevance, and representativeness.

Factor structure. If a scale contains three or more items and a composite score is computed, it is essential that the scale's factor structure be assessed using exploratory factor analysis (EFA) followed by confirmatory factor analysis (CFA). Conducting both EFA and CFA with the same sample is not recommended because doing so maximizes the likelihood that the factor solution identified may be attributable to idiosyncratic features of a single dataset.

EFA is a data-driven approach and, thus, is recommended when "a researcher has relatively little theoretical or empirical basis to make strong assumptions about how many common factors exist" (Fabrigar, Wegener, MacCallum, & Strahan, 1999, p. 277). In contrast, CFA, which is theory driven, would be advised "when there is sufficient theoretical and empirical

basis for a researcher to specify the model or small subset of models that is the most plausible” (Fabrigar et al., 1999, p. 277).

Recent systematic reviews (e.g., Morrison, Bishop, Morrison, & Parker-Taneo, 2016; Sakaluk & Short, 2017) have indicated that researchers using EFA seldom employ best practices. Specifically, in their review of 216 EFAs appearing in 139 journal articles and 24 entries in the *Handbook of Sexuality-Related Measures* (Fisher, Davis, Yarber, & Davis, 2011), Sakaluk and Short (2017) found that 59.3% of the EFAs conducted were actually principal component analysis (PCA), which is not a form of EFA; 49.5% used orthogonal rotation (typically varimax); and 51.4% used the eigenvalue greater than 1.0 “rule” to make decisions about factor retention. None of these choices are optimal. First, PCA is recommended only for data reduction since it does not take into account the underlying latent structure(s) of a group of scale items (Osborne & Costello, 2009). If the intention is to examine a scale’s dimensionality, then “true” EFA methods should be used. The three EFA methods that are recommended most often are unweighted least squares (ULS; Gaskin & Happell, 2014), maximum-likelihood (ML) estimation, and principal axis factoring (PAF), with the latter being advised when assumptions of data normality are severely violated (Sakaluk & Short, 2017). Second, orthogonal rotation assumes that generated factors are uncorrelated (Pett, Lackey, & Sullivan, 2003), which rarely applies with social scientific data (i.e., researchers typically expect “some correlation among factors, since behavior is rarely partitioned into neatly packaged units that function independently of one another”; Osborne & Costello, 2009, p. 136). Oblique rotation (e.g., oblimin and promax), which permits but does not require factors to be intercorrelated, is the preferred choice (Sakaluk & Short, 2017). Third, in terms of factor retention, no support exists for the arbitrary eigenvalue greater than 1.0 rule, with simulation studies demonstrating that it can lead to overfactoring or underfactoring (Osborne & Costello, 2009; Sakaluk & Short, 2017). Instead, parallel analysis (PA; see O’Connor, 2000 for SPSS and SAS syntax) in conjunction with other retention criteria (e.g., MAP or scree test) should be used.

Best practice recommendations for the use and reporting of confirmatory factor analysis also have been elucidated (e.g., Jackson, Gillaspay, & Purc-Stephenson, 2009).

Scale score reliability. Cronbach’s alpha, which refers to the “expected correlation between an actual test and a hypothetical alternative form of the same length” (Carmines & Zeller, 1979, p. 45), is a proximal indicator of “true” reliability (i.e., “the observed score is equal to the sum of the true score and a measurement error”; Eisinga, Grotenhuis, & Pelzer, 2013, p. 638). It is important to note, however, that reliability is based on scale scores; it is not a fixed property of the scale itself (Streiner, 2003). What this means is that

Cronbach's alpha must be computed each time a researcher uses a multi-item measure with the intention of averaging or summing the items. Although .80 is generally considered the threshold for "good" scale score reliability, there may be occasions when much lower Cronbach's alpha coefficients are acceptable (see Schmitt, 1996). It also is recommended that 95% confidence intervals be calculated for Cronbach's alpha (Fan & Thompson, 2001). The upper and lower bound estimates denote the range of plausible values for alpha should a scale be distributed to another sample drawn from the same population as the original sample (Cumming & Finch, 2005). Finally, a critical, though consistently neglected, dimension of scale score reliability is test-retest (Charter, 2003). In a systematic sample of 696 tests appearing in the American Psychological Association's *Directory of Unpublished Experimental Mental Measures*, Hogan, Benjamin, and Brezinski (2000) found that internal consistency was estimated using Cronbach's alpha for a majority (66.5%) of these tests. McCrae, Kurtz, Yamagata, and Terracciano (2011) asserted that test-retest reliability has been neglected because: (1) distributing measures to the same participants on two occasions is inconvenient; and (2) researchers appear to erroneously presume that different forms of scale score reliability, such as test-retest and Cronbach's alpha, yield equivalent information (i.e., are interchangeable). However, only test-retest quantifies a measure's reproducibility, which refers to the "degree to which a test or measure produces the same scores when applied repeatedly in the same circumstances" (Batterham & George, 2003, p. 122).

Validity. In the broadest sense, "validity" may be viewed as the degree to which a scale measures the construct it is designed to assess. There are two specific classifications of validity: criterion-related, which may be divided into concurrent and predictive forms, and construct, which may be separated into convergent and discriminant (Carmines & Zeller, 1979). Both concurrent and predictive validity focus on how well a newly developed scale correlates with another measure (ideally, a "gold standard" one) that is designed to assess the same construct. The difference, however, is that with the concurrent validity, the gold standard measure is completed along with the new scale at the same point in time, whereas with predictive validity, the gold standard is completed at some point in the future. The stronger the association between scores on the new scale and scores on the gold standard measure, the greater the level of criterion-related validity (Kimberlin & Winterstein, 2008). Oftentimes, a gold standard measure does not exist (i.e., the researcher is developing a scale that assesses a novel construct, or available measures of the construct in question are psychometrically poor). Under such circumstances, the researcher may progress directly to testing convergent and discriminant validities. Convergent validity tests whether scores on a new scale correlate—in the

predicted direction—with scores on measures of other constructs that should, for theoretical and/or empirical reasons, be associated with the new scale. Discriminant validity is tested in the same way. However, the focus is on constructs that—again, for theoretical and/or empirical reasons—should not be associated with, or should have a negligible relationship with, scores on the new scale (Springer, Abell, & Hudson, 2002). Both convergent validity and discriminant validity involve testing multiple predictions, with each supported prediction serving as one piece of evidence in support of the scale's construct validity (Carmines & Zeller, 1979). Given the complexity of a construct's nomological network (i.e., the theoretical system in which the construct is embedded), construct validation is incremental and iterative (i.e., it is an ongoing process that involves testing a series of predictions).

Method

Google Scholar in conjunction with PsycINFO, PsycARTICLES, and PsycTESTS were used to identify relevant journal articles. The References section of each identified article also was inspected to locate other materials.⁴ To be eligible for this systematic review, each article had to meet a series of inclusion criteria. First, the article had to be published in a peer-reviewed journal. Second, the article had to be written in English. Third, the authors were required to claim that they were measuring prejudice and/or discrimination toward LGBT persons as a single group. Articles that were qualitative, theoretical in nature, or focused only on a subset of the LGBT communities (e.g., assessed prejudice toward gay men and lesbian women but not bisexual persons) were excluded. After applying these criteria, 24 articles, containing 32 instruments, remained.

Each article was reviewed to determine the extent to which the authors provided evidence of the aforementioned psychometric properties—that is, content validity, scale score reliability, dimensionality, and validity (criterion-related and construct). Please see [Table 1](#) for an overview of the psychometric properties of each relevant scale appearing in each article. If evidence of the psychometric element in question was established, a check mark (✓) was entered. If no details about the psychometric element were provided, an “X” was given. A question mark (?) was issued when: (1) the psychometric property was not tested (or mentioned) directly but supportive evidence could be inferred from the results; or (2) adherence to “best practices” could not be determined due to authors providing insufficient information (e.g., the authors claimed evidence of content validity but did not indicate whether invested stakeholders and/or content experts were consulted during item development and refinement). Finally, “N/A” (not applicable) was assigned when a given psychometric element was not germane to the scale



Table 1. Psychometric properties of the reviewed scales.

Study No.	Source	Measure	Content Validity	Factor Structure	Scale Score Reliability	C-R Validity	Construct Validity
1.	Brown, Clarke, Gortmaker, and Robinson-Keilig (2004) <i>n</i> = 253 undergraduate students <i>n</i> = 105 resident assistants <i>n</i> = 126 faculty members <i>n</i> = 41 student affairs staff	4-item modified version of Student Support for GLB rights scale	X	X	✓	X	?
2.	Cabeldue et al. (2016) <i>N</i> = 403 (from Amazon Mechanical Turk)	40-item Hate Crime Beliefs Scale	?	?	✓	X	?
3.	Carlson et al. (2015) <i>N</i> = 353 long distance truck drivers	3-item scale measuring generalized prejudice against LGBT persons 4-item scale measuring individual prejudice against LGBT persons	X X	N/A N/A	N/A N/A	X X	? ?
4.	Chi et al. (2015) <i>N</i> = 172 undergraduate students	7-item Attitudes Toward Sexual and Gender Minorities	?	X	✓	X	X
5.	Cochran et al. (2007) <i>N</i> = 46 substance abuse treatment counselors	18-item measure of attitudes toward gay, lesbian, bisexual, and transgender clients	X	X	X	X	?
6.	Costa et al. (2015) <i>N</i> = 8,184 Brazilian undergraduate students	18-item Revised Prejudice Against Sexual and Gender Diversity Scale (Brazilian Context)	X	X	✓	X	?
7.	Costa et al. (2016a) <i>N</i> = 307 Brazilian health practitioners	18-item Revised Prejudice Against Sexual and Gender Diversity Scale (Brazilian Context)	?	X	✓	X	?
8.	Costa et al. (2016b) <i>N</i> = 8,184 Brazilian undergraduate students	18-item Revised Prejudice Against Sexual and Gender Diversity Scale (Brazilian Context)	?	X	✓	X	✓
9.	Dessel and Rodenberg (2017) <i>N</i> = 149 MSW students	1-item measure of attitudes about LGBT military policy 1-item measure of attitudes toward LGBT marriage equality 2-item measure of recognition of LGBT discrimination	X X X	N/A N/A N/A	N/A N/A X	X X X	? ? ?
10.	Fredriksen-Goldsen et al. (2011) <i>N</i> = 327 faculty members	4-item Social Attitudes Toward LGBT People and Issues	X	X	✓	X	?

(Continued)

Table 1. (Continued).

Study No.	Source	Measure	Content Validity	Factor Structure	Scale Score Reliability	C-R Validity	Construct Validity
11.	Gandy et al. (2013) N = 100 mental health care providers	16-item Personal Comfort Assessment Tool	X	X	X	X	?
12.	Hancock and Haskin (2015) N = 279 speech-language pathologists n = 217 American n = 53 Australian n = 7 Canadian n = 2 New Zealand	5 items measuring knowledge of LGBTQ issues 5 items measuring comfort with LGBTQ topics	X	N/A N/A	N/A N/A	X X	? ?
13.	Holland, Matthews, and Schott (2013) N = 1,768 undergraduates	11-item measure of condemnation/tolerance of LGBT people 10-item measure of LGBT social norms/morality 14-item measure of LGBT contact	X X X	X X X	✓ ✓ ✓	X X X	?
14.	Johnson and Federman (2014) N = 384 Veterans Administration psychologists	4-item measure of attitudes toward LGBT members of the military	X	X	X	X	?
15.	Lennon-Dearing and Delavega (2015) N = 235 social workers	10-item LGBT Acceptance Scale 3-item LGBT Respect Scale	X X	X X	✓ ✓	X X	? ?
16.	Logie et al. (2007) N = 173 MSW students	26-item LGBT Assessment Scale	?	X	✓	X	?
17.	Orosz et al. (2016) N = 105 Hungarian public high school students	4-item Social Distance Measure pertaining to LGBT persons 6-item modified Modern Racism Scale	X X	X X	✓ ✓	X X	? ?
18.	Passani and Debicki (2016) N = 1,371 students n = 739 (Italy) n = 335 (The Netherlands) n = 249 (Belgium) n = 48 (Estonia)	11 items measuring attitudes about LGBT rights	X	N/A	N/A	X	?
19.	Riggs and Fell (2010) N = 25 undergraduate students	20-item Modified Index of Attitudes Toward Homosexuals	X	X	✓	X	X

(Continued)



Table 1. (Continued).

Study No.	Source	Measure	Content Validity	Factor Structure	Scale Score Reliability	C-R Validity	Construct Validity
20.	Sanchez, Rabatin, Sanchez, Hubbard, & Kalet (2006) N = 248 third and fourth year medical students	13-item Attitudes toward LGBT Patients Scale	X	X	X	X	?
21.	Swank et al. (2013) n = 1,707 heterosexual undergraduate students n = 367 sexual minority undergraduate students	1 item regarding support of LGBT employment rights	X	N/A	N/A	X	?
22.	Wilson et al. (2014) N = 475 students	11-item Modified Attitudes Toward LGBT Patients Scale	X	X	✓	X	?
23.	Woodford, Brennan, Gutiérrez, and Luke (2013) N = 161 graduate teaching faculty	4-item LGBT Social Attitudes Scale	X	X	✓	X	?
24.	Woodford et al. (2012) N = 1,817 university students	9-item LGBT Social Attitudes Scale	?	?	✓	X	?

Note: ✓ = sufficient evidence of the psychometric property provided; X = no details about the psychometric element provided; ? = either supportive evidence of the psychometric property could be indirectly inferred from results or insufficient details were provided regarding adherence to best practices; N/A = the psychometric element was not applicable to the measure in question; LGBT = lesbian, gay, bisexual, and transgender; GLB = gay, lesbian, and bisexual; LGBTQ = lesbian, gay, bisexual, transgender, and queer.

in question (e.g., tests of dimensionality are unnecessary for scales containing fewer than three items).

Results

Each of the targeted psychometric properties (i.e., content validity, factor structure, scale score reliability, and construct validity) is discussed in relation to the total number of scales used across the reviewed articles. Illustrative examples are used to contextualize key points.

Content validity

None of the 32 scales reviewed had sufficient evidence attesting to their content validity. Indeed, 26 (81.2%) scales were issued an “X.” Similar to findings reported by Morrison et al. (2016), it was common for scales to use vague language in reference to the development of their instruments. The following examples illustrate this point: “based in part on” (Fredriksen-Goldsen, Woodford, Luke, & Gutiérrez, 2011, p. 22); “used an adapted six-item version of” (Orosz, Bánki, Bóthe, Tóth-Király, & Tropp, 2016, p. 512); “assessed through an adapted version of” (Riggs & Fell, 2010, p. 33); and “adapted this portion of the instrument from the survey previously mentioned” (Sanchez, Rabatin, Sanchez, Hubbard, & Kalet, 2006, p. 22). Similarly, some authors made reference to previous projects that had used certain measures but did not elucidate how the previous work influenced their new measures. For example, Carlson, Gammage, and Barrientos (2015) developed a 3-item measure that examined generalized prejudice against LGBT persons and a 4-item measure that assessed individual prejudice against LGBT persons. For both measures, the authors referred to previous surveys conducted by the International Labour Organization and briefly mentioned how these surveys inspired their new measures. The authors, however, do not indicate whether specific items were borrowed from the previous surveys or if the inspiration came from another related (yet undisclosed) source. Finally, other researchers did not mention content validity at all (e.g., Passani & Debicki, 2016; Swank, Woodford, & Lim, 2013).

The remaining six (18.8%) received a “?” since sufficient details about content validity were not provided. For example, Chi et al. (2015) described how the scales developed for their study were revised on the basis of focus groups and consultations with expert reviewers. These authors do not provide the sorts of details one would expect from a study using a focus group methodology (e.g., Was each focus group asked questions pertaining to each of the four scales? What themes emerged from the focus groups? How were the focus group data analyzed? What was the composition of the focus groups?). Similarly, Logie, Bridge, and Bridge (2007) mentioned that a survey

blueprint was used to help increase the content validity of their new 26-item LGBT Assessment Scale because they identify blueprinting as an efficient method to coordinate the construction of a scale. However, no further details on how this procedure was carried out were provided.

Factor structure

Assessments of dimensionality were relevant to 23 (71.8%) measures that consisted of three or more items. Of the nine scales for which tests of dimensionality were not relevant, five possessed three or more items but did not involve the computation of total scores. For example, Passani and Debicki (2016) reported using 11 items that measured attitudes about LGBT rights but elected to analyze each item individually. Two scales used by Carlson et al. (2015) also did not lend themselves to tests of dimensionality because scores were based on meeting or exceeding thresholds for each item.

Of the 23 measures, only seven (30.4%) reported any test of dimensionality. However, none of these seven scales had their factor structure assessed in accordance with best practices for EFA. As a result, five were issued an “X.” To illustrate: Lennon-Dearing and Delavega (2015) used the results of an EFA on their original pool of survey items, which indicated a two-factor solution, to craft their 10-item LGBT Acceptance and 3-item LGBT Respect scales. However, the authors reported using varimax rotation, which constrains factors to be uncorrelated, and no mention was made about how decisions were reached concerning factor retention.

The conflation of PCA and EFA occurred for three of these five scales. In each case, it was reported that PCA was used to investigate scale dimensionality (Costa, Machado, Bandeira, & Nardi, 2016b; Johnson & Federman, 2014; Wilson et al., 2014).

Due to insufficient details, two scales were issued a “?”: the 9-item LGBT Social Attitudes Scale developed by Woodford, Silverschanz, Swank, Scherrer, and Raiz (2012) and the 40-item Hate Crimes Beliefs Scale (Cabeldue, Cramer, Kehn, Crosby, & Anastasi, 2016).

Scale score reliability

A majority (29; 90.6%) of the scales reviewed consisted of two or more items. Of these, 24 (82.8%) computed total scale scores for one or more factors of LGBT prejudice; the remaining five (17.2%) instruments appear to have been analyzed one item at a time. Four (16.7%) of these 24 scales did not provide any evidence of scale score reliability (Cochran, Peavey, & Cauce, 2007; Gandy, McCarter, & Portwood, 2013; Johnson & Federman, 2014; Sanchez, Rabatin, Sanchez, Hubbard, & Kalet, 2006). Additionally, while a Cronbach’s alpha coefficient was provided for a 2-item scale measuring recognition of

LGBT discrimination (Dessel & Rodenberg, 2017), a Spearman-Brown coefficient would have been more appropriate (see Eisinga et al., 2013).

The remaining 19 (66.0%) instruments had Cronbach's alpha coefficients calculated for their respective samples. In addition to Cronbach's alpha, Logie et al. (2007) reported applying the Spearman-Brown prediction formula to their 26-item LGBT Assessment Scale to estimate the increase in reliability by theoretically increasing the number of items on certain subscales.

The two measures used by Carlson et al. (2015) warrant special mention since both received a rating of "N/A." The reason for this was because Carlson et al. issued a score of 0 or 1 depending on whether an individual participant's score did or did not exceed a certain threshold. To illustrate: for the item "Homosexuality (sex between men or between women) is an illness," the authors issued a numeric value of 1 if the participant's score on the 5-point Likert scale provided was 3 or higher. A numeric value of 0 was given if the score was less than 3. Taking the 4-item measure as an example, if a participant received a total score of 3 or higher using the criteria described, it would lead to a classification of "prejudiced" for that individual.

Validity

Criterion-related validity. None of the 32 scales reviewed furnished evidence pertaining to criterion-related validity. This is likely due to the absence of any measure that serves as a gold standard indicator of prejudice and discrimination toward LGBT persons.

Construct validity. None of the scales that we reviewed were tested explicitly for construct validity. A notable exception was Costa et al. (2016b) as they did mention conducting validity testing of their 18-item Scale of Prejudice Against Sexual and Gender Diversity (PASGD-R); however, they appear to have confused criterion-related and construct validity. The authors described how total scores on their instrument differed across demographic variables historically associated with greater prejudice (e.g., gender, religiosity, sexual orientation) and asserted that said differences provided evidence of the PASGD-R's criterion-related validity. The types of validation described by the authors are convergent and known-groups, both of which fall under the rubric of construct validity. Even though the type of validity was labeled incorrectly, this scale was still issued a check mark for this psychometric indicator.

Twenty-nine (90.6%) of the scales reviewed received a question mark since they provided indirect evidence of construct validity. Known-groups validity, which tests whether two or more groups expected to differ on the variable of interest do so, was the type assessed most often. For example, Passani and Debicki (2016) examined if responses to each individual item on their

measure of LGBT rights varied in accordance with participants' country of origin (Italy vs. Belgium vs. the Netherlands). As predicted, Italian participants were less likely to endorse adoption by same-sex couples than their Belgian or Dutch counterparts; the authors noted that of the three countries, Italy is the only one that has not legally granted adoption rights to same-sex couples. Meanwhile, Lennon-Dearing and Delavega (2015) reported that sexual minority participants as well as those who were Caucasian scored higher on both the 10-item LGBT Acceptance Scale and the 3-item LGBT Respect Scale (with higher scores being indicative of less prejudice and discrimination).

Finally, two (6.3%) of the scales were issued an "X" since no evidence of construct validity was provided.

Discussion

This review indicates that all of the instruments purported to measure LGBT prejudice and discrimination were, to some degree, psychometrically problematic. Expected elements of psychometric integrity such as factor structure and validity were often omitted. Further, while a majority of the studies reviewed did furnish evidence of scale score reliability, the overwhelming metric reported was Cronbach's alpha, which relies on assumptions that are rarely, if ever, met by social scientific data (see Dunn, Baguley, & Brunsten, 2014). None of the studies computed omega as recommended by psychometricists (e.g., Dunn et al., 2014).

Each of the 32 measures appearing in the 24 articles we reviewed was evaluated on the basis of five psychometric properties: (1) content validity; (2) reliability; (3) dimensionality (i.e., factor structure); (4) criterion-related validity; and (5) construct validity. As there is no gold-standard measure of prejudice and discrimination toward LGBT individuals as a monolithic group, the fourth property (i.e., criterion-related validity) could be discounted. Thus a check mark (✓) for each indicator would result in a "perfect" score of four points. Table 1 reveals that none of the reviewed scales were issued a perfect score.

Most of the measures reviewed did not adhere to best practices within the realms of item creation and development. Typically, when assessing the psychometric properties of a measure, or a series of measures, the fundamental expectation exists that the items developed, refined, and, ultimately, included in a measure will meet a certain standard of quality. Dillman (2000) provided a series of principles⁵ that can be conceptualized as best practices for item writing. First, the selection of simple over more technical terms is recommended (e.g., "work" versus "employment"). Doing so will increase the accessibility of the scale and allow for its use with a greater number of populations. Second, writing concise items will help ensure that participants

are less likely to overlook critical elements in an effort to quickly complete the scale. Third, items should be crafted as complete sentences. Although this may seem contradictory to the second principle, being too concise can lead to confusion, whereas using complete sentences will help researchers avoid this pitfall. Fourth, vague quantifiers (e.g., never, rarely, occasionally) should be avoided as a response format when more specific options are available (e.g., not at all, a few times, about once a month). Fifth, both sides of the response format should be given in a scale's instructions because doing so avoids the inadvertent encouragement of certain responses (e.g., instead of asking participants to indicate to what extent they "agree" with each item, ensure that they are asked to what extent they "agree or disagree"). Sixth, it is important to use appropriate time referents (e.g., participants are likely better able to recall their behavior over the course of the last several months as opposed to a period several years in the past). Seventh, the technical accuracy of each item should be reviewed. Eighth, double-barreled items should be avoided (i.e., the researcher should ensure that each item asks a single question). This particular element is highly relevant to composite scales using the acronym "LGBT" (i.e., the item in question may be more pertinent to one of these groups and less pertinent to others). Two sample items from the LGBT Patients Scale (Wilson et al., 2014) illustrate this point: "LGBT patients should disclose their sexual orientation to their health care providers" and "LGBT patients should only seek health care from gay and lesbian health clinics." The former is problematic because sexual orientation is not necessarily related to gender identity (i.e., some transgender individuals may self-identify as heterosexual and, thus, feel less need to "disclose" their sexual orientation). The latter item is problematic because it is unclear why a trans patient would seek care from a clinic that purportedly serves only gay men and lesbian women? Similarly, how should a bisexual respondent interpret this question?

Finally, items that are potentially objectionable should be softened (e.g., instead of asking someone if they have ever engaged in "gay bashing," one might instead ask whether they had ever had a physical confrontation with a person due to that individual's sexual orientation).

The current review suggested that none of the scales we assessed followed these guidelines. For example, Carlson et al.'s (2015) 3-item measure of general LGBT prejudice contains the following question: "A driver would lose the respect of his colleagues if he had sexual relations with a transvestite" (p. 333). The authors explicitly stated that this item was included because they deemed it important to take elements of "transphobia" into account. The term *transvestite*, however, is an archaic term that refers to cisgender men who wear female clothing in some capacity. (The more politically correct term is *crossdresser*.) The word *transvestite* is not synonymous with or comparable to the term *transgender*. Furthermore, Carlson et al.'s measures used the term *homosexual*,

which has been flagged by the American Psychological Association (APA, 2009) as problematic.⁶ Although Orosz et al. (2016) consistently used the LGBT acronym throughout their article, the only example of an LGBT-inclusive item they provided is: “Over the past few years, homosexuals have gotten more economically than they deserve.”

The 11 items used by Passani and Debicki (2016) were mentioned several times. Recall that since each item was examined in relation to other items and measures, some semblance of construct validity emerged from the data. However, the development of these items was problematic. For example, the first item states: “Gay men and lesbian women should be free to live their own life as they wish.” This statement is “double-barreled” because it combines two attitudinal targets (gay men and lesbian women). It is entirely possible that attitudes toward one target may be more or less lenient than attitudes toward the other target. To illustrate: perhaps someone who is overtly sexist may attribute fewer rights and freedoms to women regardless of their sexual orientation. Parallel items pertaining to bisexual and transgender individuals were subsequently created. However, each of the items that followed the first three regarding adoption, perceived parenting ability, and marriage made no further reference to transgender individuals.

Finally, it appears as though, in some cases, a composite measure was used to examine prejudice and discrimination toward LGBT individuals due to concerns about “excessive respondent burden, especially in a survey that was already extensive” (Woodford et al., 2012, p. 303). From our perspective, this concern does not justify collapsing the experiences of six distinct groups (i.e., gay men, lesbian women, bisexual men and women, and FtM and MtF transgender individuals) into a composite measure.

Conclusion

The results of this systematic review underscore two points. First, the use of composite measures (i.e., those that focus on “LGBT persons”) is not advisable. Second, it is essential that researchers familiarize themselves with and adhere to best-practice recommendations when developing or revising attitudinal scales. The creation of a measure is a deceptive practice. On the surface, it appears fairly simple to generate items about a given attitudinal target. However, in reality, the development of a scale is a time-intensive and quite sophisticated endeavor. It necessitates generating a large pool of items; refining and reducing those items using content and layperson experts as well as statistical methods such as inter-item correlations; and formulating a nomological network that comprises numerous tests of convergent and discriminant validity. Findings emanating from measures that treat LGBT individuals as a monolithic group; conflate prejudice and discrimination; employ outdated and potentially offensive

terminology such as *homosexual* and *transvestite*; and give limited attention to issues of reliability and validity do little to advance our understanding of the lived experiences of sexual and gender minority persons. The unintended outcome is a plethora of research from which one is unable to extract a coherent picture—an empirical testament to the dictum “garbage in, garbage out.”

Notes

1. *Homonegativity* refers to negative attitudes about and/or behaviors toward individuals that are or perceived to be gay or lesbian (Hudson & Ricketts, 1980). This term is preferable to the more commonly used word *homophobia* because it is broader in scope (i.e., the “phobia” suffix connotes an overwhelming, incapacitating, and irrational fear of gay men and lesbian women).
2. *Binegativity* refers to negative attitudes about and/or behaviors toward bisexual individuals or those perceived to be bisexual (Yost & Thomas, 2012). For reasons similar to the one articulated above, we do not endorse use of the word *biphobia*.
3. *Transnegativity* denotes negative attitudes about and/or behaviors toward individuals who do not conform to society’s current gender expectations in appearance and/or identity (King, Winter, & Webster, 2009).
4. Additional details about the Boolean search terms that were used may be obtained by contacting the senior author.
5. Dillman’s (2000) principles appear in a chapter dedicated to writing questions for surveys. However, some of these elements are relevant to the writing of scale items to ensure acceptable clarity and accessibility for respondents prior to any pilot testing. Those applicable are listed here.
6. We recommend that researchers use “gay men,” “lesbian women,” “bisexual men,” and “bisexual women” when referring to cisgender persons that self-identify as gay, lesbian, or bisexual.
7. Costa et al. (2016b) provided additional details regarding the development of the PASGD-R that were not provided in Costa, Peroni, De Camargo, Pasley, and Nardi (2015).

References

- Allport, G. W. (1979). *The nature of prejudice*. Cambridge, MA: Perseus Books.
- American Psychological Association. (2009). *Publication manual of the American psychological association*. Washington, DC: Author.
- Batterham, A. M., & George, K. P. (2003). Reliability in evidence-based clinical practice: A primer for allied health professionals. *Physical Therapy in Sport, 4*, 122–128. doi:10.1016/S1466-853X(03)00076-2
- Blosnich, J., Foyes, M. M., & Shipherd, J. C. (2013). Health disparities among sexual minority women veterans. *Journal of Women’s Health, 22*, 631–636. doi:10.1089/jwh.2012.4214
- Boehmer, U., Miao, X., Linkletter, C., & Clark, M. A. (2012). Adult health behaviors over the life course by sexual orientation. *American Journal of Public Health, 102*, 292–300. doi:10.2105/AJPH.2011.300334

- Bradford, J., Reisner, S. L., Honnold, J. A., & Xavier, J. (2013). Experiences of transgender-related discrimination and implications for health: Results from the Virginia Transgender Health Initiative Study. *American Journal of Public Health, 103*, 1820–1829. doi:10.2105/AJPH.2012.300796
- Brown, R. D., Clarke, B., Gortmaker, V., & Robinson-Keilig, R. (2004). Assessing the campus climate for gay, lesbian, bisexual, and transgender (GLBT) students using a multiple perspectives approach. *Journal of College Student Development, 45*, 8–26. doi:10.1353/csd.2004.0003
- Cabeldue, M. K., Cramer, R. J., Kehn, A., Crosby, J. W., & Anastasi, J. S. (2016). Measuring attitudes about hate: Development of the hate crime beliefs scale. *Journal of Interpersonal Violence*. Advanced online publication. doi:10.1177/0886260516636391
- Carlson, E. S., Gammage, S., & Barrientos, J. (2015). Is prejudice against LGBT persons linked to increased HIV vulnerability for heterosexual men?: Lessons for human rights advocacy from South America. *Journal of Human Rights Practice, 7*, 327–341. doi:10.1093/jhuman/huv003
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Newbury Park, CA: Sage.
- Charter, R. A. (2003). A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *Journal of General Psychology, 130*, 290–304. doi:10.1080/00221300309601160
- Chi, X., Hawk, S. T., Winter, S., & Meeus, W. (2015). The effect of comprehensive sexual education program on sexual health knowledge and sexual attitude among college students in southwest China. *Asia-Pacific Journal of Public Health, 27*, 2049–2066. doi:10.1177/1010539513475655
- Clements-Nolle, K., Marx, R., & Katz, M. (2006). Attempted suicide among transgender persons: The influence of gender-based discrimination and victimization. *Journal of Homosexuality, 51*(3), 53–69. doi:10.1300/J082v51n03_04
- Cochran, B. N., Peavey, M., & Cauce, A. M. (2007). Substance abuse treatment providers' explicit and implicit attitudes regarding sexual minorities. *Journal of Homosexuality, 53*(3), 181–207. doi:10.1300/J082v53n03_10
- Costa, A. B., Machado, W. D. L., Bandeira, D. R., & Nardi, H. C. (2016b). Validation study of the revised version of the Scale of Prejudice Against Sexual and Gender Diversity in Brazil. *Journal of Homosexuality, 63*(11), 1446–1463. doi:10.1080/00918369.2016.1222829
- Costa, A. B., Pase, P. F., De Carmago, E. S., Guaranha, C., Caetano, A. H., Kveller, D., ... Nardi, H. C. (2016a). Effectiveness of a multidimensional web-based intervention program to change Brazilian health practitioners' attitudes toward the lesbian, gay, bisexual, and transgender population. *Journal of Health Psychology, 21*, 356–368. doi:10.1177/1359105316628748
- Costa, A. B., Peroni, R. O., De Camargo, E. S., Pasley, A., & Nardi, H. C. (2015). Prejudice toward gender and sexual diversity in a Brazilian public university: Prevalence, awareness, and the effects of education. *Sexuality Research and Social Policy, 12*, 261–272. doi:10.1007/s13178-015-0191-z
- Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*, 170–180. doi:10.1037/0003-066X.60.2.170
- Dessel, A. B., & Rodenborg, N. (2017). Social workers and LGBT policies: Attitude predictors and cultural competence course outcomes. *Sexuality Research and Social Policy, 14*, 17–31. doi:10.1007/s13178-016-0231-3
- Dillman, D. A. (2000). Writing questions. In *Internet and mail surveys: The tailored design method* (pp. 32–78). New York, NY: Wiley.

- Dunn, T. J., Baguley, T., & Brunnsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*, 399–412. doi:10.1111/bjop.12046
- Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*, 637–642. doi:10.1007/s00038-012-0416-3
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299. doi:10.1037/1082-989X.4.3.272
- Fan, X., & Thompson, B. (2001). Confidence intervals for effect sizes: Confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. *Educational and Psychological Measurement*, *61*, 517–531. doi:10.1177/00131644011614001
- Fisher, T. D., Davis, C. M., Yarber, W. L., & Davis, S. L. (2011). *Handbook of sexuality-related measures* (3rd ed.). London: Routledge.
- Fredriksen-Goldsen, K. I., Woodford, M. R., Luke, K. P., & Gutiérrez, L. (2011). Support of sexual orientation and gender identity content in social work education: Results from national surveys of U.S. and Anglophone Canadian faculty. *Journal of Social Work Education*, *47*, 19–35. doi:10.5175/JSWE.2011.200900018
- Gandy, M. E., McCarter, S. A., & Portwood, S. G. (2013). Service providers' attitudes toward LGBTQ youth. *Residential Treatment for Children & Youth*, *30*, 168–186. doi:10.1080/0886571X.2013.813344
- Gaskin, C. J., & Happell, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *International Journal of Nursing Studies*, *51*, 511–521. doi:10.1016/j.ijnurstu.2013.10.005
- Gazzola, S. B., & Morrison, M. A. (2014). Cultural and personally endorsed stereotypes of transgender men and transgender women: Notable correspondence or disjunction? *International Journal of Transgenderism*, *15*, 76–99. doi:10.1080/15532739.2014.937041
- Grant, J. M., Mottet, L. A., Tanis, J., Harrison, J., Herman, J. L., & Keisling, M. (2012, September 11). *Injustice at every turn: A report of the National Transgender Discrimination Survey*. Retrieved from http://www.transequality.org/sites/default/files/docs/resources/NTDS_Report.pdf
- Grollman, E. A. (2012). Multiple forms of perceived discrimination and health among adolescents and young adults. *Journal of Health and Social Behavior*, *53*, 199–214. doi:10.1177/0022146512444289
- Hancock, A., & Haskin, G. (2015). Speech-language pathologists' knowledge and attitudes regarding lesbian, gay, bisexual, transgender, and queer (LGBTQ) populations. *American Journal of Speech-Language Pathology*, *24*, 206–221. doi:10.1044/2015_AJSLP-14-0095
- Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*, 238–247. doi:10.1037/1040-3590.7.3.238
- Hendricks, M. L., & Testa, R. J. (2012). A conceptual framework for clinical work with transgender and gender nonconforming clients: An adaptation of the minority stress model. *Professional Psychology: Research and Practice*, *43*, 460–467.
- Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, *60*, 523–531. doi:10.1177/00131640021970691
- Holland, L., Matthews, T. L., & Schott, M. R. (2013). “That’s so gay!”: Exploring college students’ attitudes toward the LGBT population. *Journal of Homosexuality*, *60*(4), 575–595. doi:10.1080/00918369.2013.760321

- Hudson, W. W., & Ricketts, W. A. (1980). A strategy for the measurement of homophobia. *Journal of Homosexuality*, 5(4), 357–372. doi:10.1300/J082v05n04_02
- Jackson, D. L., Gillaspay, J. A., Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14, 6–23. doi:10.1037/a0014694
- Johnson, L., & Federman, E. J. (2014). Training, experience, and attitudes of VA psychologists regarding LGBT issues: Relation to practice and competence. *Psychology of Sexual Orientation and Gender Diversity*, 1, 10–18. doi:10.1037/sgd0000019
- Kiebel, E. M., McFadden, S. L., & Herbstrith, J. C. (2017). Disgusted but not afraid: Feelings toward same-sex kissing reveal subtle homonegativity. *Journal of Social Psychology*, 157, 263–278. doi:10.1080/00224545.2016.1184127
- Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, 65, 2276–2284. doi:10.2146/ajhp070364
- King, M., Semlyen, J., Tai, S. S., Killaspay, H., Osborn, S., Popelyuk, D., & Nazareth, I. (2008). A systematic review of mental disorder, suicide, and deliberate self harm in lesbian, gay and bisexual people. *BMC Psychiatry*, 8, 70–86. doi:10.1186/1471-244X-8-70
- King, M. E., Winter, S., & Webster, B. (2009). Contact reduces transprejudice: A study on attitudes towards transgenderism and transgender civil rights in Hong Kong. *International Journal of Sexual Health*, 21, 17–34. doi:10.1080/19317610802434609
- Kosciw, J. G., Greytak, E. A., Giga, N. M., Villenas, C., & Danischewski, D. J. (2016, December 14). *The 2015 National School Climate Survey: The experiences of lesbian, gay, bisexual, transgender, and queer youth in our nation's schools*. Retrieved from <https://www.glsen.org/sites/default/files/2015%20National%20GLSEN%202015%20National%20School%20Climate%20Survey%20%28NSCS%29%20-%20Full%20Report.pdf>
- Lennon-Dearing, R., & Delavega, E. (2015). Policies discriminatory of the LGBT community: Do social workers endorse respect for the NASW code of ethics? *Journal of Gay & Lesbian Social Services*, 27, 412–435. doi:10.1080/10538720.2015.1087266
- Logie, C., Bridge, T. J., & Bridge, P. D. (2007). Evaluating the phobias, attitudes, and cultural competence of Master of Social Work students toward the LGBT populations. *Journal of Homosexuality*, 53(4), 201–221. doi:10.1080/00918360802103472
- Marshal, M. P., Dietz, L. J., Friedman, M. S., Stall, R., Smith, H. A., McGinley, J., . . . Brent, D. A. (2011). Suicidality and depression disparities between sexual minority and heterosexual youth: A meta-analytic review. *Journal of Adolescent Health*, 49, 115–123. doi:10.1016/j.jadohealth.2011.02.005
- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, 15, 28–50. doi:10.1177/1088868310366253
- Meyer, I. H. (1995). Minority stress and mental health in gay men. *Journal of Health and Social Behavior*, 36, 38–56.
- Morrison, T. G., Bishop, C., Morrison, M. A., & Parker-Taneo, K. (2016). A psychometric review of measures assessing discrimination against sexual minorities. *Journal of Homosexuality*, 63(8), 1086–1126. doi:10.1080/00918369.2015.1117903
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, 32, 396–402. doi:10.3758/BF03200807
- Orosz, G., Bánki, E., Bóthe, B., Tóth-Király, I., & Tropp, L. R. (2016). Don't judge a living book by its cover: Effectiveness of the living library intervention in reducing prejudice toward Roma and LGBT people. *Journal of Applied Social Psychology*, 46, 510–517. doi:10.1111/jasp.12379

- Osborne, J. W., & Costello, A. B. (2009). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pan-Pacific Management Review*, 12, 131–146. Retrieved from <http://140.133.46.32/retrieve/33379/11930.pdf>
- Passani, A., & Debicki, M. (2016). Students opinions and attitudes toward LGBT persons and rights: Results of a transnational European project. *Journal of LGBT Youth*, 13, 68–88. doi:10.1080/19361653.2015.1087927
- Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis*. Thousand Oaks, CA: Sage.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29, 489–497. doi:10.1002/nur.20147
- Riggs, D. W., & Fell, G. R. (2010). Teaching cultural competency for working with lesbian, gay, bisexual, and trans clients. *Psychology Learning and Teaching*, 9, 30–38. doi:10.2304/plat.2010.9.1.30
- Rosenkrantz, D. E., Black, W. W., Abreu, R. L., Aleshire, M. E., & Fallin-Bennett, K. (2017). Health and health care of rural sexual and gender minorities: A systematic review. *Stigma and Health*, 2, 229–243. doi:10.1037/sah0000055
- Sakaluk, J. K., & Short, S. D. (2017). A methodological review of exploratory factor analysis in sexuality research: Used practices, best practices, and data analysis resources. *Journal of Sex Research*, 54, 1–9. doi:10.1080/00224499.2015.1137538
- Sanchez, N. F., Rabatin, J., Sanchez, J. P., Hubbard, S., & Kalet, A. (2006). Medical students' ability to care for lesbian, gay, bisexual, and transgendered patients. *Family Medicine*, 38, 21–27. Retrieved from https://www.researchgate.net/publication/7393571_Medical_Students%27_Ability_to_Care_for_Lesbian_Gay_Bisexual_and_Transgendered_Patients
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 350–353. doi:10.1037/1040-3590.8.4.350
- Springer, D. W., Abell, N., & Hudson, W. W. (2002). Creating and validating rapid assessment instruments for practice and research: Part 1. *Research on Social Work Practice*, 12, 408–439. doi:10.1177/1049731502012003005
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80, 99–103. doi:10.1207/S15327752JPA8001_18
- Swank, E., Woodford, M. R., & Lim, C. (2013). Antecedents of pro-LGBT advocacy among sexual minority and heterosexual college students. *Sexuality Research and Social Policy*, 10, 317–332. doi:10.1007/s13178-013-0136-3
- Vogt, D. S., King, D. W., & King, L. A. (2004). Focus groups in psychological assessment: Enhancing content validity by consulting members of the target population. *Psychological Assessment*, 16, 231–243. doi:10.1037/1040-3590.16.3.231
- Weiss, J. T. (2004). GL vs. BT: The archaeology of biphobia and transphobia within the U. S. gay and lesbian community. *Journal of Bisexuality*, 3(4), 25–55. doi:10.1300/J159v03n03_02
- Wilson, C. K., West, L., Stepleman, L., Villarosa, M., Ange, B., Decker, M., & Waller, J. L. (2014). Attitudes toward LGBT patients among students in the health professions: Influence of demographics and discipline. *LGBT Health*, 1, 204–211. doi:10.1089/lgbt.2013.0016
- Winter, S., Webster, B., & Cheung, P. K. E. (2008). Measuring Hong Kong undergraduate students' attitudes towards transpeople. *Sex Roles*, 59, 670–683. doi:10.1007/s11199-008-9462-y
- Woodford, M. R., Brennan, D. J., Gutiérrez, L., & Luke, K. P. (2013). U.S. graduate social work faculty' attitudes toward lesbian, gay, bisexual, and transgender people. *Journal of Social Service Research*, 39, 50–62. doi:10.1080/01488376.2012.666936

- Woodford, M. R., Silverschanz, P., Swank, E., Scherrer, K. S., & Raiz, L. (2012). Predictors of heterosexual college students' attitudes toward LGBT people. *Journal of LGBT Youth*, 9, 297–320. doi:10.1080/19361653.2012.716697
- Worthen, M. G. F. (2013). An argument for separate analyses of attitudes toward lesbian, gay, bisexual men, bisexual women, MtF and FtM transgender individuals. *Sex Roles*, 68, 703–723. doi:10.1007/s11199-012-0155-1
- Yaghmaie, F. (2009). Content validity and its estimation. *Journal of Medical Education*. 3, 25–27. Retrieved from <http://www.journals.sbm.ac.ir/jme/article/download/870/769>
- Yost, M. R., & Thomas, G. D. (2012). Gender and binegativity: Men's and women's attitudes toward male and female bisexuals. *Archives of Sexual Behavior*, 41, 691–702. doi:10.1007/s10508-011-9767-8