A Systematic Review of the Psychometric Properties of Binegativity scales

Melanie A. Morrison, Mark J. Kiss, Kandice Parker, Thomas Hamp & Todd G. Morrison


To link to this article: https://doi.org/10.1080/15299716.2019.1576153

Published online: 04 Mar 2019.
A Systematic Review of the Psychometric Properties of Binegativity scales

Melanie A. Morrison, Mark J. Kiss, Kandice Parker, Thomas Hamp, and Todd G. Morrison

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

ABSTRACT
Research suggests that bisexual people encounter discrimination from heterosexual people, lesbian women, and gay men. This discrimination, henceforth labeled ‘binegativity,’ manifests itself in various forms such as denying and delegitimizing bisexuality as a sexual orientation, rendering bisexual persons invisible from the larger queer community, and promulgating various myths about bisexual persons (e.g., the myth that bisexual persons are intrinsically more promiscuous than their nonbisexual counterparts). Binegativity has the potential to compromise bisexual individuals’ mental and physical health and, consequently, their experiences of discrimination are being assessed using a wide variety of measures. To date, a review of the psychometric properties of the scales utilized to measure binegative experiences has not been published. To address this omission, a systematic review was completed to determine each scale’s adherence to best practices in psychometric development and testing. Forty-one studies using 30 unique scales were awarded a score from 0 to 5 (i.e., higher scores denote greater psychometric soundness). The Anti-Bisexual Experiences Scale emerged as the strongest measure of experienced binegativity. Other measures are identified and reviewed, and recommendations for future research are offered.

KEYWORDS
Bisexuality; binegativity; prejudice; discrimination; psychometrics; validity

A bisexual person, who may potentially be romantically and/or sexually attracted to people of more than one sex and/or gender (Ochs, 2015), challenges Western society’s entrenchment in, and comfort with, binary monosexualities (i.e., that dictate persons as either heterosexual or lesbian/gay; Diamond, 2008; Ochs & Deihl, 1992). Defiance of this normalized dichotomy can evoke discomfort and anxiety with bisexuality (Alexander & Yescavage, 2004) and affirm monosexist attitudes (i.e., those that presume that everyone is “monosexual”: attracted to no more than one gender; Eisner, 2013). This unease has been documented within heterosexual persons (e.g., Eliason, 2001), lesbian women (e.g., Rust, 1993), as well as gay
men (e.g., Mohr & Rochlen, 1999). The concept of monosexism is derived from heterosexism; a cultural ideology that exalts heterosexuality as socially and culturally superior to other sexual orientations, which are denied, delegitimized, and denigrated (Roberts, Horne, & Hoyt, 2015; Herek, 2004). Together, monosexism and heterosexism contribute to the formation of unique and pervasive negative beliefs about bisexuality (Israel & Mohr, 2004). Specifically, research demonstrates that bisexuality is oft viewed as an illegitimate sexual orientation (Eliason, 2001; Mohr & Rochlen, 1999). Further, bisexual men and women are stereotyped as being obsessed with sex, less likely to be in a monogamous relationship, and more likely to cheat (Eliason, 2001; Spaulding & Peplau, 1997). Dodge et al. (2016) conducted a nationally representative probability sample of adults in the United States and found that negative stereotypes about bisexual persons are persistent and have not improved akin to recent positive cultural attitude changes toward lesbian women and gay men. Further, monosexism toward bisexual people stemming from heterosexual and gay/lesbian cultures (Yoshino, 2000) has contributed to the cultural erasure of bisexuality (i.e., the use of restrictive “gay and lesbian” nomenclature in many legal writings; Marcus, 2018). Research comparing heterosexual men and women’s attitudes toward gay men, lesbian women, and bisexual men and women has documented that heterosexual participants evidence similar or less favorable attitudes toward bisexual targets than other sexual minorities (Herek, 2000; Israel & Mohr, 2004; Yost & Thomas, 2012). For example, Herek (2002) conducted a telephone survey in the United States assessing heterosexual men and women’s perceptions of various social groups (e.g., sexual, religious, and racial minorities) and observed that heterosexual participants endorsed greater negativity toward bisexual persons than toward gay men or lesbian women. Indeed, bisexual persons were more stigmatized than all other groups, with the exception of injection drug users. Additionally, a problematic belief about bisexual men and women concerns the oversexualization of members of this group. Stated simply, given that their sexual attraction is not limited by gender, bisexual persons may be regarded as wanton and sexually greedy (Burleson, 2005; Hayfield, Clarke, & Halliwell, 2014; Ochs, 2007). Further, bisexual persons tend to be viewed as untrustworthy in relationships and less capable of monogamy as compared to heterosexual or lesbian/gay persons (Brewster & Moradi, 2010; Burke & LaFrance, 2016; Eliason, 2001; Mohr & Rochlen, 1999).

To encapsulate the reported and observed negativity evidenced towards bisexuality, the concept of ‘binegativity’ was posited. Although the term appears to be analogous to ‘homonegativity,’ binegativity concerns prejudice and discrimination emanating not only from the heterosexual majority, but also in addition, the gay/lesbian minority (Burke & LaFrance,
The primary and central expression of binegativity is denying the very existence of a bisexual orientation (Mohr & Rochlen, 1999; Ochs, 1996). Burke and LaFrance (2016) surveyed 1379 laypersons and found the belief that bisexuality is not a stable sexual orientation was held by heterosexual and lesbian/gay participants. According to this belief, bisexuality is a temporary phase in which persons are either confused about their sexual orientation or lack sufficient courage to come out as gay or lesbian (Eliason, 2001; Hayfield, Clarke, & Halliwell, 2014; Roberts, Horne, & Hoyt, 2015). This stereotype contradicts existing evidence that demonstrates bisexuality is often a stable sexual orientation for women and men (Human Rights Commission of San Francisco, LGBT Advisory Committee, & Ulrich, 2012).

Research indicates that experiencing binegativity may have significant impacts on bisexual persons’ health. Many bisexual persons reported that they do not disclose their sexual orientation to health care providers (Human Rights Commission of San Francisco, LGBT Advisory Committee, & Ulrich, 2012; Stonewall, 2012). Moreover, bisexual persons are unlikely to access required resources and support opportunities, which may be available within (lesbian, gay, bisexual, transsexual, and queer (LGBTQ)) spaces (Flanders, Dobinson, & Logie, 2015). Pakula, Shoveller, Ratner, and Carpiano (2016), using findings from more than 220,000 Canadians who participated in the Canadian Community Health Survey between 2007 and 2012, observed that significant health disparities exist for bisexual Canadians. To illustrate: these researchers found that gay and lesbian Canadians reported more than twice the rates of anxiety and depression compared to heterosexual Canadians, and bisexual Canadians reported more than four times the rates of anxiety and depression compared to heterosexual Canadians. Bisexual Canadians further experience disproportionate rates of heavy drinking and co-occurring disorders, which coincide with reports of chronic stress, prejudice, and stigma experienced by members of the bisexual population (Pakula et al., 2016). Moreover, the public and health professionals remain generally uneducated about bisexuality, binegativity, and the unique health disparities experienced by bisexual persons (Rauch, 2016). Ross et al. (2018) conducted a meta-analysis on 52 studies that provided separate data for bisexual persons and identified that mental health disparities in bisexual persons are pervasive. Researchers observed bisexual persons consistently demonstrate significantly higher scores on indices of depression, anxiety, and other mood disorders when compared to heterosexual populations, and significantly higher or equivalent scores compared to lesbian/gay populations. Further, Salway et al. (In Press) conducted a meta-analysis of 46 studies examining the effects of binegativity and found that bisexual persons report the highest proportion of suicide
ideation or attempts, as compared to a smaller proportion reported by lesbian women and gay men, and the smallest proportion reported by heterosexual persons. Robinson and Espelage (2011) found that 33% of LGBTQ students reported thinking about suicide in the past month. This proportion was 44% among an exclusively bisexual sample of students, with 21% indicating that they had made at least one suicide attempt during the prior year.

**Purpose of study**

There is a clear need to further our understanding of bisexual persons’ self-reported experiences of binegativity due to their elevated risk of negative health outcomes. Despite this, there is a notable absence of information regarding the measurement and assessment of bisexual people’s experiences with binegativity. For example, Monro, Hines and Osborne (2017) analyzed the ways in which bisexuality has been addressed within sexuality studies from 1970 to 2015. The researchers contend that failure to address bisexual persons apart from other sexual orientations likely eclipses and conflates the needs of bisexual persons with gay men and lesbian women. This is often seen in the form of composite measurement of homonegativity where bisexual persons are considered in conjunction with gay men, lesbian women, and trans persons (Morrison, Bishop, & Morrison, In Press). This may be problematic because these types of scales rely on broad, generalized questions that fail to particularize the unique discriminatory experiences that may be associated with each group.

The current study is a systematic review assessing existing scales previously used to exclusively measure experienced binegativity. Our focus is on assessing these measures in terms of their adherence to recommended practices in scale development and refinement. This type of review is advantageous for several reasons. First, it will detail how experiences of binegativity are assessed psychometrically outside of LGBTQ composite scales. Second, in reviewing extant measures, recommendations for scale use and improvements in assessing experienced binegativity can be elucidated. In the absence of a review, researchers may continue generating ad hoc measures of varying quality or utilize scales with poor psychometric properties which may be inappropriate to use with bisexual people.

**Method**

To determine the appropriateness and utility of the scales of experienced binegativity, each measure was evaluated across five psychometric domains: (1) content validity, (2) structural validity (i.e., factor dimensionality), (3)
scale score reliability, (4) into criterion-related validity (i.e., concurrent or predictive), and (5) construct validity (i.e., convergent and discriminant validities). Below is the guide that used to make determinations whether a given measure adhered to best practice recommendations for the aforementioned domains.

**Content validity**

‘Content validity’ refers to the relevance and representativeness of the scale’s items, instructions, and response format to ensure accurate measurement of the target construct (Haynes, Richard, & Kubany, 1995). ‘Relevance’ in this context refers to the necessity for scale items to concentrate solely on the target construct while also parsing similar and nonrelevant constructs. ‘Representativeness,’ on the other hand, refers to the requirement for the scale items to address the construct’s domain in its entirety (Vogt, King, & King, 2004). Yaghmaie (2009) argues that to determine if the content validity of a scale has been properly established, three criteria must be considered: (1) relevant literature, (2) relevant stakeholders, and (3) experts. First, a systematic review of pertinent literature should be completed. Second, interviews need to occur with representatives or stakeholders from relevant populations. Finally, experts in the field should be contacted to provide any additional information that was not present in the existing literature or identified by relevant stakeholders. It is important to note that experts should not replace stakeholders; rather, both are needed to ensure the full scope of the construct (in this case, binegativity) is reflected. Content validity is often regarded as the first step needed to validate a measure because the remaining forms of validity rely on the measure’s items being relevant and representative. To assess if the scale has suitable content validity, it is recommended that stakeholders and experts assess proposed items using formalized rating scales (Haynes et al., 1995). Content validity indices (see Polit & Beck, 2006) can be determined by calculating the clarity and straightforwardness of the instructions, in addition to the representativeness and relevance of each item.

**Structural validity**

When developing or modifying a scale that contains three or more items and involves calculating a total (or average) score, it is critical to assess the scale’s factor structure using exploratory factor analysis (EFA) followed by confirmatory factor analysis (CFA). EFA and CFA also are recommended when utilizing a sample that differs, on some significant attribute (i.e., sexual orientation), from the samples used in previous literature (e.g., Kiss, Morrison, & McDonagh, In Press).
EFA is a data-driven approach that 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 comparison, CFA is theory driven and, therefore, is recommended when “when there [are] sufficient theoretical and empirical bases 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 indicate that (1) tests of dimensionality (i.e., how groups of items cluster together) are seldom conducted for scales assessing discrimination toward sexual minority persons (Morrison, Bishop, Morrison, & Parker-Taneo, 2016; Morrison et al., 2017) and (2) in sexological research, EFAs rarely follow best practice guidelines. Specifically, Sakaluk and Short (2017) reviewed 216 EFAs appearing in 139 journal articles and 24 entries in the Handbook of Sexuality-Related Measures (Fisher, Davis, Yarber, & Davis, 2011). The researchers found that 59.3% of analyses labeled EFA were, in fact, principal component analyses (PCA); 49.5% used orthogonal rotation (typically varimax); and 51.4% utilized the eigenvalue greater than 1.0 “rule” to assist with factor retention decisions. These findings are problematic for several reasons. First, PCA does not consider the underlying latent structure or structures of a group of scale items so this analysis should be used primarily for data reduction (Osborne & Costello, 2009). If the purpose is to examine a scale’s dimensionality, EFA is recommended. Second, orthogonal rotation assumes that generated factors are uncorrelated (Pett, Lackey, & Sullivan, 2003); however, this assumption rarely applies to social scientific data. Rather, social scientists generally anticipate a degree of correlation among factors because the attitudinal or behavioral items represented by the factors seldom can be partitioned into easily interpretable packets that function separately from one another (Osborne & Costello, 2009). Oblique rotation, which permits but does not require factors to be intercorrelated, is the preferred choice (Sakaluk & Short, 2017). Lastly, support for the arbitrary eigenvalue greater than 1.0 “rule” is not strong and simulation studies imply that it can lead to overfactoring or underfactoring (Osborne & Costello, 2009; Sakaluk & Short, 2017), both of which provide a distorted view of dimensionality. Rather, parallel analysis (e.g., O’Connor, 2000) in conjunction with other retention criteria (e.g., Velicer’s Minimum Average Partials [MAP] or scree test) are advised.

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 the most commonly used estimate of
When referring to reliability, it is crucial to note that it is not a fixed property of a scale; rather, it is a characteristic of scale scores (Streiner, 2003). Therefore, it is necessary to determine scale score reliability each time a researcher distributes a multi-item measure with the intention of averaging or summing the items. Although an alpha of .80 is routinely identified as the cutoff for acceptable scale score reliability, lower Cronbach’s alpha coefficients may be adequate (see Schmitt, 1996).

Another important dimension of scale score reliability is test–retest (Charter, 2003.) Unfortunately, this indicator of reliability appears to be largely neglected when researchers test the psychometric properties of scales. To illustrate: a systematic sample of 696 tests found that in 66.5% (n=533), estimates of internal consistency used Cronbach’s alpha exclusively (Hogan, Benjamin, & Brezinski, 2000). McCrae, Kurtz, Yamagata, and Terracciano (2011) argue that test–retest reliability has been disregarded because (1) it is inconvenient to distribute measures to the same participants at two different points in time due to logistics and participant attrition and (2) researchers assume that Cronbach’s alpha and test-retest are interchangeable and provide identical statistical information. According to researchers, only test–retest allows one to quantify a measure’s reproducibility (i.e., “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**

Defined broadly, ‘validity’ refers to the degree to which a scale accurately measures the construct it was designed to assess. However, the term is best conceptualized categorically: (1) criterion-related (concurrent and predictive) and (2) construct (convergent and divergent/discriminant). Criterion-related validity is determined by how the measure compares to a “gold standard” indicator of the same construct. Concurrent validation requires that the gold standard is completed contemporaneously with the new measure. In contrast, with predictive validity, the gold standard measure is distributed/assessed in the future at some point. The higher the correlation between scores on the measure being constructed and scores on the gold standard, the greater the criterion-related validity (Kimberlin & Winterstein, 2008).

There are numerous instances where a gold standard measure of a particular construct does not exist because the scale one is developing assesses a novel construct or because existing measures of the construct in question are considered to be psychometrically faulty. Therefore, it is more commonplace for researchers to determine a scale’s construct validity: convergent and divergent (or discriminant). Convergent validity assesses whether
constructs theoretically related to the scale’s construct correlate as hypothesized. Divergent validity identifies constructs that are predicted to have a negligible relationship with scores on the new scale (Springer, Abell, & Hudson, 2002). Researchers must provide strong evidence from previous research suggesting the construct’s linkage – or lack thereof. Convergent and divergent validity require researchers to test multiple related constructs, with each constituting an additional strand of evidence in support of the scale’s construct validity (Carmines & Zeller, 1979). Due to the complicated nature of a construct’s nomological network (i.e., the theoretical system in which the construct is embedded), construct validation is incremental and iterative. Thus, construct validation is an ongoing process that involves testing an array of predictions.

**Systematic review**

To locate applicable journal articles pertaining to binegativity, a systematic review of the existing literature was conducted. The following scholarly search engines were used: PsycINFO, PsycARTICLES, PsycTESTS, ERIC, SocINDEX, and Google Scholar. The keywords primarily investigated were 'binegativity OR binegative OR biphobia OR biphobic OR biprejudice' in combination with 'Measure* OR Instrument* OR Scale OR Index OR Inventor* OR Questionnaire* OR Test*.' The number of results identified was 4931. To gather as many results as possible, no specific date ranges were used.

**Inclusion criteria**

To be eligible for the current systematic review, articles were required to (1) be published in a peer-reviewed journal. (2) be published in English. (3) be categorized as quantitative. and (4) examine bisexual individuals’ experiences with discrimination due to their sexual orientation. Articles that referred to participants’ “sexual orientation” rather than their bisexuality explicitly (e.g., “Do you feel discriminated against because of your sexual orientation?”) also were included. However, as per the recommendation of Morrison et al. (In Press), composite LGBTQ measures were not utilized (e.g., “Do you feel discriminated against because you are LGBTQ?”).

**Article selection**

Of the 4931 articles identified in the database search, 699 duplicates were removed. The number of articles that were excluded because their topic or participants did not include bisexual people was 2068. The remaining articles were inspected for eligibility and excluded on the following basis: 1159 articles conflated lesbian, gay and bisexual individuals and did not
contain a relevant scale; 385 articles pertained to bisexual persons but did not contain a relevant scale; 321 articles reported using a composite LGBT-negativity measure and/or did not differentiate among these groups; 139 articles pertained to bisexual individuals but were commentaries rather than studies; 64 articles pertained to bisexual individuals but were qualitative in nature; and 55 articles pertained to measures assessing prejudicial attitudes towards bisexual persons. After excluding these articles, 41 published articles were retained which used a total of 30 measures focusing on experienced binegativity (see Figure 1).
Review
Each of the 41 retained articles using 30 unique measures of binegativity were individually evaluated for their scales’ adherence to best practices in psychometric testing. The information gathered from the systematic review is present in Table 1. If evidence of the psychometric property in question is present in the article, a check mark (√) was entered. An “X” signified that the authors did not include details about the psychometric element of interest. A question mark (?) was utilized in cases where (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 insufficient information (e.g., the authors claimed evidence of content validity but did not indicate whether bisexual persons or content experts informed item development/refinement). Finally, “N/A” (not applicable) was used to signify when the specific psychometric element was not applicable to the scale in question (e.g., reliability coefficients cannot be calculated for single-item measures and tests of dimensionality are not required for scales containing fewer than three items).

There are cases where √ were granted because the authors utilized a measure that had been rigorously evaluated in earlier studies. For instance, if content validation had been previously completed for a measure, it was not expected that future researchers repeat the content analysis process if they were using a similar configuration of participants (e.g., a sample of bisexual persons). This point, however, does not apply to measures’ scale score reliability and factor structure because they are score specific and not instrument specific. Thus, both require repeated testing with each new sample. We recognize that some researchers, peer reviewers, or even journal editors may regard testing a measure’s dimensionality as unnecessary “bulk” in a results section. In such cases, it would be appropriate to provide these details, succinctly, in an appendix, as an endnote or in the supplemental files commonly archived by journals online.

Results
Content validity
Out of the 30 scales used, four were deemed to adhere to best practice guidelines for content validity and were used by 15 (37%) out of the 41 studies we reviewed. Based on the systematic review, the most commonly used measure of experienced binegativity is the Anti-Bisexual Experiences Scale (ABES; Brewster & Moradi, 2010). In the article detailing the development of the ABES, Brewster and Moradi (2010) provide numerous strains of evidence in support of the content validity of the instrument. First, a
<table>
<thead>
<tr>
<th>#</th>
<th>Measure</th>
<th>C</th>
<th>FS</th>
<th>R</th>
<th>C-R</th>
<th>CV</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17-item Anti-Bisexual Experiences Scale [1]</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>17-item Anti-Bisexual Experiences Scale [2]</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>8-item Brief Anti-Bisexual Experiences Scale [2]</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>38-item Bisexual Microaggression and Microaffirmation Scales for Women (Microaggression subscale only) [3]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>17-item Anti-Bisexual Experiences Scale [6]</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>17-item Anti-Bisexual Experiences Scale [3]</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>17-item Anti-Bisexual Experiences Scale [12]</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>17-item Anti-Bisexual Experiences Scale [13]</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>17-item Anti-Bisexual Experiences Scale [14]</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>17-item Anti-Bisexual Experiences Scale [15]</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>17-item Anti-Bisexual Experiences Scale [16]</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>8-item measure of everyday discrimination [17]</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>14-item Heterosexist, Harassment, Rejection, and Discrimination Scale [18]</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>14-item Heterosexist, Harassment, Rejection, and Discrimination Scale [19]</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>14-item Heterosexist, Harassment, Rejection, and Discrimination Scale [7]</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>10-item Bisexual Minority Stress Scale [20]</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>45-item Homonegative Microaggressions Scale [21]</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>6-item SO victimization measure [22]</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>11-item measure of bisexual stigma [23]</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>11-item measure of external homophobic discrimination [24]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>7-item sexual orientation victimization scale [25]</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>X</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>1-item measure of discrimination and victimization due to SO [26]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>22-item microaggression questionnaire for bisexual individuals [27]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>16-item measure of lifetime victimization and discrimination [28]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>1-item measure of victimization experiences (if yes, asked to answer 4 additional items) [29]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>1-item measure of negative reactions to one’s same-sex attraction [30]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>34</td>
<td>2 items measured homophobic verbal abuse (last 12 months or ever) [31]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>2 items measured homophobic physical abuse (last 12 months or ever) [31]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>36</td>
<td>5-item Violence and Harassment subscale from modified version of Measure of Gay-Related Stress (MOGS) [32]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>?</td>
<td>0</td>
</tr>
</tbody>
</table>

(continued)
Table 1. Continued.

<table>
<thead>
<tr>
<th>#</th>
<th>Measure</th>
<th>C</th>
<th>FS</th>
<th>R</th>
<th>C-R</th>
<th>CV</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>3-item General Discrimination subscale from MOGS [32]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>38</td>
<td>3 item measure of verbal stigmatization at workplace due to SO [33]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>39</td>
<td>11-item measure of sexual orientation barriers [34]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>23 items measuring discrimination experiences among LGB [35]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>1 item measure of having ever experienced negative reactions from healthcare providers due to SO [36]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>42</td>
<td>1 item measuring being discriminated against due to SO in past year [37]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>43</td>
<td>1 item measuring experiences with anti-LGBQ crimes (if yes, additional items were completed) [38]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>44</td>
<td>4-item violent incidents in the last 12 months due to sexual orientation identity [39]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>5 items measuring perceived experiences of victimization in the last 12 months [40]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>46</td>
<td>14-item Heterosexist, Harassment, Rejection, and Discrimination Scale [41]</td>
<td>X</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>17-item Schedule of Heterosexist Events [42]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>?</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. C = content validity; FS = factor structure; R = reliability; C-R = criterion-related validity; CV = construct validity. TS = total score based on how many psychometric properties a measure was evaluated on using best practices; ✓ = 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; LGBQ = lesbian, gay, bisexual, or questioning OR lesbian, gay, bisexual, or queer; LGB = lesbian, gay, and bisexual; SO = sexual orientation.

Scales 1, 2, 3, and 4 received the highest overall rating. Numbers appearing in square brackets at the end of the listings in the Reference section correspond to the entries in the table. To illustrate, the first entry in the Reference section is for Brewster and Moradi (2010).
pool of items was generated based on a review of the existing literature pertaining to bisexual persons’ experiences with sexual orientation-related discrimination. The literature review also considered bisexual persons’ views on discrimination enacted by heterosexual individuals as well as lesbian women and/or gay men. The resulting item pool was reviewed by four expert scholars in bisexuality research. The auditing process required the experts to provide feedback about item clarity and content validity; furthermore, members of the panel were asked to suggest new items – if applicable – or to remove items they deemed to be unsuitable. Following feedback from these experts, five bisexual individuals assessed the resultant item pool on the dimensions of clarity and relevance based on their personal experiences. This process also informed the subsequent development of the Brief ABES (B-ABES; Dyar, Feinstein, & Davila, In Press). Similarly, researchers developing the Bisexual Microaggression and Microaffirmation Scales for Women (BMMS-W; Flanders, LeBreton, & Robinson, In Press) utilized community-based focus groups to generate their items. The questions were synthesized and guided based on existing research pertaining to women’s experiences of binegativity. An advisory committee of bisexual women who were activists and academics were recruited to serve as content experts.

The Bisexual Minority Stress Scale (BMSS; Molina et al., 2015) has strands of evidence suggesting content validity. However, the description detailing the process used by the researchers to assess content validity is brief, and suggests it was developed concurrently with The Daily Heterosexist Experiences Questionnaire (DHEQ; Balsam, Beadnell, & Molina, 2013). The DHEQ is a composite scale for LGBT persons’ self-report experiences with discrimination. The sample consisted of LGBT individuals, including bisexual men and women. The construction of the item pool for the DHEQ – and consequently the BMSS – was not informed by a literature review. Rather, the questions were generated through semi-structured focus groups with LGBT individuals. The DHEQ article implies that separate focus groups with experts (activists and leaders) were also conducted; however, the BMSS article reports that only stakeholders were interviewed. It is unclear if this was an oversight or if experts were not used in the development of the BMSS item pool. These sorts of discrepancies make it difficult to ascertain if the scale follows best practices for content validity.

Twenty-six of the 30 (87%) measures reviewed did not furnish evidence for the content validity of their scales and were used by 24 (59%) of the 41 studies. For instance, the second most commonly used measure of binegativity was the Heterosexist, Harassment, Rejection, and Discrimination Scale (HHRD; Szymanski, 2006). Problematically, the study describing the
construction and validation of the HHRD fails to provide adequate details that would attest to the content validity of the measure. Instead, the author simply claims that the scale is reflective of the construct, stating “The scale includes 14-items that reflect the frequency with which lesbians report having experienced heterosexist harassment, rejection, and discrimination within the past year” (Szymanski, 2006, p. 229). Despite the HHRD lacking an important psychometric property, it was utilized by authors of four of the 41 (10%) studies reviewed. Additionally, Flanders (2015) also developed and utilized a 22-item bisexuality microaggression questionnaire without furnishing evidence of its content validity.

More concerningly, a large number of studies employed ad hoc scales (i.e., those created solely for use in a specific study, e.g., Engler et al., 2011; Herek, Gillis, & Cogan, 1999; Lea, de Wit, & Reynolds, 2014), used a scale assessing nonspecific sexual orientation-related discrimination (e.g., Durso & Meyer, 2013; Lewis, Derlega, Brown, Rose, & Henson, 2009), or assessed binegativity by modifying an existing scale designed to measure a (purportedly) similar construct (e.g., discrimination experienced by gay men; see, for example, Bostwick, 2012; Sarno & Wright, 2013). Regardless of the sexual orientation measure employed, none of these scales offered compelling evidence of content validity.

**Structural validity**

To ensure that binegativity scales adhered to best practices in psychometric testing, it was required that measures containing three or more items have their factor structure tested. According to Osborne and Costello (2009), factors with fewer than three-items tend to be unstable. Thus, if a scale were found to be unidimensional, a three-item minimum was considered acceptable. If the scale contained more than one factor, a minimum of six items (i.e., three items per factor) is recommended. In the current review, three out of the 41 (7%) studies assessed the factor structure of the scales used to measure binegativity. Further, all structural validity efforts occurred within the context of scale development. ABES (Brewster & Moradi, 2010) and BMMS-W (Flanders et al., In Press) illustrated partial adherence to best practices in factor analysis. In these two studies, the authors conducted an EFA using principle-axis factoring with oblique rotation (promax) which is recommended. Problematically, the number of factors retained was determined using the arbitrary eigenvalue greater than 1.0 “rule” which is not advisable (Osborne & Costello, 2009; Sakaluk & Short, 2017). Factor retention techniques such as parallel analysis would have been preferable (O’Connor, 2000). Based on these results, the researchers then conducted a CFA using a sample that did not overlap with the one used for their EFA.
In addition, the B-ABES (Dyar et al., In Press), which serves as a short version of the aforementioned scale by Brewster and Moradi (2010), reduced the ABES from 17 items to 8 items using best practices in EFA. Further, researchers conducted a CFA to ensure the short form of the measure retained the ABES factor structure.

Out of the 41 studies in our review, 25 (61%) did not examine dimensionality for one or more of the scales used and were flagged with an “X.” Thus, some researchers appear to have been operating from the assumption that scales were unidimensional or had items that formed logically coherent groupings. For instance, 15 studies (37%) used the ABES (Brewster & Moradi, 2010), but 12 (29%) of these did not assess its factor structure (other than the two previously mentioned studies that validated the long and short versions). This is problematic because certain studies use samples consisting of numerous categories such as pansexual, queer, demisexual, two-spirit, fluid, heterosexual, gay, and lesbian in addition to bisexual persons (e.g., Flanders et al., In Press; Watson, Morgan, & Craney, 2018). Other studies used a wide range of specifically nonmonosexual orientations such as queer, questioning, demisexual, and pansexual (e.g., Katz-Wise, Mereish, & Woulfe, 2017; MacLeod, Bauer, Robinson, MacKay, & Ross, 2015; Watson et al., 2018). Further, in some cases, these participants were collapsed into a single grouping for statistical analysis. Therefore, studies with samples differing from Brewster and Moradi’s (2010) should assess the ABES’ factor structure to determine its structural validity. Dyar et al. (In Press) provided evidence that the factor structure of the B-ABES did not differ according to sexual identity, accounting for several non-monosexual configurations (e.g., bisexual, pansexual, lesbian, gay, queer, omnisexual, mostly heterosexual, heterosexual). This finding may suggest that combining certain non-monosexual categories may be appropriate. However, rather than operating from the assumption of factorial invariance, researchers should assess the factor structure of the ABES across different configurations of non-monosexual orientations.

The HHRD (Szymanski, 2006) was used by four of the 41 (10%) studies despite the author of the original article assessing the scale’s factor structure using a sample of lesbian women (e.g., Katz-Wise et al., 2017; Lambe, Cerezo, & O’Shaughnessy, 2017; Watson, Grotewiel, Farrell, Marshik, & Schneider, 2015). It is certainly possible that the factor structure of the HHRD, when completed by bisexual individuals, may differ; thus, the rationale for omitting this type of analysis is unclear.

Lastly, many of the studies we reviewed employed measures of victimization that did not specify bisexuality but rather used the general concept of sexual orientation (e.g., Lewis et al., 2009; van Lisdonk & Keuzenkamp, 2017). Out of the nine measures that employed this generic term and
contained three or more items, none of the researchers utilizing them to measure binegativity assessed their factor structure.

Eleven (27%) of the 41 studies used 13 measures that were coded “N/A.” Five of the measures possessed more than two items; however, each item was examined individually, effectively rendering them single-item indicators (e.g., Lea et al., 2014; Pereira & Costa, 2016; van Lisdonk & Keuzenkamp, 2017). The remaining eight measures contained two items or fewer. Finally, a “?” was issued for the Sexual Orientation Victimization scale (SOV; Darwich, Hymel, & Waterhouse, 2012) and the BMSS (Molina et al., 2015) because insufficient details were provided. The authors of these two measures did not specify (1) the type of factor analysis that was performed, (2) the factor extraction method that was used, (3) the type of rotation that was used, (4) the size of the eigenvalues for the retained factors, and (5) the proportion of common variance accounted for by the factor solution. Without these details, it is impossible to assess whether the authors’ use of factor analysis adhered to “best practice” guidelines.

**Scale score reliability**

Twenty-nine (70%) of the 41 studies included in the review employed measures consisting of three or more items; therefore, assessing scale score reliability is necessary. Twenty-two (54%) of the studies assessed scale score reliability by calculating Cronbach’s alpha for each scale when multiple scales were present in a single study. Two (5%) studies provided no scale score reliability indicants. In a single study (2%), Cronbach’s alpha coefficient was reported based on previous research; however, the author did not compute it for their own sample (e.g., Weber, 2008). As previously stated, indicators of reliability such as Cronbach’s alpha are data dependent and, thus, must be computed each time a summative scale is used with a new sample. Three (7%) of the 41 studies (Darwich et al., 2012; Fredriksen-Goldsen et al., 2012; Lewis et al., 2009) computed overall Cronbach alpha for items that were not inter-related. Doing so does not make conceptual sense (i.e., Cronbach’s alpha assumes that how a participant responds to one item should have logical implications for how they respond to the other items). For instance, Fredriksen-Goldsen et al. (2012) calculated an overall Cronbach’s alpha for the 16-item Lifetime Victimization Scale and Discrimination Scale. However, experiencing property damage may or may not relate to experiencing denial of healthcare. A single study (2%) was labeled “?” because it was unclear if the computed alpha was for the current study or that of a previous study.

Eleven of the 41 studies (27%) were coded as “N/A.” Seven (17%) studies received this rating because they used measures that consisted of two or
fewer questions. Five (12%) studies were given this rating because, while they used multi-item measures that contained three or more items, each item was analyzed individually. (It should be noted that one study had both a 1-item and 3+ item scale, which resulted in the sum total of 12 rather than the expected 11.) As a result, there was no need to compute a Cronbach’s alpha coefficient.

Of the 41 studies, only one (2%) furnished evidence of test-retest reliability in addition to Cronbach’s alpha. Brewster and Moradi (2010) conducted a 2-week follow-up for the online survey that was used in the development of the 17-item ABES. These researchers observed strong correlations between the two studies (LG subscale: \( r = .89, p < .001 \); H subscale: \( r = .78, p < .001 \)); thus, providing a strand of evidence of test–retest reliability for the measure.

**Validity**

*Criterion-related validity*

Out of the 41 studies using 30 measures of binegativity, one study explicitly tested for criterion-related validity. However, it is important to note that the omission of criterion-related validation testing is not problematic if, (1) the authors are developing a scale that measures a heretofore unexamined construct or (2) the authors do not believe that a measure presumed to be a “gold standard” is psychometrically sound. In the case of binegativity scales, there are very few existing instruments that specifically measure the discrimination that bisexual people may experience. However, none of the studies we reviewed justified omitting criterion-related analyses for these reasons. The ABES (Brewster & Moradi, 2010), which is the most commonly used measure of binegativity, was treated as a gold standard measure by Flanders et al. (In Press) in their assessment of the criterion-related validity of the BMMS-W. Although the status of the ABES as a psychometrically optimal measure of binegativity may be questioned, we treated it as such in the Flanders et al.’s study, due to their explicit suggestion that the ABES criterion validity was achieved (Table 1). Indeed, the researchers observed a fairly strong correlation between scores on the ABES and scores on the BMSS-W (\( r = .65, p < .001 \)).

*Construct validity*

Of the 41 studies reviewed, two (5%) provided direct evidence of construct validity. For a measure to be considered as possessing construct validity in the current review, the authors must have explicitly hypothesized a relationship between experienced binegativity and other constructs (convergent
or divergent) and provided adequate evidence from existing literature to support their prediction(s). To illustrate: Brewster and Moradi (2010) tested the convergent and discriminant validity of the ABES. Specifically, it was predicted that scores on the ABES would positively correlate with awareness of stigma (i.e., the awareness and personal salience of social stigma against one’s group) and provided previous research attesting to this association. The full ABES measure correlated positively ($r = .53, p < .001$) with the Stigma Consciousness Questionnaire (SCQ; Pinel, 1999), suggesting convergent validity. Also, as hypothesized, these researchers found that scores on the ABES did not correlate significantly with scores on a measure of social desirability bias – specifically, the 20-item Impression Management scale of the Balanced Inventory of Desirable Responding (Paulhus & Reid, 1991). The researchers suggest that this nonsignificant association provides evidence of discriminant validity; however, it would be more accurate to assert that it furnishes evidence attesting to the ABES’ (apparent) lack of susceptibility to response bias. The construct validity of the short-form version of the ABES (i.e., the B-ABES) also has been tested. For example, Dyar et al. (In Press) hypothesized that scores on the ABES and B-ABES would correlate positively with scores on depression and anxiety as measured by the 9-item Patient Health Questionnaire (PHQ; Kroenke, Spitzer, & Williams, 2001) and the 7-item Generalized Anxiety Disorder measure (GAD-7; Spitzer, Kroenke, Williams, & Löwe, 2006), respectively. Moreover, using previous research, they provided detailed evidence concerning how these constructs relate. Dyar et al. (In Press) observed modest correlations with anxiety for the short and full ABES measure, but not for depression. Despite only partially supporting their hypotheses, this finding adds additional evidence to the construct validity of the measure, and potentially suggests that depression does not correlate with scores on the ABES. However, future research using a battery of depression measures other than the PHQ would be necessary to make this determination.

Fourteen studies (34%) that received an “X” for this psychometric indicator provided no evidence of convergent or discriminant validity. Considering the dearth of scales that evaluate binegativity, it is especially problematic that novel scales are neglecting this important indicator of psychometric soundness. Flanders (2015), for instance, used a 22-item micro-aggression questionnaire for bisexual individuals but did not provide evidence of construct validity. Furthermore, in the development of the 38-item BMMS-W (Flanders et al., In Press), criterion-related validity was tested but no pertinent evidence of construct validity was provided. Construct validity was also neglected in Bostwick’s (2012) article assessing an 11-item measure of bisexual stigma.

Twenty-five (61%) studies were coded with a “?” because construct validity was not explicitly mentioned; thus, making this determination requires
effort from the reader to gauge if the tested constructs provide evidence of the measure’s construct validity. As previously mentioned, the ABES (Brewster & Moradi, 2010) is the most commonly used measure of binegativity with 16 studies, thus far, employing the scale. Of the 16 studies using the ABES, 10 of these studies provided strands of support for the construct validity of this scale. Watson et al. (2018) hypothesized that binegativity and sexist events would be correlated due to theoretical overlap. Specifically, it was posited that greater experiences of binegativity should predict more experiences of sexism for bisexual women. Furthermore, the authors drew upon relevant literature to generate this hypothesis. A moderate positive correlation was observed ($r = .52$, $p < .001$), possibly furnishing additional evidence in support of the convergent validity of the ABES. Using previous research, Roberts et al. (2015) predicted that familial and friendship support would be negatively associated with binegativity, measured using the ABES. The findings of the study supported the prediction (Family: $\beta = -0.24$, $p < .001$; Friends: $\beta = -0.18$, $p < .001$), unveiling a potential strand of evidence for convergent validity.

The HHRD (Szymanski, 2006) was used by four (10%) of the reviewed studies to assess binegativity, none of which explicitly provided evidence of construct validity for the scale. This is cause for concern when considering the lack of psychometric evidence for the scale’s applicability to bisexual samples (i.e., it was developed for lesbian women). However, Katz-Wise et al. (2017) did observe a correlation between scores on the HHRD and binegativity with the ABES ($r = .59$, $p < .05$). This suggests that greater experiences of sexual orientation-related stigma are associated with greater experienced binegativity. Unfortunately, this finding was neither formally predicted nor stated explicitly as evidence for the construct validity of the HHRD for use with bisexual persons.

Twenty-two (54%) of the 41 studies used scales without psychometric rigor (i.e., previous studies had not determined content validity, factor structure, criterion-related validity, or construct validity). This is especially evident in scales created on an ad hoc basis (e.g., Pereira & Costa, 2016; Saewyc, Poon, Homma, & Skay, 2008). In such cases, it is difficult to judge whether indirect evidence of construct validity can be immediately determined.

Lastly, one study (2%) provided indirect evidence for two of the three scales it used but no evidence was provided for the third scale.

**Discussion**

In the current systematic review, we found that the majority of the studies assessing experienced binegativity use measures that do not adhere to best practice recommendations for scale development and validation. Tests of
structural, criterion-related, convergent, and discriminant validities were often omitted despite being crucial elements of psychometric testing.

A total of 41 studies with 30 measures of binegativity were assessed based on their adherence to four psychometric domains: (1) content validity, (2) structural validity (i.e., factor structure), (3) scale score reliability, (4) criterion-related validity, and (5) construct validity. The latter was partitioned into criterion-related and construct resulting in five points of evaluation. A checkmark (√) for each of the five properties would signify a “perfect” score of five points. The current systematic review reveals that none of the reviewed scales fulfilled all five criteria (see Table 1). The scales that were awarded the most points (four out of five) were the ABES (Brewster & Moradi; 2010), the B-ABES (Dyar et al., in Press), and the BMMS-W (Flanders et al., in Press). The ABES, B-ABES, and BMSS-W represent the strongest measures evaluated for use with North American, English-speaking samples.

The results of this systematic review serve to highlight the importance of adhering to best practice psychometrics when developing, revising, or utilizing measures. Furthermore, on the basis of this review, the following broad recommendations are offered to aid in future research pertaining to experienced binegativity.

**Recommendation 1:** The purpose of the systematic review was to use scientific research databases to locate peer-reviewed studies centered exclusively on bisexual individuals’ experiences with binegativity. Thus, broad keywords were utilized to understand how binegativity was being measured (e.g., ‘binegativity,’ ‘biphobia’). The findings of the current review suggest that experienced binegativity has traditionally been measured using scales that were not intended for or validated to measure bisexual persons’ encounters with discrimination. More problematically, a large number of these scales were ad hoc measures that had not been validated with bisexual samples. Thus, going forward, it is recommended that scales be used with content informed by bisexual individuals.

**Recommendation 2:** In the absence of a compelling rationale, we suggest that researchers abandon creating ad hoc measures of binegativity. The ABES Scale (Brewster & Moradi; 2010), in addition to the reduced 8-item B-ABES (Dyar et al., in Press), appear to hold considerable promise as measures of this construct. Thus, we advocate for the continued use and incremental refinement of the ABES and B-ABES.

**Recommendation 3:** The conceptualization and definition of bisexuality requires further consideration because the parameters of this sexual orientation can differ greatly depending on the study. A portion of the articles used samples of self-identified bisexual individuals, whereas some also used bisexuals in combination with lesbian, gay male, heterosexual participants.
but considered them as a singular group. As well, some studies that used purportedly “bisexual” samples included participants who identified as pansexual (i.e., attraction regardless of sex or gender identity; Morandini, Blaszczynski, & Dar-Nimrod, 2017), demisexual (i.e., sexual attraction appears only after establishing a meaningful emotional bond; Haefner, & Plante, 2015), two-spirit (i.e., indigenous identity as a “third” or nonbinary gender; Wilson, 1996), queer (i.e., rejection of heterosexuality, gender binaries and sexual orientations labels; Levy & Johnson, 2012), fluid (i.e., changes over time to sexual orientations or attractions; Katz-Wise & Hyde, 2015) and nonmonosex (i.e., sexual identities outside of a heterosexual/homosexual binary; Ferguson & Gilmour, 2018). Collapsing these diverse participants into a singular category of “bisexual” is conceptually problematic. For example, in its traditional sense, ‘two-spirit’ refers primarily to gender nonbinary dress or identity in Indigenous communities and is not considered to reflect the Westernized view of sexual orientation (Estrada, 2011; Wilson, 1996). Also, the word ‘queer’ may differ from individual to individual because the term is somewhat ambiguous. Therefore, this category may not be strictly limited to individuals with nonmonosex attraction but rather a wide assortment of personal meanings about gender and attraction (e.g., a queer “man” with sole attraction to gender-fluid individuals who have penises). As well, lumping bisexual persons in with a plethora of other nonheterosexual configurations contributes to the erasure of bisexual persons and, in so doing, may reinforce the view that their identity lacks “legitimacy.” Researchers need to carefully justify their decision to combine these groups theoretically (i.e., would these groups deem it appropriate to be considered together?) and psychometrically (i.e., do nomological networks differ according to group?).

It is important to note that we recognize certain exigencies may prevent researchers from following best practice in measure development and refinement (e.g., journal space limitations, peer-reviewers’ feedback, and editors’ requests). However, a number of these psychometric requirements can be conveyed in a succinct manner (e.g., one or two sentences). Further, more space consuming requirements, such as factor analysis, can be furnished as supplemental materials. However, we strongly believe that journal editors should require authors to provide evidence of measures’ scale score validity and reliability as doing so increases researchers and laypersons’ confidence in the findings we produce. Adherence to clearly articulated best practice recommendations also enhances replicability – an issue that has recently proven contentious in psychology (e.g., Open Science Collaboration, 2015).

In conclusion, the findings of our systematic review suggest that the ABES was the strongest measure of experienced binegativity. Rather than
continuing to generate psychometrically dubious ad hoc measures, we advocate that researchers focus on testing the ABES and its short-form (B-ABES) and, in so doing, expand the nomological network of these instruments. This is more than an esoteric “call to arms.” Suboptimal measurement of bisexual persons’ experiences of discrimination may have profound implications for their psychological and physical well-being.

**ORCID**

Mark J. Kiss [http://orcid.org/0000-0002-4040-7839](http://orcid.org/0000-0002-4040-7839)

**References**


