Systematic review of the psychometric properties of transphobia scales

Melanie A. Morrison, CJ Bishop, Stephanie B. Gazzola, Jessica M. McCutcheon, Kandice Parker & Todd G. Morrison

Taylor & Francis
Taylor & Francis Group

Check for updates

Systematic review of the psychometric properties of transphobia scales

Melanie A. Morrison, CJ Bishop, Stephanie B. Gazzola, Jessica M. McCutcheon, Kandice Parker, and Todd G. Morrison

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

ABSTRACT

Introduction: Prejudice and discrimination against transgender individuals (i.e., transphobia) is pervasive and has been shown to have pernicious effects on the physical (e.g., substance abuse and other self-harm behaviors) and psychological (e.g., depression and suicidal ideation) well-being of those targeted.

Aims: To date, a review of the psychometric properties of scales assessing transphobia has not been conducted; this gap compromises researchers' ability to make informed measurement decisions.

Methods: In the current study, 61 articles that contained 83 scales were identified, and their adherence to best practices in psychometric testing was evaluated.

Results: Most of the transphobia scales included in the current review did not provide sufficient information about item generation and refinement, scale dimensionality, scale score reliability, or validity. Each scale was entered into a table and was ranked on the basis of its total score. A score of 1 was issued for each psychometric feature that adhered to best practice guidelines, with total scores ranging from 0 to 5 (i.e., higher scores denote greater psychometric soundness).

Discussion: Properties of the reviewed scales are summarized and recommendations are made for future transphobia scale development and validation. The measures that emerged as possessing the highest scores and, subsequently, the greatest utility are identified.

The term *transphobia* is traditionally defined as an irrational fear or hatred of, or an emotional disgust toward individuals who do not conform to society's gender expectations (Hill & Willoughby, 2005). This traditional definition captures attitudes and behaviors that denote a "phobic response." In the context of this article, "transphobia" should be conceptualized in a more comprehensive way—that is, stereotypes, prejudice, and discrimination directed toward individuals that are or are perceived to be transgender. Although limited, available research suggests that substantial proportions of individuals are transphobic. For instance, using data obtained from a nationally representative sample of Americans, Norton and Herek (2013) evaluated respondents' feelings toward various marginalized groups including gay men, lesbian women, and transgender persons. Evaluation thermometers were used. This technique requires that individuals assign a "temperature" reflective of their feelings toward the group in question (i.e., possible values ranged from 0 degrees to 100 degrees with

lower temperatures reflecting colder, less favorable attitudes). The mean temperature allocated for transgender individuals was 32 degrees, which was well below the neutral midpoint (50 degrees) and significantly lower than temperatures assigned to other stigmatized groups (e.g., gay and bisexual men). Focusing specifically on beliefs about trans women, Winter and colleagues (2009) surveyed undergraduate students (*N* = 841) attending universities located in seven different countries (China, Malaysia, Philippines, Singapore, Thailand, United Kingdom, and United States of America). The researchers found that participants (1) socially rejected trans women (i.e., agreed that should their brother, son, or teacher disclose a transgender identity, the participant would be unable to accept that transition) and (2) evidenced mild endorsement of the belief that trans women are mentally ill.

These two studies are illustrative of the published research on transphobia—a corpus of work that is notable for being small in number. For example, Worthen's (2016) recent narrative review of the

literature details only 18 quantitative studies that examine attitudes toward transgender individuals. It should be noted that (1) all of the research identified by Worthen (2016) was published between 1983 and 2013; (2) 15 of the 18 studies used convenience samples of undergraduate students; and (3) 9 of the 15 studies recruited participants from the United States. Such a dearth of research is problematic given the documented sequelae of transphobia. Adapting Meyer's minority stress model (1995, 2003) to better capture the experiences of transgender individuals, Hendricks and Testa (2012) explicate the varying sources of these negative outcomes. Aside from the general life stressors that all individuals experience (e.g., concerns about money), Hendricks and Testa suggest that transgender individuals are especially susceptible to external stressful events that manifest as attitudinal or behavioral negativity (e.g., hate speech, social ostracism, and physical and/or sexual assault). For example, in a recent survey of victimization among trans persons residing in the United Kingdom ($N = 600$), Ellis, Bailey, and McNeil (2016) found that episodes of physical assault were "fairly common" (p. 216), with approximately 20% of respondents stating that they had been "hit" or "beaten up" because of their trans status. Larger proportions reported experiencing physical intimidation (39.5%), sexual harassment (40.3%), objectification/fetishization (52%), name calling (75%), and "silent harassment" (84.9%). The latter term refers to being stared at or talked about which, in turn, contributes to a sense of marginalization and plays a role in establishing a climate of fear. The authors also noted that much of this victimization occurred recently (i.e., within the past 12 months). To illustrate, 24.3% of those subjected to physical intimidation/threats and 51.9% of those who had been the recipient of silent harassment had experienced this victimization during the 12-month period preceding their involvement in the survey. The consequences of transphobia can be devastating and include depression, substance abuse, and suicidal ideation (e.g., Clements-Noelle, Marx, & Katz, 2006; Grant et al., 2012).

Minority stress theory (Meyer, 1995) also elucidates how sexual minority persons may internalize the negativity to which they are subjected. Hendricks and Testa (2012) contend that a similar process may occur for transgender individuals. Indeed, such internalized transphobia is the result of the negative messages consistently experienced by transgender persons and is compounded by the belief that one must consciously refrain from disclosing such an identity. Since transphobia evidenced by the cisgender majority contributes to a social climate that is harmful to transgender persons' overall wellness, more research on this topic is needed.

A major impediment to studying transphobia is the absence of information detailing the psychometric integrity of measures used to assess this construct and whether one or more of these measures best approximates a "gold standard." Without a systematic review that clearly elucidates which indices of transphobia most closely adhere to recommended practices in scale development and refinement, researchers may continue generating ad hoc measures of varying psychometric quality—a situation that makes it difficult to form a coherent interpretation of the research landscape.

The purpose of the current study is to address this gap by reviewing all instruments that have been used to measure transphobia and, in so doing, identify which are the strongest psychometrically. To assist in making this determination, each measure was evaluated across five psychometric properties: (1) content validity; 2) factor structure (i.e., dimensionality); (3) scale score reliability; (4) criterion-related validity (typically, concurrent); and (5) construct validity (i.e., convergent and discriminant validities). The best practice guidelines that were used when evaluating these five components are delineated, followed by a systematic review of transphobia measures.

Content validity

This term refers to the extent to which all features of a scale, including its response format, items, and instructions, are "relevant to and representative of the targeted construct" (Haynes, Richard, & Kubany, 1995, p. 238). In this context, relevance means that scale items tap into the construct of interest rather than related constructs and representativeness refers to the entire domain of content being captured by the scale items (Vogt, King, & King, 2004). Yaghmaie (2009) suggests that content validity is established from three distinct sources: literature, relevant stakeholders, and experts. First, an exhaustive review of literature germane to the construct should occur. Second, representatives or stakeholders from relevant

populations need to be consulted for the unique insights they can provide about the construct. Finally, individuals with expertise relevant to the construct being assessed need to be targeted as they provide a wealth of knowledge that likely exceeds what can be gleaned from a literature review. While consulting experts is undoubtedly important, it should not be viewed as a suitable replacement for consultations with representatives from target populations. Content validity may be regarded as a critical first step in the validation of any measure (i.e., the intelligibility of other forms of validation depends upon scale items being content valid). To assess content validity, it is recommended that experts and members of the target population 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 the dimensions of relevance and representativeness for content experts and on elements such as item clarity and straightforwardness of instructions for members of the target population.

Factor structure

If a scale contains three or more items, it is critical to assess its factor structure using exploratory factor analysis (EFA) and/or confirmatory factor analysis (CFA). 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 [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) in reference to scales assessing discrimination toward sexual minority persons, tests of dimensionality were seldom conducted (Morrison, Bishop, Morrison, & Parker-Taneo, 2016); and (2) within the broader realm of sexological studies, researchers using EFA rarely adopted best practices for this statistical technique (Sakaluk & Short, 2017). 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 analyses were actually principal component analysis (PCA), which is not a form of EFA; 49.5% used orthogonal rotation (typically varimax); and 51.4% utilized the eigenvalue greater than 1.0 "rule" to assist with factor retention decision-making. None of these choices is optimal. First, PCA are recommended for data reduction as it does not take into account the underlying latent structure or structures of a group of scale items (Osborne & Costello, 2009). If the purpose is to examine a scale's dimensionality, then "true" EFA methods should be employed. The two EFA methods that are recommended most often are maximum-likelihood (ML) estimation and principal axis factoring (PAF), with the latter being advised when assumptions of data normality are "violated severely" (Sakaluk & Short, 2017, p. 3). Second, orthogonal rotation assumes that generated factors are uncorrelated (Pett, Lackey, & Sullivan, 2003); this assumption seldom 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, which permits but does not require factors to be intercorrelated, is the preferred choice (Sakaluk & Short, 2017). Third, in terms of factor retention, there is no support for the arbitrary eigenvalue greater than 1.0 "rule," with simulation studies revealing that it can lead to overfactoring or underfactoring (Osborne & Costello, 2009; Sakaluk & Short, 2017). Rather, parallel analysis (see O'Connor, 2000) for SPSS and SAS syntax) in conjunction with other retention criteria (e.g., interpretability and adherence to current theorizing) should be used.

Scale score reliability

The most popular reliability estimate is Cronbach's alpha, which may be conceptualized as the "expected correlation between an actual test and a hypothetical alternative form of the same length" (Carmines & Zeller, 1979, p. 45). It should be noted that reliability is not a fixed property of a scale but, rather, a characteristic of scale scores (Streiner, 2003). Thus, it must be calculated *each* time a researcher distributes a multi-item measure with the intention of averaging or summing the items. Although .80 is routinely identified as the cut-off for "good" scale score reliability,

there may be instances for which appreciably lower Cronbach's alpha coefficients are acceptable (see Schmitt, 1996). It also is recommended that 95% confidence intervals be computed for Cronbach's alpha (Fan & Thompson, 2001). The upper- and lower-bound estimates represent the range of plausible values for the alpha coefficient should the scale in question be distributed to the larger population from which the sample was drawn (Cumming & Finch, 2005). Finally, a critical, though routinely overlooked, 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 in 66.5% ($n = 533$) of these tests internal consistency was estimated using Cronbach's alpha. McCrae, Kurtz, Yamagata, and Terracciano (2011) contend that test-retest reliability has been neglected because: (1) it is inconvenient to distribute measures to the same participants at two different points in time and (2) researchers appear to operate from the misapprehension that forms of scale score reliability such as Cronbach's alpha and test-retest are interchangeable (i.e., they provide duplicate information). However, this is not the case, as 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

Globally speaking, validity refers to the degree to which a scale measures the construct it was designed to assess. Carmines and Zeller (1979) note that there are specific categories of validity: criterion-related, which may be partitioned into concurrent and predictive forms, and construct, which may be divided into convergent and divergent (or discriminant). Both forms of criterion-related validity focus on how well a newly developed measure correlates with a "gold standard" indicator of the *same* construct. The difference between concurrent and predictive validation is that, with the former, the gold standard is completed contemporaneously with the new measure, whereas with the latter, the gold standard is distributed at some future point in time. The stronger the association between scores on the new measure and scores on the

gold standard, the greater the level of criterion-related validity (Kimberlin & Winterstein, 2008). There are many times when a gold standard measure does not exist either because the scale one has developed assesses a novel construct or because extant measures of the construct are psychometrically deficient. Under such circumstances, convergent and divergent (or discriminant) validity are tested.

Convergent validity examines whether scores on the new scale correlate—in the predicted direction—with other constructs that, for theoretical and/or empirical reasons, should be associated with the new scale. To test divergent validity, one identifies constructs that—again, for theoretical and/or empirical reasons—should have a negligible relationship with scores on the new scale (Springer, Abell, & Hudson, 2002). Both convergent and divergent validity involve testing multiple predictions, with each supported prediction constituting one piece of evidence in support of the scale's construct validity (Carmines & Zeller, 1979). Given the dynamism of a construct's nomological network (i.e., the theoretical system in which the construct is embedded), construct validation is incremental and iterative. In other words, construct validation is an ongoing process that involves testing an array of predictions.

Method

Search strategy

A search of academic databases (Google Scholar, PsycINFO, PsycARTICLES, and PsycTESTS) was conducted to identify relevant transphobia studies. The authors' home institution allows for searching PsycINFO, PsycARTICLES, and PsycTESTS simultaneously using the same Boolean search protocol. This protocol included "transnegativity OR transnegative OR transphobia OR transphobic OR transprejudice OR gender nonconforming phobia OR gender nonconforming negativity OR gender variant phobia OR gender variant negativity." The number of results identified was 674. For Google Scholar, a similar Boolean search strategy was employed. The search term entered was "transnegative OR transphobic OR transnegativity OR transphobia OR transprejudice AND Measure* OR Instrument* OR Scale OR Index OR Inventor* OR Questionnaire* OR Test*"; this search yielded 2,081 results. (It is important to note that Google Scholar cannot display more than 1,000 search

results. Thus, entries from 1,001 to 2,081 could not be viewed. However, as other databases were used [e.g., PsycINFO] and the reference sections for all identified articles were inspected, it is unlikely that influential measures of transphobia were overlooked as a result of Google Scholar's limitation.) No specific dates were entered into either search so that each would yield as many results as possible. In addition, using advanced-search options, any search results that were not written in English and were not published in a peer-reviewed outlet were automatically excluded.

Inclusion and exclusion criteria

Articles were retained if they met the following criteria: (1) published in a peer-reviewed outlet; (2) published in English; and (3) quantitatively measured transphobia as a singular construct. Articles that quantitatively measured transphobia but were never published, measured transphobia as part of a composite LGBT prejudice and/or discrimination scale, were theoretical in nature, or were qualitative were excluded.

Article selection

Following the application of these inclusion criteria, of the 674 articles identified in the database search, 589 were identified as being theoretical in nature, qualitative, or not related to transgender issues. Twenty-four of these articles measured transphobia as part of a composite instrument examining prejudice and discrimination against LGBT persons as a whole. This yielded a total of 60 published articles from the database search. The same 60 articles were located during review of the first 1,000 search results using Google Scholar along with a single additional article that also quantitatively measured transphobia. (The same 24 articles that measured transphobia as part of a larger composite measure were identified.) In total, 61 peer-reviewed articles were located which used a total of 83 transphobia measures. See Figure 1.

Review procedure

Each retained article was reviewed to determine the extent to which it adhered to best practices in relation to the aforementioned psychometric elements (i.e., content validity, factor structure, scale score reliability, and validity [criterion-related and construct]). Please
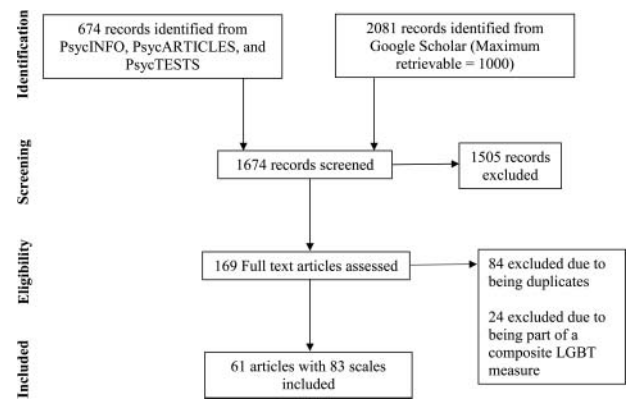
Identification
674 records identified from PsycINFO, PsycARTICLES, and PsycTESTS
2081 records identified from Google Scholar (Maximum retrievable = 1000)

Screening
1674 records screened
1505 records excluded

Eligibility
169 Full text articles assessed
84 excluded due to being duplicates
24 excluded due to being part of a composite LGBT measure

Included
61 articles with 83 scales included

Figure 1. Flowchart detailing literature search.

see Table 1 for an overview of the psychometric properties of the measures of transphobia included in this systematic review. If evidence of the psychometric property in question was established, a check mark (√) was entered. An "X" signified that no details about the psychometric element in question were provided. A question mark (?) was utilized in cases wherein (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 transgender persons or content experts informed item development/refinement). Finally, "N/A" (not applicable) was assigned in cases wherein a given psychometric element was not germane 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).

Given that psychometric testing is an incremental process, it was anticipated that researchers using a given scale would build upon the work of their predecessors. Thus, there were instances where a "√" or "?" was granted because the authors were utilizing a measure that had been rigorously assessed in previous studies. To illustrate, if Research Team A conducted a thorough content validation assessment of its measure of transphobia, it was not expected that other researchers using the same measure, within a similar cultural context, would need to start from scratch (i.e., they would not have to repeat the various steps involved in testing content validity). A measure's scale score reliability and factor structure, however, are score specific (rather than instrument specific) and,

Table 1. Psychometric properties of the reviewed scales.

#	Measure	C	FS	R	C-R	CV	TS
1	Genderism and Transphobia Scale–Revised	?	✓	✓	✓	✓	4
2	Genderism and Transphobia Scale–Short Form	?	✓	✓	✓	✓	4
3	Negative intentions subscale	?	X	✓	✓	✓	3
4	Negative attitudes subscale	?	X	✓	✓	✓	3
5	Attitudes Toward Hijra Scale	✓	?	✓	X	✓	3
6	Attitudes Toward Transgendered Individuals Scale	?	?	✓	✓	✓	3
7	Transphobia Scale	?	X	✓	X	✓	2
8	Genderism and Transphobia Scale–Short Form (Spanish)	?	?	✓	X	✓	2
9	Attitudes Toward Transgendered Individuals Scale	?	X	X	✓	✓	2
10	Genderism and Transphobia Scale	?	X	✓	X	✓	2
11	Transgender Attitudes and Beliefs Scale	?	?	✓	N/A	✓	2
12	Chinese Attitudes Towards Transgenderism and Transgender Civil Rights Scale	✓	X	✓	X	?	2
13	Transphobia Scale	?	X	✓	X	✓	2
14	Attitudes Toward Transgendered Individuals (Thai)	?	X	✓	X	✓	2
15	Transphobia Scale	?	?	✓	X	✓	2
16	Transphobia Scale	?	X	✓	X	✓	2
17	Genderism and Transphobia Scale	?	X	✓	X	✓	2
18	Genderism and Transphobia Scale (Chinese)	?	?	✓	X	✓	2
19	Attitudes Toward Transgender Students Questionnaire	?	X	✓	X	?	1
20	Attitudes Toward Transgender Students Questionnaire	?	X	✓	X	?	1
21	Social Distance Scale (in response to a trans target)	X	X	✓	X	?	1
22	Negative Attitudes Toward Transsexuals Scale	X	X	✓	X	?	1
23	Beliefs in Myths About Transsexuals Scale	X	X	✓	X	?	1
24	Predicted Discriminatory Behaviors Against Transsexuals Scale	X	X	✓	X	?	1
25	Genderism and Transphobia Scale (Chinese)	?	X	✓	X	?	1
26	Negative Attitudes Toward Transgender Individuals	?	X	✓	X	?	1
27	Attitudinal measure of perceived appropriateness of transgender behavior in children	X	X	✓	X	?	1
28	Behavioral intentions if respondents were various authority figures in a hypothetical (trans) child's life	X	X	✓	X	?	1
29	Attitude Toward Transgendered Individuals Scale	?	X	X	X	✓	1
30	Modified Transphobia Scale	?	?	✓	X	?	1
31	Genderism and Transphobia Scale	?	X	✓	X	?	1
32	Genderism and Transphobia Scale	?	X	X	X	✓	1
33	Transgender Attitudes and Beliefs Scale	?	X	X	N/A	✓	1
34	Attitudes Toward Transgender Individuals Scale	?	X	X	✓	X	1
35	Transgender Attitudes and Beliefs Scale	?	X	X	N/A	✓	1
36	Behavioral, cognitive, and affective dimensions of attitudes toward trans persons (4 each)	X	X	✓	X	?	1
37	Adapted Counselor Attitudes Toward Transgender Scale	?	X	✓	X	?	1
38	Adapted Attitudes Toward Transgender Individuals Scale	?	X	✓	X	?	1
39	Adapted Counselor Attitudes Toward Transgender Scale	?	X	✓	X	?	1
40	Adapted Attitudes Toward Transgender Individuals Scale	?	X	✓	X	?	1
41	Attitudes Towards the Inclusion of Transgender Women in Domestic Violence Services Scale	?	X	✓	X	?	1
42	Measure of comfort working with transgender women	?	X	✓	X	?	1
43	Revised Genderism and Transphobia Scale (Australian)	?	?	✓	X	?	1
44	Adapted Counselor Attitudes Toward Transgender Scale	?	X	✓	X	?	1
45	Revised Genderism and Transphobia Scale (Australian)	?	X	✓	X	?	1
46	Music Teacher's Attitudes Toward Transgender Individuals Scale	?	X	✓	X	?	1
47	Transphobia Scale	?	X	✓	X	?	1
48	Trans Persons Beliefs Scale	X	X	✓	X	X	1
49	Trans Persons Civil Rights Scale	X	X	✓	X	?	1
50	Genderism and Transphobia Scale	?	X	✓	X	?	1
51	Genderism and Transphobia Scale	?	X	✓	X	?	1
52	Social Distance Scale (in response to a trans target)	X	X	✓	X	?	1
53	Genderism and Transphobia Scale	?	X	✓	X	?	1
54	Transphobia Scale	?	X	✓	X	?	1
55	Frequency with which participants hear transphobic language at school	✓	N/A	N/A	X	X	1
56	Attitudes toward and discrimination against trans persons	?	X	✓	X	?	1
57	Acceptance of trans persons using specific gendered spaces	X	X	✓	X	?	1
58	Modified Genderism and Transphobia Scale	?	X	✓	X	?	1
59	Genderism and Transphobia Scale (Filipino)	?	X	✓	X	?	1
60	Attitudes toward and discrimination against trans women	?	X	✓	X	X	1
61	Attitudes Toward Trans Individuals Scale	?	X	✓	X	?	1
62	Attitudes Toward Transgender Individuals Scale	?	X	✓	X	?	1
63	Attitudes Toward Trans Persons Scale	?	X	✓	X	?	1
64	Attitudes Toward Trans Individuals Scale	?	X	✓	X	?	1
65	Attitudes Toward Transgender Individuals Scale	?	X	✓	X	?	1
66	Attitudes toward MtF trans individuals subscale	?	X	✓	X	?	1
67	Attitudes toward FtM trans individuals subscale	?	X	✓	X	?	1
68	Attitudes toward transindividuals subscale	?	X	✓	X	?	1
69	Measure of perceived responses to a parent disclosing a trans identity	X	X	X	X	?	0
70	Feeling thermometer (in response to a trans target)	N/A	N/A	N/A	X	?	0

Table 1. (*Continued*)

#	Measure	C	FS	R	C-R	CV	TS
71	Feeling thermometer measuring feelings toward trans persons	N/A	N/A	N/A	X	?	0
72	Attitudes toward transgender rights	X	?	?	X	?	0
73	Genderism and Transphobia Scale	?	X	X	X	?	0
74	Genderism and Transphobia Scale	?	X	X	X	X	0
75	Questionnaire About Transsexualism	X	N/A	N/A	X	?	0
76	Transphobia Inventory	X	X	X	X	?	0
77	Feeling thermometer measuring feelings toward trans persons	N/A	N/A	N/A	X	?	0
78	Single-item measure of transphobia	X	N/A	N/A	X	?	0
79	Genderism and Transphobia Scale	?	X	X	X	X	0
80	Genderism and Transphobia Scale	?	X	X	X	?	0
81	Measure of transphobia and adherence to masculine norms	?	?	X	X	?	0
82	Adapted Attitudes Toward Transsexualism Questionnaire	X	N/A	N/A	X	X	0
83	Measure of gender prejudice	X	X	X	X	?	0

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. Scales 1 and 2 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 Agee-Aguayo and associates. The number *19* appears in brackets at the end of this reference, which corresponds to item number 19 in Table 1 (Attitudes Toward Transgender Students Questionnaire). The number appearing appearing in brackets at the end of the Apperson and associates entry is 69, which corresponds to item number 69 in Table 1 (Measure of perceived responses to a parent disclosing a trans identity).

thus, should be tested with each new sample. When a researcher's focus is nonpsychometric, testing a measure's dimensionality may be perceived as adding unnecessary "clutter" to a results section. In such cases, it is recommended that these details be provided in an appendix. Finally, tests of construct validity did not have to be formally conducted as they were often nested within studies' key predictions. For example, if, as hypothesized, a researcher found a positive association between scores on a scale assessing transphobia and scores on variable *x*, this correlation would offer one strand of evidence supporting the construct (i.e., convergent) validity of the transphobia measure.

Results

Each of the targeted psychometric characteristics (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 then employed to contextualize key points.

Content validity

Of the 83 scales used across the 61 articles reviewed, only three (3.6%) used best practices to assess content validity. For example, to study transphobia in a Chinese context, King, Winter, and Webster (2009; see Table 1, entry 12) used the 38-item Chinese Attitudes Towards Transgenderism and Transgender Civil

Rights Scale (CATTCRS). In the article detailing the development of this instrument, King (2008) describes how semistructured interviews were conducted with 17 relevant stakeholders following an in-depth literature review. Some interviewees were cisgender individuals providing their personal understandings of transphobia and gender nonconformity in a Chinese context, while others were transgender individuals sharing their experiences with transphobia and their interpretation of the factors involved with this type of prejudice. Transcripts of the interviews were then analyzed for content with emergent themes being used to inform the development of content domains. Pilot testing and item revision occurred following these steps. Similarly, when creating the 32-item Attitudes Toward Hijra Scale, Jami and Kamal (2015; see Table 1, entry 5) conducted 10 focus-group discussions with self-identified hijra individuals and members of the general public. (*Hijra* is an umbrella term used to describe transgender, gender nonconforming, and other gender minority men in a Pakistani and Indian context.) A content analysis of transcribed discussions from these focus groups was used to inform the generation of scale items. The resultant item pool was refined under the guidance of a three-member expert panel before being pilot tested.

Thirty-one of the 83 scales (37.3%) earned a question mark since the studies did not provide sufficient details regarding their content validity. For example, Tebbe, Moradi, and Ege (2014; see Table 1, entries 1 &

2) created shorter, revised forms of Hill and Willough-by's (2005; see Table 1, entry 10) Genderism and Transphobia Scale (GTS) and Riggs, Webber, and Fell (2012; see Table 1, entry 45) modified the GTS to ensure it was suitable for use within an Australian context. Tebbe et al. (2014) employed statistical criteria such as factor loadings and inspection of the modification indices that accompany confirmatory factor analysis to reduce the number of items on the GTS. In contrast, Riggs et al. (2012) employed more subjective criteria. They altered items that had "problematic wording" (p. 55) and removed one question because it "was too broad for a survey focused exclusively on trans persons" (p. 56). None of these modifications, however, addressed the absence of clarity about the GTS's content validity. In the source article for the GTS, literature on "anti-trans sentiments and the difficulties trans persons have on a day-to-day basis" was reviewed (Hill & Willoughby, 2005, p. 534). While this is one facet of content validity, no mention is made of the initial or refined pool of items being reviewed by (1) experts in psychometric testing; (2) transgender individuals; or (3) members of the population to whom the scale would be distributed (i.e., cisgender persons). Given that the content validity of the GTS is questionable, subsequent changes to this measure—in the absence of consultation with content experts and laypersons—do not clarify whether the measure is or is not content valid.

Most the scales reviewed (46 out of 83; 55.4%) did not furnish any evidence about content validity. Instead, vague language was used such as, "adapted from" (e.g., Agee-Aguayo, Bloomquist, Savage, & Woitaszewski, 2017, p. 156; Costa & Davies, 2012, p. 1430); "a modified version of" (Riggs & Sion, 2017, p. 189); "adapted items from existing measures" (e.g., Barbir, Cohn, & Vandevender, 2017, p. 160); and "an adapted version of" (e.g., Riggs & Bartholomaeus, 2015, pp. 161 & 162; 2016, p. 11).In some cases, the adaptation involved taking a scale that was designed to assess prejudice toward a different social group and simply modifying the attitudinal target. To illustrate, Bowers, Lewandowski, Savage, and Woitaszewski's (2015; see Table 1, entry 20) Attitudes Toward Transgender Students Questionnaire (ATTSQ) was adapted from the Attitudes Toward Lesbians and Gay Men Scale (ATLG; Herek & McLemore, 2010). However, given that there is limited overlap in terms of item content, it is unclear how the ATLG was "adapted" to

become the ATTSQ. Of course, one might challenge the implicit assumption underlying this practice; namely, that the shared variance between homonegativity and transphobia is of such magnitude that measures designed to examine the former can, with only minor changes to item text, be used to examine the latter. For example, Hill and Willoughby (2005) found that transphobia and homonegativity were modestly intercorrelated ($r = .34$; Study 3), which is not congruent with the view that prejudice toward sexual- and gender-minority persons is interchangeable.

Finally, three (3.6%) scales were flagged as not applicable, "N/A." The measures in question were evaluation thermometers (i.e., single-item indicators where participants were asked to indicate their feelings toward transgender individuals on a scale from 0 to 100: Carroll, Güss, Hutchison, & Gauler, 2012; Cragun & Sumereau, 2015; Norton & Herek, 2013; see Table 1, entries 70, 71, & 77). Given that evaluation thermometers can be used with any attitudinal target, evidence of content validity was not deemed necessary.

Factor structure

The issue of scale dimensionality was relevant to any measure of transphobia that contained three or more items. Osborne and Costello (2009, p. 138) note that a factor with fewer than three items is "generally weak and unstable." Therefore, a three-item minimum is acceptable provided the scale in question is unidimensional. If a scale consists of more than one factor, a minimum of six items is recommended (i.e., three items per factor). Although 78 scales (94%) met this criterion, details about factor analysis were available for a much smaller number of measures ($k = 36$ out of 78; 46.2%). Inspecting this subset reveals that, akin to earlier reviews of psychometric "best practices" in the areas of sexology (Sakaluk & Short, 2017) and discrimination against gay and lesbian persons (Morrison et al., 2016), most researchers made suboptimal decisions when testing the factor structure of their measures of transphobia. Reliance on principal component analysis, varimax rotation, and/or the eigenvalue greater than 1 rule to assist with factor retention was common (e.g., Barbir, Cohn, & Vandevender, 2017; Costa & Davies, 2012; Elischberger, Glazier, Hill, & Verduzco-Baker, 2016; Hill & Willoughby, 2005; Nagoshi et al., 2008; Tee & Hegarty, 2006; Winter et al., 2009). Based on these observations, only two

scales (5.6%) (i.e., the 22-item *Genderism and Transphobia Scale–Revised* [GTS-R] and 13-item *Genderism and Transphobia Scale–Short Form* [GTS-SF]; Tebbe et al., 2014) had their factor structure assessed using best practices. In their article, Tebbe et al. (2014) explicitly describe how they conducted an exploratory factor analysis using principal-axis factoring with promax rotation and used parallel analysis as a means of calculating factor retention. On the basis of these results, the authors conducted a CFA during creation of the GTS-SF.

Since the remaining 34 scales either did not adhere to best practices or inadvertently conflated principal component and factor analyses, they were flagged with an "X." An additional 29 (37.2%) scales were not examined for factor structure at all. This omission is problematic since many of the researchers who used these scales assumed they were either unidimensional or that items formed logically coherent groupings without any evidence supporting these assumptions. For example, using Hill and Willoughby's (2005) GTS, Grigoropoulos and Kordoutis (2015) examined transphobia among 238 undergraduate students attending various universities in Athens, Greece. The authors did not test the dimensionality of the GTS but, rather, treated the two-factor structure identified by Hill and Willoughby as a fixed property of the scale. It is unfortunate that Grigoropoulos and Kordoutis (2015; see Table 1, entry 73) did not test the replicability of the GTS's factor structure, especially as they were using this measure within a different cultural context (western Europe versus North America). A similar concern arises with Fisher et al.'s (2016; see Table 1, entry 29) use of the Attitudes Toward Transgendered Individuals Scale (ATTI: Walch, Ngamake, Francisco, Stitt, & Shingler, 2012a). The former researchers seem to operate from the assumption that the unidimensionality identified by the latter group will "hold" despite the use of culturally diverse samples (European versus American, respectively).

Seven of the scales were coded as "N/A." Two possessed more than a single item; however, each item was examined individually, effectively rendering them individual single item measures: the 15-item Attitudes Toward Transsexualism Questionnaire (ATTQ), used by Landén and Innala (2000; see Table 1, entry 75), and its adapted version used by Willoughby et al. (2010; see Table 1, entry 82). The remaining five scales were one-item measures. Finally, a "?" was issued in

cases for which insufficient details were provided ($k = 11$). For example, Watgen and Mitchell (2013; see Table 1, entry 81) report performing a "factor analysis" (p. 144) and give factor loadings for their attitudinal measure. They do 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

Seventy-eight (94%) of the scales reviewed consisted of two or more items. Of these, no indicator of scale score reliability was provided for 14 of the 78 (17.9%) measures. In two additional studies, the authors reported Cronbach's alpha coefficients obtained in previous research but did not compute them for their own sample (e.g., see Table 1, entries 34 & 74). As noted earlier, indicators of reliability such as Cronbach's alpha coefficient are data dependent and, thus, must be computed each time a summative scale is used.

Total scores and Cronbach's alpha coefficients were reported for 61 (98.4%) of the 62 remaining instruments. However, not all authors referred to Cronbach's alpha by name. To illustrate, Agee-Aguayo, Bloomquist, Savage, and Woitaszewski (2017; see Table 1, entry 19) mentioned "internal consistency reliability" (p. 5); however, it was clear they were describing Cronbach's alpha since they compared their coefficient to the alpha value computed in an earlier study. Similarly, Cabeldue, Cramer, Kehn, Crosby, and Anastasi (2016; see Table 1, entry 7) used the term "internal reliability" in reference to the coefficient obtained for their measure of transphobia. However, since they explicitly refer to Cronbach's alpha for every other scale, it seems reasonable to presume that the "internal reliability" coefficient is, in fact, Cronbach's alpha.

The last of the 62 scales was issued a "?" since, while the author contends that his three-item measure of "attitudes toward transgender rights" had "high internal consistency and reliability" (p. 403), no reliability coefficient was computed (Flores, 2015; see Table 1, entry 72). Instead, the researcher appears to infer

reliability on the basis of the three items having strong factor loadings (.69 to .82). It should be noted that items evidencing higher loadings on a given factor are more representative of the latent construct reflected by that factor than items having weaker loadings (Pett et al., 2003). Loadings, however, do not serve as a proxy for scale score reliability.

Finally, seven scales (8.4%) were coded as "N/A." Five received this rating because they consisted of a single question. Two were given this rating because, while they were multi-item measures, each scale item was analyzed individually, thereby, eliminating the need for computation of Cronbach's alpha coefficient (ATTQ: Landén and Innala (2000); Willoughby et al., 2010).

Validity

Criterion-related validity

Of the 83 scales being reviewed, none explicitly tested for criterion-related validity. However, seven scales (8.4%) did provide evidence of concurrent validity but referred to their analyses as indicants of construct validity. Nonetheless, these scales were flagged as having tested criterion-related validity even though their methods were mislabeled. For example, Walch et al. (2012a; see Table 1, entry 6) found strong negative correlations between their newly developed ATTI and the overall GTS score and its transphobia subscale score (lower scores on the ATTI denote greater transphobia compared to higher scores on the GTS). The authors also reported a moderate negative correlation between the ATTI and the gender-bashing subscale. One group of researchers claimed to be providing evidence of concurrent validity (see Table 1, entry 14) by reporting a strong negative correlation between their Thai version of the Attitudes Toward Transgendered Individuals Scale and scores on an indicant of homonegativity. However, this type of association is rightfully classified as a test of convergent validity. 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. For example, when Hill and Willoughby (2005) developed the GTS, it was the first measure of its kind to assess elements of transphobia. Kanamori, Cornelius-White, Pegors, Daniel, and Hulgus (2016a)

acknowledge that, while a small number of transgender attitudes scales have undergone psychometric testing, none of them prove entirely satisfactory. First, Kanamori et al. (2016a) expressed reservations about researchers' reliance on college students when constructing transphobia scales. Second, Kanamori et al. (2016a) noted that existing measures were simplistic; that is, their factor structures were one- or two-dimensional, which suggests they do not capture the complexity of cisgendered persons' attitudes toward transgender individuals. Given their misgivings about extant transphobia scales, Kanamori and colleagues' omission of criterion-related validity is not surprising (see Table 1, entry 11). Despite their concerns about existing measures of transphobia, Kanamori et al. (2016a) tested the association between scores on their newly developed scale (Transgender Attitudes and Beliefs Scale [TABS]) and Hill and Willoughby's (2005) GTS and Walch et al.'s (2012a) ATTI. The resultant correlations were .88 and .95, respectively, which Kanamori et al. (2016a) erroneously claim provides evidence of the construct validity of the TABS.

Construct validity

Of the scales reviewed, 21 (25.3%) contained direct evidence regarding construct validity, while 55 (66.3%) contained indirect evidence (i.e., while it was not mentioned explicitly, the testing of predictions furnished pieces of evidence that were adduced to support the construct validity of the transphobia measures). Only seven (8.4%) of the instruments did not offer any findings that could be used for construct validation purposes.

The following examples illustrate the differences between studies that were classified as providing direct versus indirect evidence of construct validity. Due to the emphasis placed on social norm adherence, Nagoshi et al. (2008; see Table 1, entry 13) hypothesized that scores on their newly developed Transphobia Scale (TS) would correlate strongly with scores on a measure of "homophobia" (i.e., individuals more inclined to derogate sexual-attraction variance also would be more likely to derogate gender variance). As predicted, a strong correlation was obtained for both male and female participants (rs $=$ .56). Thus, one strand of evidence was furnished in support of the TS's convergent validity. Also, as hypothesized, the authors found that scores on the TS did not correlate with theoretically unrelated constructs, such as

locus of control, self-esteem, impulsivity, and self-monitoring, thereby offering one strand of discriminant validation. Similarly, Carrera-Fernández, Lameiras-Fernández, Rodríguez-Castro, and Vallejo-Medina (2014; see Table 1, entry 8) predicted that scores on the two subscales of their short-form Spanish version of the GTS (transphobia/genderism and gender-bashing) would correlate with scores on measures of modern homophobia and ambivalent sexism. Results supported both predictions (i.e., as respondents' transphobia increased so, too, did their negativity toward gay and lesbian persons and women). Such findings provide separate strands of support for the convergent validity of Carrera-Fernández et al.'s Spanish version of the GTS.

The seven studies that received an "X" for this psychometric indicator provided no evidence of convergent or discriminant validity. To illustrate, Winter et al. (2009) acknowledge that their 30-item questionnaire exploring attitudes toward and discrimination against transgender people has not yet undergone any empirical testing of its construct validity (see Table 1, entry 60).

The remaining 55 studies provided some indirect evidence of construct validity. For example, Apperson, Blincoe, and Sudlow (2015) developed an eight-item measure of perceived responses to a parent disclosing a transgender identity. While the authors do not state explicitly that they are testing the psychometric properties of their scale, inferences about known-groups validity can be made on the basis of the findings obtained. More specifically, based on past research suggesting that heterosexual men are less tolerant of non-heterosexuals, the authors predicted that "a male adult child would have more difficulty accepting transgender disclosure, particularly a father's disclosure, than a female adult child" (p. 434). As expected, female participants reported more positive attitudes than did male participants across all conditions. This gender difference provides one strand of support for the construct validity of the researchers' scale (see Table 1, entry 69).

Costa and Davies (2012) do not mention validity other than in the context of validation efforts that occurred in previous studies. However, the results showed that their 25-item scale, entitled Negative Attitudes Toward Transgender Individuals (NATI), strongly correlated with measures of affective reactions to gay men and lesbian women and attitudes toward gender roles (i.e., greater levels of transphobia were associated with greater levels of homonegativity

and stronger endorsement of traditional gender beliefs). The confirmation of these sorts of predictions may be used to make inferences about the construct validity of the NATI (see Table 1, entry 26).

Discussion

This systematic review indicates that most of the scales designed to measure transphobia did not adhere to best practice recommendations for scale development and validation. Expected elements of psychometrically sound measures such as tests of dimensionality and validity (criterion-related, convergent, and discriminant) were often omitted. While a majority of the studies reviewed did furnish evidence of scale-score reliability, the overwhelming metric reported was Cronbach's alpha, which relies upon various assumptions that are seldom met by social scientific data (see Dunn, Baguley, & Brunsden, 2014). None of the studies computed Omega or the Greatest Lower Bound estimate, two indices of reliability that are recommended by psychometrists (e.g., Dunn et al., 2014; Peters, 2014).

Each of the 83 measures appearing in the 61 articles reviewed were evaluated on the basis of five psychometric properties: (1) content validity; (2) factor structure; (3) scale score reliability; (4) criterion-related validity (typically, concurrent); and (5) construct validity (i.e., convergent and divergent). A check mark (✓) for each indicator would result in a "perfect" score of five points. Table 1 reveals that none of the reviewed scales received a perfect score. The scales that were awarded the most points (four out of five) are the Genderism and Transphobia Scale–Revised (GTS-R) and the Genderism and Transphobia Scale–Short Form (GTS-SF) (Tebbe et al., 2014) (see Table 1, entries 1 & 2). The GTS-R and GTS-SF represent the strongest measures evaluated for use with North American, English-speaking samples because they provide sufficient strands of evidence attesting to scales' factor structure, reliability, criterion-related validity, and construct validity. These scales fall short in the area of content validation as they were developed based on the original GTS, a scale whose content validity was indeterminate (see Table 1, entry 10).

The results of this systematic review underscore the importance of becoming familiar with and adhering to best practice recommendations when developing or revising attitudinal measures. Therefore, researchers

wishing to create novel measures of transphobia are recommended to proceed through the following series of steps. Please note that these guidelines are not sponsored, endorsed, or officially recognized by any governing bodies. They are presented here as best practice guidelines as identified by the authors of this article.

First, a large pool of items should be created based on a thorough review of the relevant literature and input from transgender persons. Second, the item pool should be evaluated, both qualitatively and quantitatively, by context experts (e.g., academics publishing in the areas of prejudice and discrimination toward trans men and trans women and toward individuals that have been the recipient of transphobia). Input from members of the population that would be completing the scale (i.e., cisgender individuals) also would be valuable to ensure they understand the terminology used in the items and scale instructions. Third, the revised pool should be pilot tested and then winnowed based on the results (i.e., items with low interitem correlations should be removed as should items with restricted frequency distributions). Fourth, the penultimate pool of items should be given to a new sample along with measures designed to test convergent and divergent validity. Fifth, at this juncture, exploratory factor analysis (EFA) should be conducted. The recommended extraction method is ML estimation or principal axis factoring (PAF); parallel analysis also should be used to assist with factor retention decision-making; and oblique rotation is advised. To maximize the likelihood of obtaining simple structure, the application of rigorous factor loading criteria is helpful (e.g., for any item, its minimum factor loading is .50 and it has no cross loadings greater than .29). Sixth, the retained items in conjunction with other measures designed to test construct validity should be distributed to a new sample. Ideally, a subset of this sample would be used for test-retest purposes. Finally, if large enough, it is recommended that the sample be split in two. EFA should be conducted with the first half, and confirmatory factor analysis (CFA) should be conducted with the second half. The latter analysis would permit researchers to determine the reproducibility of the scale's factor structure.

While practical exigencies may constrain some researchers' ability to proceed through all of these steps, it is recommended that, prior to omitting one of the psychometric elements detailed in this systematic review, researchers consider the costs of suboptimal scale development—costs that extend beyond the statistical and theoretical and permeate the lives of transgender persons everywhere.

Funding

References

Agee-Aguayo, J., Bloomquist, E., Savage, T. A., & Woitaszewski, S. A. (2017). School psychologists in support of transgender and gender diverse students in light of California's AB 1266 (School Success and Opportunity Act): Implications and opportunities. *Contemporary School Psychology*, *21*(2), 152–165. doi:10.1007/s40688-016-0101-2 [19](The number in square brackets corresponds to its listing in Table 1.)

Apperson, J. M., Blincoe, S., & Sudlow, J. L. (2015). An exploratory study of young adults' attitudes toward parental disclosure of LGBT identity. *Psychology of Sexual Orientation and Gender Diversity*, *2*(4), 492–496. doi:10.1037/sgd0000111 [69]

Barbir, L. A., Cohn, T. J., & Vandevender (2017). Friendship, attitudes, and behavioral intentions of cisgender heterosexuals toward transgender individuals. *Journal of Gay & Lesbian Mental Health*, *21*(2), 154–170. doi:10.1080/19359705.2016.1273157 [3, 4]

Batterham, A. M., & George, K. P. (2003). Reliability in evidence-based clinical practice: A primer for allied health professionals. *Physical Therapy in Sport*, *4*(3), 122–128. doi.org/10.1016/S1466-853X(03)00076-2

Bowers, S., Lewandowski, J., Savage, T. A., & Woitaszewski, S. A. (2015). School psychologists' attitudes toward transgender students. *Journal of LGBT Youth*, *12*(1), 1–18. doi:10.1080/19361653.2014.930370 [20]

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*. Advance online publication. doi:10.1177/0886260516636391 [7]

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Newbury Park, CA: Sage.

Carrera-Fernández, M. V., Lameiras-Fernández, M., Rodríguez-Castro, Y., & Vallejo-Medina, P. (2014). Spanish adolescents' attitudes toward transpeople: Proposal and validation of a short form of the Genderism and Transphobia Scale. *The Journal of Sex Research*, *51*(6), 654–666. doi:10.1080/00224499.2013.773577 [8]

Carroll, L., Güss, D., Hutchison, K. S., & Gauler, A. A. (2012). How do U.S. students perceive trans persons? *Sex Roles*, *67*(9), 516–527. doi:10.1007/s11199-012-0193-8 [21, 70]

Case, K. A., & Stewart, B. (2013). Intervention effectiveness in reducing prejudice against transsexuals. *Journal of LGBT Youth*, *10*(1–2), 140–158. doi:10.1080/19361653.2012.718549 [22, 23, 24]

Charter, R. A. (2003). A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *The Journal of General Psychology*, *130*, 290–304. doi.org/10.1080/00221300309601160

Chen, B., & Anderson, V. N. (2017). Chinese college students' gender self-esteem and trans prejudice. *International Journal of Transgenderism*, *18*(1), 66–78. doi 10.1080/15532739.2016.1256803 [25]

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

Costa, P. A., & Davies, M. (2012). Portuguese adolescents' attitudes toward sexual minorities: Transphobia, homophobia, and gender role beliefs. *Journal of Homosexuality*, *59*(10), 1424–1442. doi:10.1080/00918369.2012.724944 [26]

Cragun, R. T., & Sumerau, E. (2015). The last bastion of sexual and gender prejudice? Sexualities, race, gender, religiosity, and spirituality in the examination of prejudice toward sexual and gender minorities. *Journal of Sex Research*, *52*(7), 821–834. doi:10.1080/00224499.2014.925534 [71]

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, *60*(2), 170–180. doi:10.1037/0003-066X.60.2.170

De Jong, D. (2015). Transgender issues and BSW programs: Exploring faculty perceptions, practices, and attitudes. *Journal of Baccalaureate Social Work*, *20*(1), 199–218. doi:10.18084/1084-7219.20.1.199 [9]

Dunn, T. J., Baguley, T., & Brunsden, 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.

Elischberger, H. B., Glazier, J. J., Hill, E. D., Verduzco-Baker, L. (2016). "Boys don't cry"—or do they? Adult attitudes toward and beliefs about transgender youth. *Sex Roles*, *75*(5), 197–214. doi:10.1007/s11199-016-0609-y [27, 28]

Ellis, S. J., Bailey, L., & McNeil, J. (2016). Transphobic victimisation and perceptions of future risk: A large-scale study of the experiences of trans people in the UK. *Psychology & Sexuality*, *7*(3), 211–224. doi:10.1080/19419899.2016.1181669

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*(3), 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*(4), 517–531. doi:10.1177/0013164401614001

Fisher, A. D., Castellini, G., Fanni, E., Casale, H., Tagliagambe, M., Benni, L., & Maggi, M. (2016). Transphobia and homophobia levels in gender dysphoric individuals, general population and health care providers. *Journal of Sexual Medicine*, *13*(5–S2), S124. Retrieved from http://www.jsm.jsexmed.org/article/S1743-6095(16)00472–0/pdf [29]

Fisher, T. D., Davis, C. M., Yarber, W. L., & Davis, S. L. (2011). *Handbook of sexuality-related measures* (3rd ed.). New York, NY: Routledge.

Flores, A. R. (2015). Attitudes toward transgender rights: Perceived knowledge and secondary interpersonal contact. *Politics, Groups, and Identities*, *3*(3), 398–416. doi:10.1080/21565503.2015.1050414 [72]

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*(2), 76–99. doi:10.1080/15532739.2014.937041 [30]

Gerhardstein, K. R., & Anderson, V. N. (2010). There's more than meets the eye: Facial appearance and evaluations of transsexual people. *Sex Roles*, *62*(5), 361–373. doi:10.1007/s11199-010-9746-x [31]

Grant, J. M., Mottet, L. A., Tanis, J., Harrison, J., Herman, J. L., & Keisling, M. (2012). *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

Grigoropoulos, I., & Kordoutis, P. (2015). Social factors affecting antitransgender sentiment in a sample of Greek undergraduate students. *International Journal of Sexual Health*, *27*(3), 276–285. doi:10.1080/19317611.2014.974792 [73]

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 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*(5), 460–467. doi:10.1037/a0029597

Herek, G. M., & McLemore, K. A. (2010). Attitudes toward lesbians and gay men scale. In T. D. Fisher, C. M. Davis, W. L. Yarber, & S. L. Davis (Eds.), *Handbook of sexuality-related measures* (3rd ed., pp. 415–417). New York, NY: Routledge.

Hill, D. B., & Willoughby, B. L. B. (2005). The development and validation of the Genderism and Transphobia Scale. *Sex Roles*, *53*(7/8), 531–544. doi:10.1007/s11199-005-7140-x [10]

Hill, D. B., Menvielle, E., Sica, K. M., & Johnson, A. (2010). An affirmative intervention for families with gender variant children: Parental ratings of child mental health and gender. *Journal of Sex and Marital Therapy*, *36*(1), 6–23. doi:10.1080/00926230903375560 [32]

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

Jami, H., & Kamal, A. (2015). Measuring attitudes toward *hijras* in Pakistan: Gender and religiosity in perspective. *Pakistan Journal of Psychological Research*, *30*(1), 151–187. Retrieved from https://www.researchgate.net/publication/282322976_Measuring_Attitudes_toward_Hijras_in_Pakistan_Gender_and_Religiosity_in_Perspective [5]

Kanamori, Y., & Cornelius-White, J. H. D. (2016). Big changes, but are they big enough? Healthcare professionals' attitudes toward transgender persons. *International Journal of Transgenderism*, *17*(3–4), 165–175. doi:10.1080/15532739.2016.1232628 [33]

Kanamori, Y., Cornelius-White, J. H. D., Pegors, T. K., Daniel, T., & Hulgus, J. (2016a). Development and validation of the transgender attitudes and beliefs scale. *Archives of Sexual Behavior*. Advance online publication. doi:10.1007/s10508-016-0840-1 [11, 34, 74]

Kanamori, Y., Pegors, T. K., Hulgus, J. F., & Cornelius-White, J. H. D. (2017). A comparison between self-identified evangelical Christians' and nonreligious persons' attitudes toward transgender persons. *Psychology of Sexual Orientation and Gender Diversity*, *4*(1), 75–86. doi:10.1037/sgd0000166 [35]

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. E. (2008). *Transprejudice in Hong Kong: Chinese attitudes towards transgenderism and transgender civil rights* (Doctoral Dissertation). Retrieved from The University of Hong Kong, U Scholars Hub at: doi:10.5353/th_b3984880

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*(1), 17–34. doi:10.1080/19317610802434609 [12]

Landén, M., & Innala, S. (2000). Attitudes toward transsexualism in a Swedish national survey. *Archives of Sexual Behavior*, *29*(4), 375–388. doi:10.1023/A:1001970521182 [75]

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*(1), 28–50. doi:10.1177/1088868310366253

Menon, P., & Sreenidhi, M. (2015). In a transphobic society. *International Journal of Scientific Research*, *4*(7), 528–529. Retrieved from http://worldwidejournals.in/ojs/index.php/ijsr/article/view/7261/7309 [76]

Meyer, I. H. (1995). Minority stress and mental health in gay men. *Journal of Health and Social Behavior*, *36*(1), 38–56. doi:10.2307/2137286

Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay and bisexual populations: Conceptual issues and research evidence. *Psychological Bulletin*, *129*(5), 674–697. doi:10.1037/0033-2909.129.5.674

Morrison, E. G. (2010). Transgender as ingroup or outgroup? Lesbian, gay, and bisexual viewers respond to a transgender character in daytime television. *Journal of Homosexuality*, *57*(5), 650–665. doi:10.1080/00918361003712103 [36]

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

Nagoshi, J. L., Adams, K. A., Terrell, H. K., Hill, E. D., Brzuzy, S., & Nagoshi, C. T. (2008). Gender differences in correlates of homophobia and transphobia. *Sex Roles*, *59*(7), 521–531. doi:10.1007/s11199-008-9458-7 [13]

Ngamake, S. T., Walch, S. E., & Raveepatarakul, J. (2013). Validation of the attitudes toward transgendered individuals scale in Thailand. *International Journal of Transgenderism*, *14*(4), 158–170. doi:10.1080/15532739.2013.834810 [14]

Norton, A. T., & Herek, G. M. (2013). Heterosexuals' attitudes toward transgender people: Findings from a national probability sample of U.S. adults. *Sex Roles*, *68*(11), 738–753. doi:10.1007/s11199-011-0110-6 [77]

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*(3), 396–402.

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*(2), 131–146.

Peters, G.-J. Y. (2014). The alpha and omega of scale reliability and validity. *The European Health Psychologist*, *16*(2), 56–69.

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*(5), 489–497. doi:10.1002/nur.20147

Riggs, D. W., & Bartholomaeus, C. (2015). The role of school counsellors and psychologists in supporting transgender people. *Australian Educational and Developmental Psychologist*, *32*(2), 158–170. doi:10.1017/edp.2015.19 [37, 38]

Riggs, D. W., & Bartholomaeus, C. (2016). Australian mental health professionals' competencies for working with trans clients: A comparative study. *Psychology and Sexuality*, *7*(3), 225–238. doi:10.1080/19419899.2016.1189452 [39, 40]

Riggs, D. W., Fraser, H., Taylor, N., Signal, T., & Donovan, C. (2016). Domestic violence service providers' capacity for supporting transgender women: Findings from an Australian workshop. *British Journal of Social Work*, *46*(8), 2374–2392. doi:10.1093/bjsw/bcw110 [41, 42]

Riggs, D. W., & Sion, R. (2017). Gender differences in cisgender psychologists' and trainees' attitudes toward

transgender people. *Psychology of Men and Masculinity*, *18*(2), 187–190. doi:10.1037/men0000047 [43, 44]

Riggs, D. W., Webber, K., & Fell, G. R. (2012). Australian undergraduate psychology students' attitudes towards trans people. *Gay & Lesbian Issues and Psychology Review*, *8*(1), 52–62. Retrieved from http://www.genderidentityaustralia.com/blog/wp-content/uploads/2013/06/Attitudes-Towards-Trans-People-GLIPR.pdf [45]

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. *The Journal of Sex Research*, *54*(1), 1–9. doi:10.1080/00224499.2015.1137538

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. doi:10.1037/1040-3590.8.4.350

Silveira, J. M., & Goff, S. C. (2016). Music teachers' attitudes toward transgender students and supportive school practices. *Journal of Research in Music Education, 64*(2), 138–158. doi:10.1177/0022429416647048 [46]

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*(3), 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*(1), 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*(4), 317–332. doi:10.1007/s13178-013-0136-3 [78]

Tebbe, E. A., & Moradi, B. (2012). Anti-transgender prejudice: A structural equation model of associated constructs. *Journal of Counseling Psychology*, *59*(2), 251–261. doi:10.1037/a0026990 [15]

Tebbe, E. A., Moradi, B., & Ege, E. (2014). Revised and abbreviated form of the genderism and transphobia scale: Tools for assessing anti-trans* prejudice. *Journal of Counselling Psychology*, *61*(4), 581–592. doi:10.1037/cou0000043 [1, 2, 47]

Tee, N., & Hegarty, P. (2006). Predicting opposition to the civil rights of trans persons in the United Kingdom. *Journal of Community & Applied Social Psychology*, *16*, 70–80. doi:10.1002/casp.851 [48, 49]

Thomas, D. M., Amburgey, J., & Ellis, L. (2016). Anti-transgender prejudice mediates the association of just world beliefs and victim blame attribution. *International Journal of Transgenderism*, *17*(3–4), 176–184. doi:10.1080/15532739.2016.1232627 [50]

Tompkins, T. L., Shields, C. N., Hillman, K. M., & White, K. (2015). Reducing stigma toward the transgender community: An evaluation of a humanizing and perspective-taking intervention. *Psychology of Sexual Orientation and Gender Diversity*, *2*(1), 34–42. doi:10.1037/sgd0000088 [51, 52]

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*(3), 231–243. doi:10.1037/1040-3590.16.3.231

Walch, S. E., Ngamake, S. T., Francisco, J., Stitt, R. L., & Shingler, K. A. (2012a). The attitudes toward transgendered individuals scale: Psychometric properties. *Archives of Sexual Behavior*, *41*(5), 1283–1291. doi:10.1007/s10508-012-9995-6 [6, 79]

Walch, S. E., Sinkkanen, K. A., Swain, E. M., Francisco, J., Breaux, C. A., & Sjoberg, M. D. (2012b). Using intergroup contact theory to reduce stigma against transgender individuals: Impact of a transgender speaker panel presentation. *Journal of Applied Social Psychology*, *42*(10), 2583–2605. doi:10.1111/j.1559-1816.2012.00955.x [80]

Walters, A. S., & Rehma, K. (2013). Avenue T: Using film as entrée in teaching about transgender. *Sex Education*, *13*(3), 336–348. doi:10.1080/14681811.2012.743460 [53]

Warriner, K., Nagoshi, C. T., & Nagoshi, J. L. (2013). Correlates of homophobia, transphobia, and internalized homophobia in gay or lesbian and heterosexual samples. *Journal of Homosexuality, 60*(9), 1297–1314. doi:10.1080/00918369.2013.806177 [54]

Watgen, J., & Mitchell, R. W. (2013). College men's concerns about sharing dormitory space with a male-to-female transsexual. *Sexuality & Culture*, *17*(1), 132–166. doi:10.1007/s12119-012-9143-4 [81]

Weiner, B. A., & Zinner, L. (2015). Attitudes toward straight, gay male, and transsexual parenting. *Journal of Homosexuality*, *62*(3), 327–339. doi:10.1080/00918369.2014.972800 [16]

Wernick, L. J., Kulick, A., & Inglehart, M. H. (2014). Influences of peers, teachers, and climate on students' willingness to intervene when witnessing anti-transgender harassment. *Journal of Adolescence*, *37*(6), 927–935. doi:10.1016/j.adolescence.2014.06.008 [55]

White, C. R., & Jenkins, D. D. (2017). College students' acceptance of trans women and trans men in gendered spaces: The role of physical appearance. *Journal of Gay & Lesbian Social Services, 29*(1), 41–67. doi:10.1080/10538720.2016.1261749 [56, 57]

Willoughby, B. L. B., Hill, D. B., Gonzalez, C. A., Lacorazza, A., Macapagal, R. A., Barton, M. E., & Doty, N. D. (2010). Who hates gender outlaws? A multisite and multinational evaluation of the Genderism and Transphobia Scale. *International Journal of Transgenderism*, *12*(4), 254–271. doi:10.1080/15532739.2010.550821 [17, 58, 59, 82]

Winter, S., Chalungsooth, P., Teh, Y. K., Rojanalert, N., Maneerat, K., Wong, Y. W., & Macapagal, R. A. (2009). Transpeople, transprejudice and pathologization: A seven-country factor analytic study. *International Journal of Sexual Health*, *21*, 96–118. doi:10.1080/19317610902922537 [60]

Winter, S., Webster, B., & Cheung, P. K. E. (2008). Measuring Hong Kong undergraduate students' attitudes towards transpeople. *Sex Roles*, *59*(9), 670–683. doi:10.1007/s11199-008-9462-y [18]

Worthen, M. G. F. (2012a). Heterosexual college student sexual experiences, feminist identity, and attitudes toward LGBT

individuals. *Journal of LGBT Youth*, *9*(2), 77–113. doi:10.1080/19361653.2012.649613 [61]

Worthen, M. G. F. (2012b). Understanding college student attitudes toward LGBT individuals. *Sociological Focus, 45*(4), 285–305. doi:10.1080/00380237.2012.712857 [62]

Worthen, M. G. F. (2014a). The interactive impacts of high school gay-straight alliances (GSAs) on college student attitudes toward LGBT individuals: An investigation of high school characteristics. *Journal of Homosexuality*, *61*(2), 217–250. doi:10.1080/00918369.2013.839906 [63]

Worthen, M. G. F. (2014b). Blaming the jocks and the Greeks? Exploring collegiate athletes' and fraternity/sorority members' attitudes toward LGBT individuals. *Journal of College Student Development*, *55*(2), 168–195. doi:10.1353/csd.2014.0020 [64]

Worthen, M. G. F. (2016). Hetero-cis-normativity and the gendering of transphobia. *International Journal of Transgenderism, 17*(1), 31–57. doi:10.1080/15532739.2016.1149538 [65, 66, 67]

Worthen, M. G. F., Lingiardi, V., & Caristo, C. (2016). The roles of politics, feminism, and religion in attitudes toward LGBT individuals: A cross-cultural study of college students in the USA, Italy, and Spain. *Sexuality Research and Social Policy*. Advance online publication. doi:10.1007/s13178-016-0244-y [68]

Yaghmaie, F. (2009). Content validity and its estimation. *Journal of Medical Education*, *3*, 25–27. Retrieved from http://www.journals.sbmu.ac.ir/jme/article/download/870/769

Yost, M. R., & Gilmore, S. (2011). Assessing LGBTQ campus climate and creating change. *Journal of Homosexuality, 58*(9), 1330–1354. doi:10.1080/00918369.2011.605744 [83]