

Psychology & Sexuality

ISSN: 1941-9899 (Print) 1941-9902 (Online) Journal homepage: http://www.tandfonline.com/loi/rpse20

What is the best measure of discrimination against trans people?: A systematic review of the psychometric literature

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

To cite this article: Melanie A. Morrison, CJ Bishop & Todd G. Morrison (2018) What is the best measure of discrimination against trans people?: A systematic review of the psychometric literature, Psychology & Sexuality, 9:3, 269-287, DOI: 10.1080/19419899.2018.1484798

To link to this article: https://doi.org/10.1080/19419899.2018.1484798

Accepted author version posted online: 06 Jun 2018.
Published online: 06 Jul 2018.

Submit your article to this journal

Article views: 82

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=rpse20

PSYCHOLOGY & SEXUALITY
2018, VOL. 9, NO. 3, 269–287
https://doi.org/10.1080/19419899.2018.1484798

Routledge
Taylor & Francis Group

Check for updates

What is the best measure of discrimination against trans people?: A systematic review of the psychometric literature

Melanie A. Morrison[a], CJ Bishop[a] and Todd G. Morrison[a]

[a]Department of Psychology Saskatoon, University of Saskatchewan, SK, Canada

ABSTRACT

To understand the levels and types of discrimination experienced by a minoritised group such as trans people, it is essential that researchers have access to psychometrically sound indicators of discrimination. While a surfeit of measures exist assessing trans individuals' experiences of transnegativity, to date, no systematic review of these instruments has been conducted. In the current study, 116 scales were evaluated on the basis of their adherence to best practices in psychometric testing. The findings indicated that, for most of the instruments assessed, limited information was provided about their psychometric properties (in particular, item development and refinement, factor structure, and scale score reliability and validity). The measures that evidenced strongest adherence to best practice recommendations in scale development are identified, and recommendations are made for the creation of new instruments assessing trans people's experiences of transnegativity.

ARTICLE HISTORY
Received 16 January 2018
Accepted 1 June 2018

KEYWORDS
Trans; transgender; transphobia; discrimination; stigma; psychometric; validity; reliability

Researchers have found that transgender and gender nonconforming persons (i.e. those experiencing a disjunction between their gender identity and their sex assigned at birth: Shulman et al., 2017) are subjected to alarming levels of discrimination. For example, in a recent report summarising key findings from the Trans Pulse Project (a study focusing on the social determinants of health among trans people residing in Ontario, Canada), Bauer and Scheim (2015) outline a number of sobering results. In terms of 'everyday' transnegativity (i.e. a multidimensional construct that includes the negative beliefs, attitudes and behaviours towards individuals who are perceived correctly, or incorrectly, as trans), 96% of participants (N = 433 ages 16+) reported overhearing that trans people were not 'normal' and 73% had been teased for being trans. Violence was also a common occurrence: 20% indicated that they had been physically or sexually assaulted for being trans and 34% reported being 'verbally threatened or harassed but not assaulted' (p. 4). The researchers also documented institutional discrimination. Thirteen percent of participants reported that they were fired due to their trans status; 18% were turned down for a job because they were trans; and 17% declined a job that had been offered to them, due to their perception that their future workplace environment was unsafe for trans people. Respondents also outlined discriminatory experiences within the institution of health, particularly within medical care settings. To illustrate: 10% had been refused emergency care or had care ended prematurely because they were trans; 25% stated that they had been 'belittled or ridiculed' (p. 3) by an emergency care provider; and 40% reported at least one episode of discrimination from their family physician.

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

This manuscript was made possible by an Insight Grant awarded to the senior author from the Social Sciences and Humanities Research Council of Canada (435-2016-1485).

© 2018 Informa UK Limited, trading as Taylor & Francis Group

These findings are echoed in the large-scale U.S. Transgender Survey conducted in 2015 by the National Center for Transgender Equality. Participants were 27,715 individuals (ages 18+), who occupied a range of identities along the transgender identity spectrum (e.g. trans, genderqueer and non-binary). Results indicated that a substantial proportion of respondents who were either out or were perceived as transgender while attending school (kindergarten to grade 12) experienced some form of discrimination: 54% had been verbally harassed; 52% had been prohibited from dressing in a way that was congruent with their gender identity; 24% had been physically attacked; and 13% had been sexually assaulted. (Seventeen percent reported leaving school because the discrimination they experienced was intolerable.) Within the employment domain, 49% of those reporting that they were denied a promotion in the year preceding the survey attributed this outcome to their trans status; 43% of those that were fired believed their trans identity was the causal factor as did 39% of those that had applied for, but were unsuccessful in obtaining, a given job. In terms of medical care, 33% of participants reported having one or more negative experiences from their healthcare provider in the year preceding the survey. These experiences included: having to teach their health practitioner about trans-related issues in order to receive adequate care (24%); being asked invasive/unnecessary questions about their transgender status (15%); and being denied care that was transition-related (8%). Overall, in the year preceding the survey, 46% of respondents had been verbally harassed; 9% had been physically attacked; and 10% had been sexually assaulted. It should be noted that such instances of victimisation were more likely to occur among trans people of colour.

Theorising by academics as to why prejudice, discrimination and violence are directed towards trans people, and perpetrated primarily by cisgender men and women (i.e. non-trans men and women), has yielded several plausible explanations. First, Stewart, O'Halloran, and Oates (2018) emphasize that individuals are placed in vulnerable positions and become susceptible to prejudice and discrimination when they 'stray' from societal norms. Indeed, in their study examining anti-trans prejudice amongst cisgender heterosexual psychologists, Riggs and Sion (2017) state that anti-trans attitudes are often maintained by cisgender individuals in an effort to retain 'traditional' gender paradigms (e.g. maintenance of rigid boundaries designed to separate genders) and promulgate the notion that one's gender is the direct result of one's assigned sex at birth. Secondly, Connell and Messerschmidt (2005) indicate that anti-trans attitudes and behaviours are most prevalent amongst cisgender men due to the confines of hegemonic masculinity, wherein cisgender men are taught to compare themselves with those who are more vulnerable in society, namely, women who are cisgender, men who are gay, and individuals who are transgender. Thirdly, anti-transgender prejudiced attitudes and discriminatory behaviours are often predicated upon the negative stereotyping that cisgender women and men engage in towards transgender women and transgender men. For instance, in the first study examining the content of cisgender men and women's stereotypes towards transgender women and men, Gazzola and Morrison (2014) identified stereotypes that emanated from eight themes (e.g. perceptions of mental illness, abnormality and rejection by society). In all, it appears that considerable variance in transnegativity might be theorised to be a result of the widespread hegemonic messages received about gender, with the implication being that any person who deviates from 'tradition', be outrightly rejected via stereotyping, prejudice and discrimination.

The sequelae of trans-based discrimination have been documented. For example, Miller and Grollman (2015) employed the Minority Stress Model (Meyer, 1995) to elucidate the possible linkages among being trans, experiencing discrimination, and decrements in psychological and physical well-being. The Minority Stress Model contends that individuals exposed to recurrent group-specific stressors (e.g. transnegativity) are more likely to engage in health-injurious behaviours, such as self-harm, smoking and drug use due to the depletion of coping resources which ensues from experiencing ongoing stress. Congruent with the Minority Stress Model, Miller and Grollman (2015) found that, among their sample of American trans participants (N = 4115), those experiencing more types of discrimination (either episodes of major[1] discrimination or everyday

discrimination) were more likely than those reporting fewer types of discrimination (or no discrimination at all) to have attempted suicide. Reporting more types of trans-based major or everyday discrimination was also associated with being more likely to abuse drugs/alcohol and smoke cigarettes.

In a series of findings emanating from the Trans Pulse Project (Bauer & Scheim, 2015), the deleterious sequelae of transnegativity are also underscored. To illustrate: approximately 67% of participants reported avoiding public spaces or situations 'due to fears of harassment, being perceived as trans, or being "outed" as trans' (p. 5) and 21% stated that, due to their trans status, they did not seek emergency care even when medically necessary. The proportion that evidenced symptomatology consistent with a diagnosis of clinical depression was 66.4%, with odds ratio analysis suggesting that each 'one point increase' on a measure of transnegativity[2] was 'associated with a 12% increase in the odds of depression' (Rotondi et al., 2011a, p. 142). Sixty-five percent of participants reported that they had seriously contemplated suicide because they were trans; 43% had attempted suicide, with 10% attempting suicide in the year preceding the survey (Bauer, Pyne, Francino, & Hammond, 2013). Akin to the relationship noted with depression and congruent with the Minority Stress Model, participants that reported experiencing a greater range of violence including physical or sexual assault due to their trans status were more likely to have considered or attempted suicide in the 12 months preceding the survey (Bauer et al., 2013). Similar findings have been obtained by other researchers (e.g. Bockting, Miner, Swinburne Romine, Hamilton, & Coleman, 2013; Clements-Noelle, Marx, & Katz, 2006; Grant et al., 2012; Rood, Puckett, Pantalone, & Bradford, 2015).

When conducting research on transgender and gender nonconforming persons' experiences of discrimination, a key issue involves measurement selection, that is, from an array of available indicators of transnegativity, which ones offer the most compelling evidence of psychometric integrity? Further, in the absence of manually reviewing all published studies on trans-based discrimination, how are researchers to make an optimal choice given the objectives of their research? The purpose of the current study is to address these queries by providing a systematic review of indicators of discrimination against transgender and gender nonconforming people and, in so doing, assist researchers to make informed measurement choices.

Each of the instruments that satisfies our inclusion criteria (to be outlined later) will be evaluated on five dimensions, which reflect core elements of classical test theory. These are: (1) content validity (CV); (2) factorial validity; (3) scale score reliability; (4) criterion-related validity (CRV); and (5) construct validity (CTV). Brief definitions of these elements as well as the best practice guidelines that we recommend researchers adopt when making determinations about a measure's psychometric integrity are outlined in the next section.

Content validity. Following examination of the theoretical and empirical work on the topic of interest, a researcher will look to develop scale items. Generating content for a scale is an initial step in scale construction, and one must develop scale content in a comprehensive manner in order to ultimately furnish evidence that the developed scale possesses CV. CV is concerned primarily with items' content relevance and representativeness (Lynn, 1986). Relevance pertains to whether the proposed instrument measures the construct of interest as opposed to other conceptually similar or related constructs (Vogt, King, & King, 2004). The issue of content relevance becomes particularly important in cases where researchers attempt to measure several phenomena in a single scale. For instance, in the field of sexual and gender minority studies, numerous researchers have developed scales designed to tap the discrimination experiences of 'LGBT' (i.e., lesbian, gay, bisexual, and transgender persons) people. Ostensibly, researchers treating 'LGBT' persons as a uniform group render the complexities facing the distinct social groups that comprise this acronym invisible (Morrison, Bishop, & Morrison, 2018). As such, content relevance is best attested to when a single group, topic or dimension are to formulate the construct of the scale.

Representativeness attempts to ensure that the items developed for a new measure reflect as much of the targeted construct as possible and that important aspects of the construct are not inadvertently overlooked (Lynn, 1986). To determine whether a measure is content valid, a researcher should conduct

an extensive review of the pertinent literature and should also consult with: (1) experts who can evaluative the initial pool of items and (2) stakeholders who have a vested interest in the proposed instrument and can give experiential details concerning the targeted construct (Haynes, Richard, & Kubany, 1995; Yaghmaie, 2009). With respect to measures of transnegativity, CV would be evident if researchers briefly noted that, to assist with item generation, they: (1) conducted a thorough review of the literature documenting transgender and gender nonconforming people's experiences with discrimination and 2) consulted experts in relevant fields of scholarship (e.g. trans studies and psychometrics), as well as laypersons (i.e. transgender and gender nonconforming individuals). The consultation with experts and laypersons would enable researchers to gauge the measure's content relevance and representativeness, and also determine more practical matters such as item clarity and ease of completion. It should be noted that, when developing a scale, researchers should always develop far more scale items than one will need (i.e. err on the side of over-representation). In many ways, CV represents the cornerstone of all subsequent validation testing; that is, if a new instrument is not content valid, then determinations of construct validity, for example, are immaterial. Indeed, regarding anti-transgender prejudice measures, Billard (2018) states that 'all the statistics in the world cannot redeem a scale that, at its most basic level, fails to reflect the construct it purports to measure' [p. 2]. Billard (2018) further asserts that there is a crisis in CV that supercedes all other statistical concerns.

Factorial validity. Once a researcher has developed a measure using a comprehensive scale item development strategy, and evidence of CV can be furnished, a researcher will engage in the collection of the desired amount of data. Once the necessary data are in hand, the researcher will follow with an examination of the scale's factor structure and assess whether the scale possesses factorial validity. In order for the assessment of a scale's factor structure to be appropriate, the scale in question must contain three or more scale items. (If the measure contains fewer than three items, there is no compelling reason to evaluate a measure's dimensionality.) If the three-or-more item criterion is satisfied, researchers should then utilise statistical software (e.g. SPSS) to conduct an analysis entitled exploratory factor analysis (EFA). EFA allows researchers to determine the ways in which the various scale items 'group' together either as a unidimensional (i.e. where all items load on a single factor) or multidimensional measure (i.e. where all items are spread across more than one factor which, ultimately, might result in a series of subscales being formed). In all, EFA is a data-driven approach and 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). The factor structure of a measure should *always* be assessed using EFA first, which can then be followed by what is referred to as confirmatory factor analysis (CFA). CFA is theory-driven and is advised 'when there is sufficient theoretical and empirical basis for a researcher to specify the model or small subset of models that is the most plausible' (Fabrigar et al., 1999, p. 277). Morrison, Morrison, and McCutcheon (2017b) indicate that a confirmatory approach is adopted when researchers are ultimately aware of the interrelationships that are theorised to exist amongst the items or variable(s) of interest, which enables researchers to test how well the theorised, and empirically tested, model fits the newly obtained (i.e. sample) data. For example, if a scale is theorised to have only one factor (i.e. all of the scale items are theorised as being unitary [wherein no subscales are presumed to exist], and empirical research shows that the scale is at least preliminarily unidimensional), then researchers would test a one-factor CFA model. A variety of CFA-related fit statistics would then be utilised to assess how well the theorised and empirically derived unidimensional model fits the sample data. The model may require 'improvement'; researchers are cautioned, however, that any modifications must be made based on theoretical, statistical or practical grounds. Ideally, researchers conduct an EFA and then confirm the resultant factor structure using CFA. Separate samples should be used for these statistical procedures as conducting both analyses with the same group of participants maximise the likelihood that the factor solution obtained may reflect idiosyncrasies of a given sample.

Previous systematic reviews suggest that few researchers adhere to best practice recommendations when conducting an EFA[3] (see Gaskin & Happell, 2014; Morrison et al., 2018; Morrison, Bishop, Morrison, & Parker-Taneo, 2016; Sakaluk & Short, 2017). To illustrate, 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). They found that 59.3% of the EFAs conducted were actually principal component analyses (PCA), which is *not* a form of EFA; almost half (49.5%) used orthogonal rotation (typically varimax) which constrains factors in multifactorial solutions to be uncorrelated; and 51.4% employed the eigenvalue greater than 1.0 'rule', which is used to help researchers make decisions about the number of factors to retain. None of these choices is recommended. First, as PCA maximises the variance explained in the fewest number of components possible (Osborne & Costello, 2009), it is better suited for data reduction (i.e. when the desire is to reduce the original number of scale items to a more manageable number) than for structure detection (i.e. determining how [i.e. which ones, how many and to what extent] the items load on various components). PCA also constrains components to be uncorrelated (i.e. orthogonal, meaning that no correlation is assumed to exist amongst the separate components), an assumption that is problematic because researchers seldom expect that social scientific data can be 'partitioned into neatly packaged units that function independently of one another' ([Osborne & Costello, 2009, p. 136]). EFA provides the option of oblique rotation which permits but does not require generated factors to be correlated (Schmitt, 2011). Again, many researchers would assume that there is some correlation, however miniscule, amongst different factors, each of which is comprised of a series of items thought to cohere with one another. Finally, with respect to the eigenvalues greater than one 'rule', simulation studies have revealed that it can lead to over- or under-factoring (Osborne & Costello, 2009; Sakaluk & Short, 2017). This means that too many factors or not enough factors will appear as the accurate number to retain. Parallel analysis, in conjunction with other retention criteria (e.g. the scree test), is recommended (O'Connor, 2000; Preacher & MacCallum, 2003).

Scale score reliability. Cronbach's alpha refers to the "expected correlation between an actual test and a hypothetical alternative form of the same length (Carmines & Zeller, 1979, p. 45), or 'the mean of all split-half coefficients resulting from different splitting of a test' (Cronbach, 1951, p. 297). Moreover, Cronbach's alpha is a 'proximal indicator of "true" reliability (i.e. the observed score is equal to the sum of the true score and a measurement error'; Eisinga, Grotenhuis, & Pelzer, 2013, p. 638). Based on the long-standing endorsement by significant researchers in the field of psychometrics, Cronbach's alpha is, overwhelmingly, the most popular indicant of scale score reliability.

One crucial aspect of reliability that is often overlooked by researchers is that it is a property of *scale scores* which change from sample to sample, and not a fixed characteristic of a measure (Caruso, 2000). This means that each time an instrument, in which total scale scores will be calculated, is distributed to participants, an indicator of reliability (e.g. Cronbach's alpha) *must* be computed. The generally accepted threshold indicating 'good' scale score reliability is .80; however, there are circumstances where less desirable Cronbach's alpha coefficients are acceptable (see Schmitt, 1996). Fan and Thompson (2001) note that, when computing Cronbach's alpha coefficients, it is also informative to include 95% confidence intervals. The upper and lower values for these intervals delimit the plausible range of alpha coefficients that should be obtained if the measure were distributed to another demographically similar sample (Cumming & Finch, 2005). Ideally, the confidence intervals will be fairly narrow (i.e. the broader they are, the greater the amount of random measurement error).

Test-retest reliability, which denotes the temporal stability of measurement scores, is valuable, though under-utilised (Hogan, Benjamin, & Brezinski, 2000). Hogan et al., for example, conducted a systematic review of 696 tests appearing in the American Psychological Association's *Directory of Unpublished Experimental Mental Measures*. Their results indicated that, whereas 66.5% of the instruments assessed used Cronbach's alpha to estimate internal consistency, only 19% employed test-retest (usually in combination with other indicators of scale score reliability). It is possible that test-retest

reliability may be avoided because: (1) it is more resource intensive (i.e. it necessitates that the *same* participants be measured at two points in time) and (2) researchers may operate from the erroneous assumption that different forms of reliability yield equivalent information (McCrae, Kurtz, Yamagata, & Terracciano, 2011). In actuality, only test-retest is capable of assessing temporal stability (Charter, 2003).

Validity. Once you have a scale that possesses content and factorial validity, and previous studies have provided evidence of your measure's scale score reliability, one is ready to test additional key indicants of psychometric soundness. Very simply, validity may be viewed as the degree to which a scale measures the construct it is designed to measure. Two broad types of validity are *criterion-related* and *construct* validity, each of which may be partitioned into distinct subtypes. For CRV, the subtypes are *concurrent* and *predictive* validity, whereas for construct validity, the subtypes are *convergent* and *discriminant* (Messick, 1980). Concurrent validity seeks to determine how well a new measure correlates with a 'gold standard' indicator of the same construct when both are completed at the same time. Predictive validity is similar; however, the new measure is completed *before* the 'gold standard' test (i.e. the researcher gauges how well scores on the former predict scores on the latter). The stronger the correlation between scores on the gold standard and scores on the new measure, the greater the CRV (Kimberlin & Winterstein, 2008). Since an instrument can only be considered a 'gold standard' when multiple strands of evidence in support of its psychometric soundness have been adduced, there will be numerous instances where such a measure is unavailable (e.g. a scale has been established for a construct that heretofore has never been tested). When this occurs, CRV may be omitted and the focus is on investigating the measure's convergent and discriminant validities.

Discriminant validity is established via the *absence* of an association between scores on a new measure and scores on other variables that are theoretically unrelated (Campbell & Fiske, 1959). In contrast, convergent validity tests whether scores on a new scale correlate – in the predicted direction – with scores on measures of other constructs that should, for theoretical and/or empirical reasons, be associated with the new scale (Campbell & Fiske, 1959). Both forms of construct validity involve testing many predictions, with each supported prediction offering another strand of evidence which attests to the measure's psychometric soundness (Carmines & Zeller, 1979). As the ultimate goal of discriminant and convergent validities is to flesh out a construct's nomological network (i.e. the theoretical system in which the construct is embedded), these validation processes are incremental and iterative.

Method

Search strategy

Academic databases (i.e. PsycINFO, PsycARTICLES and PsycTESTS[4]) were searched to identify research that used at least one measure of transgender individuals' experiences with transnegativity. A Boolean search strategy was employed and the search term was (Transgender OR Transsexual OR Agender OR Genderqueer OR Two-Spirit* OR Gender Minorit* OR Gender Nonconform*) AND (Stigma* OR Discriminat* OR Prejudic* OR Bias* OR Internal OR Opinion* OR Perception* OR Attitud* OR Belief* OR Transprejudic* OR Transnegativ* OR Transphobi* OR Negativ* OR Violence OR Harassment OR Aggress*) AND (Measur* OR Instrument* OR Scale OR Index OR Inventor* OR Questionnaire* OR Test*). A simultaneous search of all three databases yielded 2081 results. Specific publication years were not included as part of the search to maximise the number of potential hits.

Inclusionary/exclusionary criteria

Each measure had to meet a series of criteria to be retained in this systematic review. First, the measure had to appear in either a peer-reviewed journal or a doctoral dissertation. Second, the article or thesis featuring the measure(s) had to be written in English. Third, the scale(s) used

must have quantitatively measured the discrimination experiences of transgender and gender nonconforming individuals. Fourth, articles that were qualitative, theoretical in nature or did not measure discrimination experienced by transgender and gender nonconforming persons were excluded. Also excluded were articles that used generic indices of experienced discrimination (i.e. non-transgender and/or gender minority specific) or did not separate gender identity from sexual orientation (i.e. measures querying whether respondents' discrimination experiences were due to their sexual orientation *and/or* gender identity).

Article selection

After 160 duplicate articles were removed, 1921 entries remained. The flow chart, appearing in Figure 1, details the various criteria that resulted in the elimination of 1847 articles, resulting in a final sample of 74 papers, which contained 116 measures.

Review procedure

Each of the 116 measures was examined to gauge adherence to best practice guidelines regarding CV, factorial validity, reliability and validity (criterion-related and construct). See Table 1 for a summary of the psychometric properties of the measures reviewed. If a source article articulated the use of best practices when assessing a particular psychometric feature, a check mark (✓) was provided. When a measure's source article provided insufficient evidence regarding a psychometric property or provided details that do not adhere to best practices, an 'X' was entered. A question mark (?) was issued when: (1) the source article did not directly test (or, in some cases, mention) a specific psychometric element but supportive evidence could be deduced from the findings or (2) use of best practices could not be determined due to the provision of insufficient details (e.g. the authors noted evidence of construct validity but did not mention the way in which it was tested). Finally, whenever a psychometric element was irrelevant to a particular measure (e.g. an assessment of factor structure is not required for scales consisting of fewer than three items), 'N/A' (not applicable) was entered.

As scale development and refinement is an incremental process, it was anticipated that researchers would further contribute to the psychometric properties of a given measure. Therefore, on occasion, a '✓' or '?' was issued when a measure's psychometric features had been previously evaluated by other researchers. For example, if a Canadian researcher decided to use a measure that was developed for use with a Canadian sample and the CRV of this measure had been tested, it was not expected that this researcher would further assess the instrument's CRV. However, if an Italian researcher decided to use a translated version of the same scale, CRV should be investigated as the measure is being employed in a novel cultural context. Finally, construct validity did not need to be directly assessed as details about this type of validation were often embedded within an article's hypotheses. For example, if, as predicted, a researcher found an inverse association between scores on a measure of experienced transnegativity and scores on another variable (e.g. ability to pass as one's chosen gender), then this finding would offer one strand of support for the construct validity of the transnegativity scale.

Results

All five of the aforementioned psychometric properties (i.e. CV, factor structure, scale score reliability and construct validity) are reviewed for the 116 scales included in our assessment. Illustrative examples are provided to elucidate important points.

Content validity

Eight of the 116 (6.9%) scales provided sufficient evidence in support of their content validity. For example, to develop the 21-item *Experiences of Transgender Discrimination* Scale (ETDS), Poteat (2012;

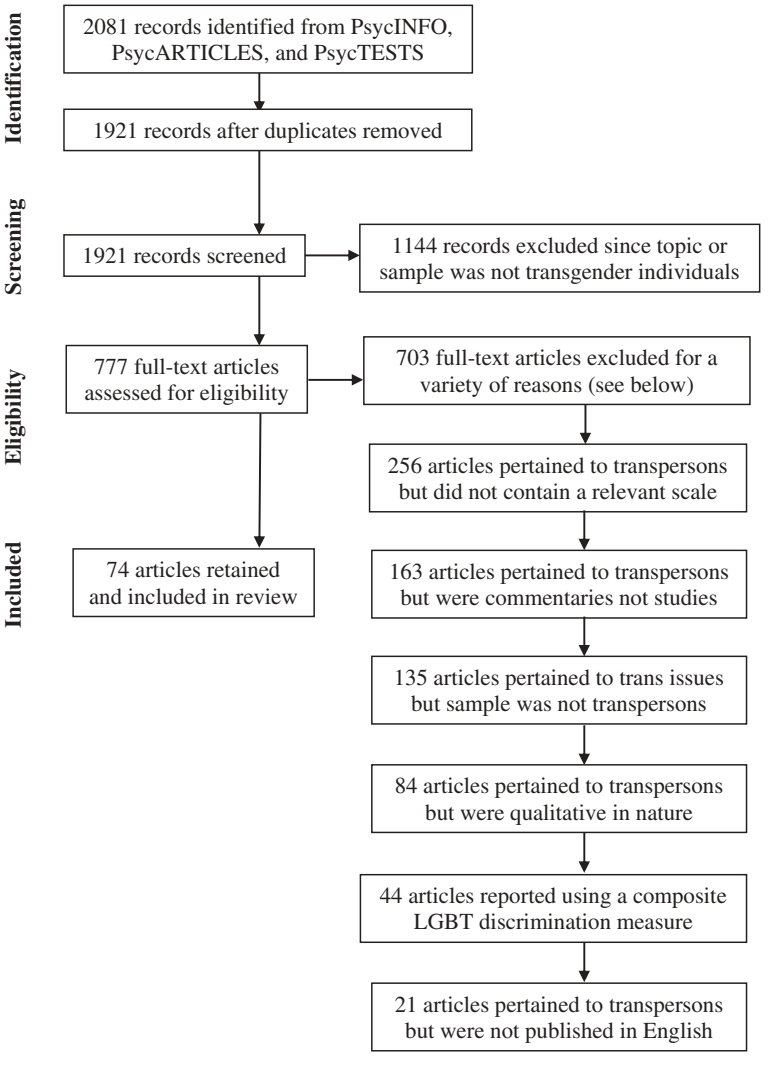


Figure 1. Flow chart describing literature search and exclusionary criteria.

see Table 1, entry 1) conducted an in-depth literature review on discrimination against transgender individuals. This researcher also developed the initial 33 items based on data from the National Transgender Discrimination Survey's (NTDS; n = 5949) original 58 items, which were already content valid since the NTDS is overseen by content experts and actively seeks input and guidance from relevant stakeholders in the development of their materials.

One (0.9%) of the measures received a question mark since insufficient details regarding the development of the scales were provided. To illustrate: in an attempt to measure relationship stigma experienced when one of the partners is a transgender individual, Gamarel, Reisner, Laurenceau, Nemoto, and Operario (2014; see Table 1, entry 20) developed nine items based on the results of focus groups with transgender women and their male partners. Given the previous research conducted by some of the authors, it is reasonable to assume that they may be classified as content experts. The authors also mentioned that both partners (i.e. transgender women and male partner) completed the same items; however, it remains unclear if this statement is referring to initial item testing or the populations this measure was developed for

Table 1. Psychometric properties of the reviewed measures.

#	Measure	CV	FS	R	CRV	CTV	TS
1	**Experiences of Transgender Discrimination Scale**	✓	✓	✓	**X**	✓	**4**
2	**Discrimination subscale (due to being TGNC)**	✓	✓	✓	**X**	✓	**4**
3	**Rejection subscale (due to being TGNC)**	✓	✓	✓	**X**	✓	**4**
4	**Victimisation subscale (due to being TGNC)**	✓	✓	✓	**X**	✓	**4**
5	Discrimination subscale (due to being TGNC)	✓	✓	✓	X	✓	4
6	Rejection subscale (due to being TGNC)	✓	✓	✓	X	✓	4
7	Victimisation subscale (due to being TGNC)	✓	✓	✓	X	✓	4
8	Adapted Workplace Discrimination Experiences scale	X	✓	✓	X	✓	3
9	Experienced transphobic events during previous year	X	X	✓	X	✓	2
10	Experienced transphobic events during lifetime	X	X	✓	X	✓	2
11	How stressful these transphobic events were	X	X	✓	X	✓	2
12	Gender Expression subscale	✓	X	✓	X	?	2
13	Experiences of Transphobia Scale	X	X	✓	X	?	1
14	Modified Turkish version of the Perceived Discrimination Scale	X	X	✓	X	X	1
15	Discrimination experienced due to GI	X	X	✓	X	X	1
16	Enacted stigma due to GI	X	X	✓	X	?	1
17	Adapted Heterosexist Harassment Rejection and Discrimination Scale	X	X	✓	X	?	1
18	Transgender Perception of Loss Scale	X	X	✓	X	?	1
19	Adapted Everyday Discrimination Scale	X	X	✓	X	?	1
20	Relationship stigma due to one partner being MtF	?	X	✓	X	?	1
21	Adapted version of the Schedule for Heterosexist Events	X	X	✓	X	?	1
22	Adapted personalised stigma subscale	X	X	✓	X	?	1
23	Experiences with Transphobia Scale	X	X	✓	X	?	1
24	Military enacted stigma due to TGI	X	X	✓	X	?	1
25	Measure of discrimination (could specify if experiences were due to GIP)	X	X	✓	X	?	1
26	Discrimination subscale from Transgender Stigma Scale	X	X	✓	X	?	1
27	(Experienced) Transphobia Scale	X	X	✓	X	?	1
28	(Experienced) Transphobia Scale	X	X	✓	X	X	1
29	Adapted Everyday Discrimination Scale	X	X	✓	X	?	1
30	Adapted Everyday Discrimination Scale	X	X	✓	X	?	1
31	(Experienced) Transphobia Scale	X	X	✓	X	?	1
32	Perceived stigma due to TGI	X	X	✓	X	X	1
33	Adapted Everyday Discrimination Scale	X	X	✓	X	?	1
34	Everyday Discrimination Scale (could specify if discrimination due to TGI)	X	X	✓	X	?	1
35	(Experienced) Transphobia Scale	X	X	✓	X	?	1
36	(Experienced) Transphobia Scale	X	X	✓	X	?	1
37	Experienced discrimination due to being GV	X	X	✓	X	?	1
38	Experiences of Transphobia Scale	X	X	✓	X	X	1
39	Victimisation by police officers	X	X	✓	X	?	1
40	(Experienced) Transphobia Scale	X	X	✓	X	?	1
41	Modified Discrimination/Harassment subscale	X	X	✓	X	?	1
42	Modified Victimisation subscale	X	X	✓	X	?	1
43	Adapted Stigma Scale	X	X	✓	X	?	1
44	Modified Heterosexist Harassment Rejection and Discrimination Scale	X	X	✓	X	?	1
45	Victimisation due to GIP	X	X	✓	X	X	1
46	Stigma Scale (for TGNC persons)	X	X	✓	X	?	1
47	Family rejection due to TGI	X	X	✓	X	?	1
48	Adapted Everyday Discrimination Scale	X	X	✓	X	?	1
49	Experiences of discrimination or harassment	X	N/A	N/A	X	X	0
50	Experiences of Transphobia Scale	X	X	X	X	X	0
51	Experienced discrimination, harassment or violence due to their TGI	X	N/A	N/A	X	X	0
52	Problems obtaining health or medical services due to TGI	X	N/A	N/A	X	X	0
53	Problems securing housing due to TGI	X	N/A	N/A	X	X	0
54	Experienced harassment, violence and economic discrimination	X	X	X	X	?	0
55	Modified measure of victimisation due to TGI	X	N/A	N/A	X	?	0
56	Measure of intimate partner violence due to TGI	X	N/A	N/A	X	?	0
57	Measuring discrimination due to GIP	X	N/A	N/A	X	?	0
58	Measure of verbal victimisation due to GIP	X	N/A	N/A	X	?	0
59	Measure of physical victimisation due to GIP	X	N/A	N/A	X	?	0
60	Perceptions of the Aversiveness of Discrimination Scale	?	N/A	X	X	X	0
61	Frequency of transnegative microaggressions by friends across 9 domains	X	X	X	X	X	0
62	Severity of transnegative microaggressions by friends across 9 domains	X	X	X	X	X	0

(*Continued*)

Table 1. (Continued).

#	Measure	CV	FS	R	CRV	CTV	TS
63	GI-based victimisation in school	X	N/A	N/A	X	?	0
64	Impact of GI-based victimisation in school	X	N/A	N/A	X	X	0
65	Whether gender-affirming health could be accessed	X	N/A	N/A	X	X	0
66	Verbal harassment at school over past year due to GE	X	N/A	N/A	X	?	0
67	Physical harassment at school over past year due to GE	X	N/A	N/A	X	?	0
68	Physical assault in school over past year due to GE	X	N/A	N/A	X	?	0
69	Experienced discrimination in healthcare due to GIP	X	N/A	N/A	X	X	0
70	Verbal harassment due to GE	X	N/A	N/A	X	?	0
71	Physical harassment due to GE	X	N/A	N/A	X	?	0
72	Physical assault due to GE	X	N/A	N/A	X	?	0
73	Housing discrimination due to TGI	X	N/A	N/A	X	?	0
74	Employment discrimination due to TGI	X	N/A	N/A	X	?	0
75	Experienced violence or harassment due to TGI	X	N/A	N/A	X	?	0
76	Experienced economic discrimination due to TGI	X	N/A	N/A	X	?	0
77	9 forms of transphobic victimisation experienced during the last thirty days, previous year, or lifetime	X	N/A	N/A	X	X	0
78	Healthcare discrimination due to TGI	X	N/A	N/A	X	?	0
79	Exposure to violence from police officers	X	N/A	N/A	X	?	0
80	Experiences of major discrimination due to TGI	X	X	X	X	?	0
81	Everyday discrimination due to TGI	X	X	X	X	?	0
82	Verbal abuse or harassment due to GIP	X	N/A	N/A	X	?	0
83	Physical abuse or violence due to GIP	X	N/A	N/A	X	?	0
84	Experienced verbal and/or physical abuse due to GIP	X	N/A	X	X	?	0
85	Measure of anti-transwomen discrimination	X	N/A	N/A	X	?	0
86	Measure of discrimination and victimisation following gender confirming surgery	X	X	X	X	?	0
87	Lifetime victimisation due to TGI	X	X	X	X	?	0
88	Recent victimisation due to TGI	X	X	X	X	?	0
89	Experienced healthcare discrimination due to TGNCI	X	N/A	N/A	X	?	0
90	Experienced transphobic discrimination in public spaces	X	X	X	X	?	0
91	Enacted stigma when seeking healthcare	X	N/A	N/A	X	?	0
92	Discrimination in healthcare, housing and employment due to GIP	X	X	X	X	?	0
93	Measure of gender-related discrimination	X	X	X	X	?	0
94	Denial of equal treatment or service in a domestic violence programme due to being TGNC	X	N/A	N/A	X	?	0
95	Denial of equal treatment or service in a rape crisis centre due to being TGNC	X	N/A	N/A	X	?	0
96	Healthcare discrimination due to participants' TGI	X	N/A	N/A	X	?	0
97	Perceived discrimination from healthcare workers	X	N/A	N/A	X	?	0
98	Perceived discrimination from other patients	X	N/A	N/A	X	?	0
99	Been attacked since age 13 (could specify if due to GE)	X	N/A	N/A	X	?	0
100	Experienced sexual violence since age 13 (could specify if due to GE)	X	N/A	N/A	X	?	0
101	Adapted Everyday Discrimination Scale (can indicate if was due to GI)	X	X	X	X	?	0
102	Experiences with transphobia	X	X	X	X	X	0
103	Victimisation due to perceived TGI	X	X	X	X	X	0
104	Experiences of Transphobia Scale	X	X	X	X	X	0
105	Discrimination due to TGI	X	X	X	X	?	0
106	Discrimination due to GIP	X	N/A	N/A	X	X	0
107	Parents' perceptions of respondents' GI	X	N/A	N/A	X	X	0
108	Discrimination from family due to TGI	X	N/A	N/A	X	X	0
109	Discrimination from friends due to TGI	X	N/A	N/A	X	?	0
110	Societal discrimination due to TGI	X	N/A	N/A	X	X	0
111	Discrimination from family due to TGI	X	N/A	N/A	X	X	0
112	Discrimination from friends due to TGI	X	N/A	N/A	X	?	0
113	Societal discrimination due to TGI	X	N/A	N/A	X	X	0

(*Continued*)

Table 1. (Continued).

#	Measure	CV	FS	R	CRV	CTV	TS
114	Discrimination from family due to TGI	X	N/A	N/A	X	X	0
115	Discrimination from friends due to TGI	X	N/A	N/A	X	?	0
116	Societal discrimination due to TGI	X	N/A	N/A	X	X	0

Note: Numbers appearing in square brackets at the end of the listings in the Reference section correspond to entries in the table. To illustrate, the first entry in the Reference section is for Arcelus et al. (2016). The number 13 appears in a square bracket at the end of this reference, which corresponds to item 13 in Table 1 (*Experiences of Transphobia Scale*). CV = content validity; FS = factor structure; R = reliability; CRV = criterion-related validity; CTV = construct validity; TS = total score based on how many psychometric properties a measure was evaluated 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 in bold received the highest overall rating; TGI = transgender identity; GIP = gender identity or presentation; GI = gender identity; TGNC = transgender or gender-nonconforming; TGNCI = transgender or gender-nonconforming identity; GE = gender expression; GV = gender variant; MTF = male-to-female; FTM = female-to-male; there are two different versions of the *(Experienced) Transphobia Scale* and the *Experiences of Transphobia Scale* in use based on two different measures being cited for each.

in general. The paper this measure appears in also clearly demonstrates that a literature review of the critical domains of this construct was completed. However, given the lack of specificity of the aforementioned steps in assessing CV, it remains unclear whether this particular measure is or is not content valid.

Finally, 107 (92.2%) scales received an 'X' since insufficient details regarding their CV were provided by their respective authors. Echoing the findings of Morrison et al. (2016) and Morrison et al. (2017a), it was common for the development of instruments to be described using imprecise and/or vague language. The following examples illustrate this point: 'The domains included in the scale were, in part, based on…' (Erich, Tittsworth, Colton Meier, & Lerman, 2010, p. 300); '…which we adapted from a scale originally developed by…' (Başar, Öz, & Karakaya, 2016, p. 1135); '…adapted it to reflect the language and experiences…' (Brennan et al., 2012, p. 1753); 'items were reworded to be more germane to…' (Brewster, Velez, DeBlaere, & Moradi, 2012, p. 63); and 'adapted slightly to assess discriminatory events' (Operario, Nemoto, Iwamoto, & Moore, 2011, p. 676).

It also was common practice for a modified scale to be used repeatedly by different groups of researchers. For example, the (*Experienced) Transphobia Scale (ETS)* was a verbatim transplant of Diaz, Ayala, Bein, Henne, and Marin (2001) 11-item *Homophobia Scale* (i.e. the only difference between the two measures was that while the latter focused on gay men, the former targeted transgender persons). From a content validation standpoint, this practice is ill-advised because it assumes that the discriminatory experiences of these minoritised groups are identical. Despite this limitation, the ETS has appeared in numerous studies: Nemoto, Bödeker, and Iwamoto (2011), Nemoto, Cruz, Iwamoto, and Sakata (2015), Pyne, Bauer, and Bradley (2015), Rotondi et al. (2011a), Rotondi et al. (2011b), Sugano, Nemoto, and Operario (2006) and Wilson et al. (2015).

It should also be noted that some researchers did not mention CV at all (e.g. Socías et al., 2014; Yadegarfard, Meinhold-Bergmann, & Ho, 2014).

Factor structure

Assessments of dimensionality were relevant to 65 (56%) measures that consisted of at least 3 items. The remaining 51 (44%) instruments did not require tests of dimensionality because 42 contained fewer than 3 items and 9 did not involve the computation of total scale scores. For example, Beemyn and Rankin (2011; see Table 1, entry 49) used seven items that measured experiences of discrimination or harassment due to one's transgender or gender nonconforming identity. Each of the seven items was examined independently as they looked at different kinds of discriminatory experiences. Similarly, Clements-Noelle et al. (2006; see Table 1, entry 57) included

four items that assessed discrimination due to participants' gender identity/presentation but analysed these items separately with their outcome variables (e.g. attempted suicide).

Of the 65 measures for whom EFA was relevant, only 8 (12.3%) followed best practice guidelines. The remainder either did not assess the dimensionality of their measure (e.g. Arcelus, Claes, Witcomb, Marshall, & Bouman, 2016; Breslow et al., 2015; Lombardi, 2009; Miller & Grollman, 2015; Nemoto et al., 2015) or did not adhere to best practice recommendations. As an illustration of the latter, Jefferson, Neilands, and Sevelius (2013) investigated the factor structure of a 13-item 'transphobic events measure' (p. 125) using principal axis factor (PAF) analysis and promax rotation. Both of these choices reflect sound decision-making (i.e. PAF is advised when data are not normally distributed and it is reasonable to use oblique rotation as it permits factors to be intercorrelated). The authors noted that 'three factors were extracted; however, only one was retained per examination of the scree plot, eigenvalues, and the interpretability of rotated factor loadings' (p. 125). The fact that Jefferson et al. did not rely solely on the eigenvalue greater than one 'rule' reflects best practice in EFA. (Of particular note is their decision to use interpretability to assist with factor retention.) Unfortunately, the authors do not report the size of the three eigenvalues, the range of the factor loadings and the presence or absence of double-loadings. They do not also indicate whether a forced one-factor solution was subsequently tested and found to be suitable for all items. In the absence of doing so, it remains unclear if each of the 13 items should have been retained (i.e. did one or more items load strongly on the second or third factors but weakly on the first one?). Gamarel et al. (2014) created a measure of relationship stigma (i.e. the negativity directed at couples when one person occupies a stigmatised social identity [in this case, transgender women with their male cisgender partners]). The authors used PCA with varimax rotation to 'examine the underlying factor [*sic*] structure of the nine items that were generated' (p. 440). Neither choice is recommended. In addition, the authors used the eigenvalue greater than one 'rule' as evidence in support of their decision to treat this measure as having a single 'factor' (read: component). It would have been advisable for the authors to use EFA, with factor extraction being determined by parallel analysis in conjunction with the scree plot and visual inspection of interpretability. Finally, it should also be noted that the authors' decision to conduct a CFA of this instrument using the *same* group of participants is ill-advised.

Scale score reliability

A majority (74; 63.8%) of the measures that were assessed contained two or more items. Of these, 63 (85.1%) computed total scale scores for one or more factors of experienced transnegativity. Nineteen (30.2%) of these 63 scales did not furnish any evidence of scale score reliability (e.g. Bouman et al., 2016; Boza & Perry, 2014; Miller & Grollman, 2015; Prunas et al., 2016).

Cronbach's alpha was calculated for 44 measures (69.8%). However, in certain cases, the rationale for furnishing this indicator of scale score reliability was not clear. To illustrate: Erich et al. (2010), authors of the 16-domain instrument entitled *Perceptions of the Aversiveness of Discrimination Scale* (PADS), reported a Cronbach's alpha coefficient of .97. The authors subsequently treated each domain as a separate indicator of social life relevant to transgender people's experiences of discrimination; thus, there is little need to provide an estimate of internal consistency. (The PADS was issued an 'X' since the computation of an alpha coefficient as evidence of its reliability was inappropriate.)

Brennan et al.'s (2012) 10-item modified measure of victimisation due to one's identity as transgender (see Table 1, entry 55) warrants special attention since, ordinarily, scale score reliability should be calculated for such measures. However, the authors did not calculate a total score per se, but rather recoded the total scores depending on whether a participant's score did or did not exceed a certain threshold (i.e. participants were issued a score of 0 if their scale score indicated they had not experienced any victimisation, or 1 if they had reported being victimised one or more times).

Validity

Criterion-Related Validity. None of the 116 scales provided evidence of CRV. This omission is (likely) due to the absence of a 'gold standard' measure of experienced transnegativity by transgender persons.

Construct Validity. One scale (0.9%) explicitly tested for evidence of construct validity; specifically, Poteat (2012; see Table 1, entry 1) assessed the convergent validity of the newly developed 21-item *Experiences of Transgender Discrimination* measure by examining the correlations between scores on this scale and scores on other theoretically related instruments. The convergent validity for the three scales (2.6%) developed by Lombardi (2009; see Table 1, entries 75–77) was examined; however, the *type* of validity was never explicitly stated anywhere in the article. Finally, for seven additional scales (6.0%), the authors appear to have conflated criterion-related and construct validity. For example, Brewster et al. (2012; see Table 1, entry 8) asserted that evidence of CRV was provided for their adapted 21-item *Workplace Discrimination Experiences Scale* via correlations in the expected direction with theoretically related constructs (i.e. job satisfaction and outness). Similarly, Testa, Habarth, Peta, Balsam, and Bockting (2015; see Table 1, entries 2–4) examined the correlations between their three scales measuring different forms of enacted stigma experienced by transgender individuals and other related yet distinct constructs (e.g. depression, mental health outcomes). In both cases, convergent validity, which falls under the broader rubric of construct validation, was being tested. Even though 10 of the 11 scales were incorrectly labelled, each was issued a check mark.

Seventy-six (65.5%) of the scales reviewed were given a question mark since indirect evidence of construct validity was available (i.e. construct validity, most often of the convergent variety, could be adduced based on the findings obtained). For example, Gleason et al. (2016; see Table 1, entry 21) observed that scores on their 17-item adapted version of the *Schedule for Heterosexist Events* correlated positively with a measure of perceived community stigma. Similarly, in a study examining risk factors for non-suicidal self-injury among transgender youth, Arcelus et al. (2016; see Table 1, entry 13) reported that lifetime presence of self-injury among participants was associated with having experienced more transnegativity. Seldom tested were known-groups validity (i.e. predicting that two or more groups will differ on the scale of interest) and discriminant validity (i.e. predicting, on the basis of theory and/or prior research, that scores on the scale of interest and scores on another measure will be negligibly associated with each other).

Finally, 29 (25%) of the scales received an 'X' since no evidence of construct validity was furnished in any capacity.

Conclusion

The findings of this systematic review revealed that the majority of the scales designed to measure experiences of transnegativity for transgender individuals did not adhere to best practice recommendations in terms of measurement construction and testing. Coinciding with the results of Morrison et al. (2016, 2017b), assessments of dimensionality and validity (criterion-related and construct) were largely absent. A majority of the instruments reviewed offered some evidence of scale score reliability, typically in the form of Cronbach's alpha. However, the use of this coefficient is problematic as it requires assumptions rarely met by social scientific data (Dunn, Baguley, & Brunsden, 2014). None of the measures examined used lesser known but recommended alternatives, such as the Greatest Lower Bound estimate or Omega (e.g. Dunn et al., 2014; Peters, 2014).

Overall, 116 measures across 74 articles and/or papers were retrieved and subsequently evaluated on the basis of 5 psychometric properties: (1) CV; (2) factor structure (i.e. dimensionality); (3) scale score reliability; (4) CRV; and (5) construct validity. A 'perfect' score of five points would occur if check marks (✓) were issued for all of the aforementioned properties. However, given the absence of a clearly defined 'gold standard' indicator of transnegativity, this element was removed from the scoring

protocol. Assuming that four points denotes an optimal measure, inspection of Table 1 reveals that one scale and three subscales received this total (i.e. four out of four). These measures are the 21-item ETDS (Poteat, 2012; see Table 1, entry 1), the 5-item discrimination subscale, 6-item rejection subscale and 6-item victimisation subscale taken from the Gender Minority Stress and Resilience Model (Testa et al., 2015, 2017; see Table 1, entries 2–4[5]). On the basis of our review, any of these four measures would serve as a suitable 'gold standard' indicator for the purposes of testing the CRV of subsequent transnegativity instruments. However, an important caveat is that these measures were developed and tested using English-speaking American participants; thus, their suitability for use with persons in other cultural contexts remains unclear and presents an avenue for future research. It should also be noted that our study did not include the search term 'non-binary'. Despite utilising seven different search terms to locate trans-directed measures of discrimination, this term was overlooked in the present study. Researchers interested in furthering the work on trans-directed measures of discrimination should consider including this term in their search strategy. Finally, the present study included the term 'two-spirit'. This is a distinct ethnocultural term that refers to people of Indigenous descent who are located in a specific geographic region (now referred to as Canada). It is recognised that other terms used to refer to additional ethnocultural groups (e.g. the non-western kathoey and the fa'a'fafine) should become a focus in future search strategies.

In closing, the results of our systematic review indicate that a majority of the measures used to assess transgender and gender nonconforming persons' experiences of discrimination are suboptimal. Many do not follow best practice guidelines in terms of item development, refinement and validation resulting in a surfeit of instruments that not only possess questionable psychometric integrity but, given their ad hoc nature, prevent researchers from formulating a coherent picture of transnegativity in contemporary societies. We recommend that the four measures highlighted in Table 1 be used with the understanding that their cultural suitability be subject to rigorous testing. We also recommend that those wishing to create new indicators of discrimination against transgender and gender nonconforming individuals follow the best practice guidelines articulated herein.

Notes

1. Rotondi et al.'s (2011a) measure of transnegativity examined 'the frequency of direct and indirect transphobic experiences, including discrimination and harassment, exposure to ideas of non-normalcy, familial embarrassment, or being fetishized' (p. 139).
2. For a brief overview of Classical Test Theory (CTT), please consult De Champlain (2010).
3. Best practice recommendations for the use and reporting of CFA also have been elucidated (e.g. Jackson, Gillaspy Jr, & Purc-Stephenson, 2009).
4. We, subsequently, conducted a search using the Education Resources Information Centre (ERIC). No additional measures were found.
5. Entries five through seven in Table 1 are identical to entries two through four; however, different predictions were tested using different samples.

Notes on contributors

Melanie A. Morrison is a Professor in Social Psychology at the University of Saskatchewan. Melanie investigates gender and anti-LGBTQ+ prejudice and discrimination, particularly covert forms, and uses feminist principles to inform her research practice.

C. J. Bishop recently obtained his PhD in Social Psychology at the University of Saskatchewan under the supervision of Dr. Todd Morrison. CJ conducts research on sexual and gender minority groups, and is currently a research associate at the University of Alberta.

Todd G. Morrison is a Professor in Social Psychology at the University of Saskatchewan. Todd conducts research in the areas of sexual and gender minority psychology, gay pornography, body image, and feminism.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by an Insight Grant awarded from the Social Sciences and Humanities Research Council (SSHRC) of Canada (435-2016-1485).

References

Arcelus, J., Claes, L., Witcomb, G. L., Marshall, E., & Bouman, W. P. (2016). Risk factors for non-suicidal self-injury among trans youth. *The Journal of Sexual Medicine*, *13*(3), 402–412. [13].

Başar, K., Öz, G., & Karakaya, J. (2016). Perceived discrimination, social support, and quality of life in gender dysphoria. *The Journal of Sexual Medicine*, *13*(7), 1133–1141. [14].

Bauer, G., Pyne, J., Francino, M., & Hammond, R. (2013). Suicidality among trans people in Ontario: Implications for social work and social justice/La suicidabilité parmi les personnes trans en Ontario: Implications en travail social et en justice sociale. *Service Social*, *59*(1), 35–62.

Bauer, G, & Scheim, A. I. (2015). *Transgender people in Ontario*. Canada: Trans PULSE.

Beemyn, G., & Rankin, S. (2011). *The lives of transgender people*. New York, NY: Columbia University Press. [49].

Benotsch, E. G., Zimmerman, R. S., Cathers, L., Heck, T., McNulty, S., Pierce, J., … Snipes, D. J. (2016). Use of the internet to meet sexual partners, sexual risk behavior, and mental health in transgender adults. *Archives of Sexual Behavior*, *45*(3), 597–605. [15].

Billard, T. J. (2018). The crisis in content validity among existing measures of transphobia. *Archives of Sexual Behavior*, *47*, 1305–1306.

Bockting, W. O., Miner, M. H., Swinburne Romine, R. E., Hamilton, A., & Coleman, E. (2013). Stigma, mental health, and resilience in an online sample of the US transgender population. *American Journal of Public Health*, *103*(5), 943–951. [16].

Bouman, W. P., Claes, L., Marshall, E., Pinner, G. T., Longworth, J., Maddox, V., … Arcelus, J. (2016). Sociodemographic variables, clinical features, and the role of preassessment cross-sex hormones in older trans people. *The Journal of Sexual Medicine*, *13*(4), 711–719. [50].

Boza, C., & Perry, K. N. (2014). Gender-related victimization, perceived social support, and predictors of depression among transgender Australians. *International Journal of Transgenderism*, *15*(1), 35–52. [51, 52, 53, 54].

Brennan, J., Kuhns, L. M., Johnson, A. K., Belzer, M., Wilson, E. C., & Garofalo, R. (2012). Syndemic theory and HIV-related risk among young transgender women: The role of multiple, co-occurring health problems and social marginalization. *American Journal of Public Health*, *102*(9), 1751–1757. [55, 56].

Breslow, A. S., Brewster, M. E., Velez, B. L., Wong, S., Geiger, E., & Soderstrom, B. (2015). Resilience and collective action: Exploring buffers against minority stress for transgender individuals. *Psychology of Sexual Orientation and Gender Diversity*, *2*(3), 253–265. [17].

Brewster, M. E., Velez, B., DeBlaere, C., & Moradi, B. (2012). Transgender individuals' workplace experiences: The applicability of sexual minority measures and models. *Journal of Counseling Psychology*, *59*(1), 60–70. [8].

Budge, S. L., Adelson, J. L., & Howard, K. A. S. (2013). Anxiety and depression in transgender individuals: The roles of transition status, loss, social support, and coping. *Journal of Consulting and Clinical Psychology*, *81*(3), 545–557. [18].

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

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

Caruso, J. C. (2000). Reliability generalization of the NEO personality scales. *Educational and Psychological Measurement*, *60*(2), 236–254.

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.

Clements-Noelle, 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. [57, 58, 59].

Connell, R. W., & Messerschmidt, J. W. (2005). Hegemonic masculinity: Rethinking the concept. *Gender & Society*, *19*(6), 829–859.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, *60*(2), 170–180.

De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education*, *44*(1), 109–117.

Diaz, R. M., Ayala, G., Bein, E., Henne, J., & Marin, B. V. (2001). The impact of homophobia, poverty, and racism on the mental health of gay and bisexual Latino men: Findings from 3 US cities. *American Journal of Public Health*, *91*(6), 927–932.

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.

Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*, 637–642.

Erich, S., Tittsworth, J., Colton Meier, S. L., & Lerman, T. (2010). Transsexuals of color: Perceptions of discrimination based on transsexual status and race/ethnicity status. *Journal of GLBT Family Studies*, *6*(3), 294–314. [60].

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.

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.

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

Galupo, M. P., Henise, S. B., & Davis, K. S. (2014). Transgender microaggressions in the context of friendship: Patterns of experience across friends' sexual orientation and gender identity.. *Psychology of Sexual Orientation and Gender Diversity*, *1*(4), 461–470. [61, 62].

Gamarel, K. E., Reisner, S. L., Laurenceau, J.-P., Nemoto, T., & Operario, D. (2014). Gender minority stress, mental health, and relationship quality: A dyadic investigation of transgender women and their cisgender male partners. *Journal of Family Psychology*, *28*(4), 437–447. [19, 20].

Gaskin, C. J., & Happell, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *International Journal of Nursing Studies*, *51*(3), 511–521.

Gazzola, S. B., & Morrison, M. A. (2014). Cultural and personally endorsed stereotypes of transgender men and transgender women: Notable correspondence or disjunction? *International Journal of Transgenderism*, *15*, 76–99.

Gleason, H. A., Livingston, N. A., Peters, M. M., Oost, K. M., Reely, E., & Cochran, B. N. (2016). Effects of state non-discrimination laws on transgender and gender-nonconforming individuals' perceived community stigma and mental health. *Journal of Gay and Lesbian Mental Health*, *20*(4), 350–362. [21].

Goldblum, P., Testa, R. J., Pflum, S., Hendricks, M. L., Bradford, J., & Bongar, B. (2012). The relationship between gender-based victimization and suicide attempts in transgender people. *Professional Psychology: Research and Practice*, *43*(5), 468–475. [63, 64].

Golub, S. A., Walker, J. J., Longmire-Avital, B., Bimbi, D. S., & Parsons, J. T. (2010). The role of religiosity, social support, and stress-related growth in protecting against HIV risk among transgender women. *Journal of Health Psychology*, *15*(8), 1135–1144. [22].

Grant, J. M., Mottet, L. A., Tanis, J., Harrison, J., Herman, J. L., & Keisling, M. (2012,September 11). Injustice at every turn: A report of the National Transgender Discrimination Survey. Retrieved from http://www.transequality.org/sites/default/files/docs/resources/NTDS_Report.pdf

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.

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.

Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*, 6–23.

Jefferson, K., Neilands, T. B., & Sevelius, J. (2013). Transgender women of color: Discrimination and depression symptoms. *Ethnicity and Inequalities in Health and Social Care*, *6*(4), 121–136. [23].

Keuroghlian, A. S., Reisner, S. L., White, J. M., & Weiss, R. D. (2015). Substance use and treatment of substance use disorders in a community sample of transgender adults. *Drug and Alcohol Dependence*, *152*, 139–146. [65].

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.

Kosciw, J. G., Palmer, N. A., & Kull, R. M. (2015). Reflecting resiliency: Openness about sexual orientation and/or gender identity and its relationship to well-being and educational outcomes for LGBT students. *American Journal of Community Psychology*, *55*(1–2), 167–178. [66, 67, 68].

Kosenko, K., Rintamaki, L., Raney, S., & Maness, K. (2013). Transgender patient perceptions of stigma in health care contexts. *Medical Care*, *51*(9), 819–822. [69].

Kull, R. M., Greytak, E. A., Kosciw, J. G., & Villenas, C. (2016). Effectiveness of school district anti-bullying policies in improving LGBT youths' school climate. *Psychology of Sexual Orientation and Gender Diversity*, *3*(4), 407–415. [70, 71, 72].

Lehavot, K., Simpson, T. L., & Shipherd, J. C. (2016). Factors associated with suicidality among a national sample of transgender veterans. *Suicide and Life-Threatening Behavior*, *46*(5), 507–524. [24, 73, 74].

Lombardi, E. (2009). Varieties of transgender/transsexual lives and their relationship with transphobia. *Journal of Homosexuality*, *56*(8), 977–992. [9, 10, 11].

Lombardi, E. L., Wilchins, R. A., Priesing, D., & Malouf, D. (2002). Gender violence: Transgender experiences with violence and discrimination. *Journal of Homosexuality*, *42*(1), 89–101. [75, 76, 77].

Luu, T. D. (2011). *Variables impacting depressive symptoms in lesbian, gay, bisexual, and transgender Asian and Pacific Islander Americans* (Doctoral Dissertation). Retrieved from http://search.proquest.com/docview/968980674 [25]
Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382–385.
Marshall, B. D. L., Socías, M. E., Kerr, T., Zalazar, V., Sued, O., & Arístegui, I. (2016). Prevalence and correlates of lifetime suicide attempts among transgender persons in Argentina. *Journal of Homosexuality*, *63*(7), 955–967. [78, 79].
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.
Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, *35*, 1012–1027.
Meyer, I. H. (1995). Minority stress and mental health in gay men. *Journal of Health and Social Behavior*, *36*, 38–56.
Miller, L. R., & Grollman, E. A. (2015). The social costs of gender nonconformity for transgender adults: Implications for discrimination and health. *Sociological Forum*, *30*(3), 809–831. [80, 81].
Mizock, L., & Mueser, K. T. (2014). Employment, mental health, internalized stigma, and coping with transphobia among transgender individuals. *Psychology of Sexual Orientation and Gender Diversity*, *1*(2), 146–158. [26].
Morrison, M. A., Bishop, C. J., Gazzola, S. B., McCutcheon, J. M., Parker, K., & Morrison, T. G. (2017a). Systematic review of the psychometric properties of transphobia scales. *International Journal of Transgenderism*, *18*, 395–410.
Morrison, M. A., Bishop, C. J., & Morrison, T. G. (2018). A systematic review of the psychometric properties of composite LGBT prejudice and discrimination scales. *Journal of Homosexuality*. doi:10.1080/00918369.2017.1422935
Morrison, T. G., Bishop, C. J., Morrison, M. A., & Parker-Taneo, K. (2016). A psychometric review of measures assessing discrimination against sexual minorities. *Journal of Homosexuality*, *63*(8), 1086–1126.
Morrison, T. G., Morrison, M. A., & McCutcheon, J. M. (2017b). Best practice recommendations for using structural equation modelling in psychological research. *Psychology*, *8*, 1326–1341.
Nemoto, T., Bödeker, B., & Iwamoto, M. (2011). Social support, exposure to violence and transphobia, and correlates of depression among male-to-female transgender women with a history of sex work. *American Journal of Public Health*, *101*(10), 1980–1988. [27].
Nemoto, T., Cruz, T. M., Iwamoto, M., & Sakata, M. (2015). A tale of two cities: Access to care and services among African-American transgender women in Oakland and San Francisco. *LGBT Health*, *2*(3), 235–242. [28].
*Numbers appearing in square brackets at the end of the listings in the Reference section correspond to entries in the table. To illustrate, the first entry in the Reference section is for Arcelus et al. (2016). The number 13 appears in a square bracket at the end of this reference, which corresponds to item 13 in Table 1 (*Experiences of Transphobia Scale*).
Nuttbrock, L., Bockting, W., Rosenblum, A., Hwahng, S., Mason, M., Marci, M., & Becker, J. (2014). Gender abuse and major depression among transgender women: A prospective study of vulnerability and resilience. *American Journal of Public Health*, *104*(11), 2191–2198. [82, 83].
Nuttbrock, L., Bockting, W., Rosenblum, A., Hwahng, S., Mason, M., Marci, M., & Becker, J. (2015). Transgender community involvement and the psychological impact of abuse among transgender women. *Psychology of Sexual Orientation and Gender Diversity*, *2*(4), 386–390. [84].
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.
Operario, D., Nemoto, T., Iwamoto, M., & Moore, T. (2011). Unprotected sexual behavior and HIV risk in the context of primary partnerships for transgender women. *AIDS and Behavior*, *15*(3), 674–682. [29].
Operario, D., Yang, M.-F., Reisner, S. L., Iwamoto, M., & Nemoto, T. (2014). Stigma and the syndemic of HIV-related health risk behaviors in a diverse sample of transgender women. *Journal of Community Psychology*, *42*(5), 544–557. [30].
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. Retrieved from http://140.133.46.32/retrieve/33379/11930.pdf
Peters, G.-J. Y. (2014). The alpha and omega of scale reliability and validity. *The European Health Psychologist*, *16*(2), 56–69.
Pinheiro Júnior, F. M. L., Kendall, C., Martins, T. A., Mota, R. M. S., Macena, R. H. M., Glick, J., & .Kerr, L. (2016). Risk factors associated with resistance to HIV testing among transwomen in Brazil. *AIDS Care*, *28*(1), 92–97. [85].
Poteat, T. (2012). *The role of stigma and discrimination in health care utilization and HIV risk among transgender adults* (Doctoral Dissertation). Retrieved from http://search.proquest.com/docview/1221009230 [1]
Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, *2*(1), 13–43.
Prunas, A., Fisher, A. D., Bandini, E., Maggi, M., Pace, V., Todarello, O., … Bini, M. (2016). Eudaimonic well-being in transsexual people, before and after gender confirming surgery. *Journal of Happiness Studies*, Advance online publication. doi:10.1007/s10902-016-9780-7 [86].
Pyne, J., Bauer, G., & Bradley, K. (2015). Transphobia and other stressors impacting trans parents. *Journal of GLBT Family Studies*, *11*(2), 107–126. [31].
Raiford, J. L., Hall, G. J., Taylor, R. D., Bimbi, D. S., & Parsons, J. T. (2016). The role of structural barriers in risky sexual behavior, victimization and readiness to change HIV/STI-related risk behavior among transgender women. *AIDS and Behavior*, *20*(10), 2212–2221. [32, 87, 88].

Reisner, S. L., Gamarel, K. E., Nemoto, T., & Operario, D. (2014). Dyadic effects of gender minority stressors in substance use behaviors among transgender women and their non-transgender male partners. *Psychology of Sexual Orientation and Gender Diversity*, *1*(1), 63–71. [33].

Reisner, S. L., Hughto, J. M. W., Dunham, E. E., Heflin, K. J., Begenyi, J. B. G., Coffey-Esquivel, J., & Cahill, S. (2015a). Legal protections in public accommodations settings: A critical public health issue for transgender and gender-non-conforming people. *The Milbank Quarterly*, *93*(3), 484–515. [89, 90].

Reisner, S. L., Hughto, J. M. W., Gamarel, K. E., Keuroghlian, A. S., Mizock, L., & Pachankis, J. E. (2016). Discriminatory experiences associated with posttraumatic stress disorder symptoms among transgender adults. *Journal of Counseling Psychology*, *63*(5), 509–519. [34].

Reisner, S. L., Pardo, S. T., Gamarel, K. E., Hughto, J. M. W., Pardee, D. J., & Keo-Meier, C. L. (2015b). Substance use to cope with stigma in healthcare among U.S. female-to-male trans masculine adults. *LGBT Health*, *2*(4), 324–332. [91].

Riggs, D. W., & Sion, R. (2017). Gender differences in cisgender psychologists' and trainees' attitudes toward transgender people. *Psychology of Men & Masculinity*, *18*(2), 187–190.

Rood, B. A., Puckett, J. A., Pantalone, D. W., & Bradford, J. B. (2015). Predictors of suicidal ideation in a statewide sample of transgender individuals. *LGBT Health*, *2*(3), 270–275. [92].

Rotondi, N. K., Bauer, G. R., Scanlon, K., Kaay, M., Travers, R., & Travers, A. (2011a). Prevalence of and risk and protective factors for depression in female-to-male transgender Ontarians: Trans PULSE project. *Canadian Journal of Community Mental Health*, *30*(2), 135–155. [35].

Rotondi, N. K., Bauer, G. R., Travers, R., Travers, A., Scanlon, K., & Kaay, M. (2011b). Depression in male-to-female transgender Ontarians: Results from the Trans PULSE project. *Canadian Journal of Community Mental Health*, *30*(2), 113–133. [36].

Rowe, C., Santos, G.-M., McFarland, W., & Wilson, E. C. (2015). Prevalence and correlates of substance use among trans*female youth ages 16–24 years in the San Francisco Bay area. *Drug and Alcohol Dependence*, *147*, 160–166. [93].

Ruggs, E. N., Martinez, L. R., Hebl, M. R., & Law, C. L. (2015). Workplace "trans"-actions: How organizations, coworkers, and individual openness influence perceived gender identity discrimination. *Psychology of Sexual Orientation and Gender Diversity*, *2*(4), 404–412. [37].

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.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353.

Schmitt, T. A. (2011). Current methodological consideration in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304–321.

Seelman, K. L. (2015). Unequal treatment of transgender individuals in domestic violence and rape crisis programs. *Journal of Social Service Research*, *41*(3), 307–325. [94, 95].

Sevelius, J. M., Saberi, P., & Johnson, M. O. (2014). Correlates of antiretroviral adherence and viral load among transgender women living with HIV. *AIDS Care*, *26*(8), 976–982. [38].

Shires, D. A., & Jaffee, K. (2015). Factors associated with health care discrimination experiences among a national sample of female-to-male transgender individuals. *Health and Social Work*, *40*(2), 134–141. [96].

Shulman, G. P., Holt, N. R., Hope, D. A., Mocarski, R., Eyer, J., & Woodruff, N. (2017). A review of contemporary assessment tools for use with transgender and gender nonconforming adults. *Psychology of Sexual Orientation and Gender Diversity*, *4*(3), 304–313.

Socías, M. E., Marshall, B. D. L., Arístegui, I., Romero, M., Cahn, P., Kerr, T., & Sued, O. (2014). Factors associated with healthcare avoidance among transgender women in Argentina. *International Journal for Equity in Health*, *13*(1), 81–88. [39, 97, 98].

Stewart, L., O'Halloran, P., & Oates, J. (2018). Investigating the social integration and wellbeing of transgender individuals: A meta-synthesis. *International Journal of Transgenderism*, *19*(1), 46–58.

Sugano, E., Nemoto, T., & Operario, D. (2006). The impact of exposure to transphobia on HIV risk behavior in a sample of transgendered women of color in San Francisco. *AIDS and Behavior*, *10*(2), 217–225. [40].

Tebbe, E. A., & Moradi, B. (2016). Suicide risk in trans populations: An application of minority stress theory. *Journal of Counseling Psychology*, *63*(5), 520–533. [12, 41, 42].

Testa, R. J., Habarth, J., Peta, J., Balsam, K., & Bockting, W. (2015). Development of the gender minority stress and resilience measure. *Psychology of Sexual Orientation and Gender Diversity*, *2*(1), 65–77. [2, 3, 4].

Testa, R. J., Michaels, M. S., Bliss, W., Rogers, M. L., Balsam, K. F., & Joiner, T. (2017). Suicidal ideation in transgender people: Gender minority stress and interpersonal theory factors. *Journal of Abnormal Psychology*, *126*(1), 125–136. [5, 6, 7].

Testa, R. J., Sciacca, L. M., Wang, F., Hendricks, M. L., Goldblum, P., Bradford, J., & Bongar, B. (2012). Effects of violence on transgender people. *Professional Psychology: Research and Practice*, *43*(5), 452–459. [99, 100].

Veldorale-Griffin, A., & Darling, C. A. (2016). Adaptation to parental gender transition: Stress and resilience among transgender parents. *Archives of Sexual Behavior*, *45*(3), 607–617. [43].

Velez, B. L., Breslow, A. S., Brewster, M. E., Cox, R. J., & Foster, A. B. (2016). Building a pantheoretical model of dehumanization with transgender men: Integrating objectification and minority stress theories. *Journal of Counseling Psychology*, *63*(5), 497–508. [44].

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.

Williams, C. C., Curling, D., Steele, L. S., Gibson, M. F., Daley, A., Green, D. C., & Ross, L. E. (2017). Depression and discrimination in the lives of women, transgender and gender liminal people in Ontario, Canada. *Health and Social Care in the Community*, *25*(3), 1139–1150. [101].

Wilson, E. C. (2007). *Male-to-female transgender youth: Stigma, discrimination and the relationship to HIV* (Doctoral Dissertation). Retrieved from http://search.proquest.com/docview/304900693 [102, 103]

Wilson, E. C., Chen, Y.-H., Arayasirikul, S., Fisher, M., Pomart, A., Le, V., … McFarland, W. (2015). Differential HIV risk for racial/ethnic minority trans*female youths and socioeconomic disparities in housing, residential stability, and education. *American Journal of Public Health*, *105*(S3), e41–e47. [104].

Wilson, E. C., Chen, Y.-H., Arayasirikul, S., Raymond, H. F., & McFarland, W. (2016). The impact of discrimination on the mental health of trans*female youth and the protective effect of parental support. *AIDS and Behavior*, *20*(10), 2203–2211. [105].

Wilson, E. C., Garofalo, R., Harris, R. D., Herrick, A., Martinez, M., Martinez, J., & Belzer, M. (2009). Transgender female youth and sex work: HIV risk and a comparison of life factors related to engagement in sex work. *AIDS and Behavior*, *13*(5), 902–913. [45, 106].

Wilson, M. S. (2013). *Violence and mental health in the transgender community* (Doctoral Dissertation). Retrieved from: http://search.proquest.com/docview/1647175438 [46]

Yadegarfard, M., Meinhold-Bergmann, M. E., & Ho, R. (2014). Family rejection, social isolation, and loneliness as predictors of negative health outcomes (depression, suicidal ideation, and sexual risk behavior) among Thai male-to-female transgender adolescents. *Journal of LGBT Youth*, *11*(4), 347–363. [47].

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.

Yang, M.-F., Manning, D., Van Den Berg, J. J., & Operario, D. (2015a). Stigmatization and mental health in a diverse sample of transgender women. *LGBT Health*, *2*(4), 306–312. [48].

Yang, X., Wang, L., Gu, Y., Song, W., Hao, C., Zhou, J., … Zhao, Q. (2016a). A cross-sectional study of associations between casual partner, friend discrimination, social support and anxiety symptoms among Chinese transgender women. *Journal of Affective Disorders*, *203*, 22–29. [108, 109, 110].

Yang, X., Wang, L., Hao, C., Gu, Y., Song, W., Wang, J., … Zhao, Q. (2015b). Sex partnership and self-efficacy influence depression in Chinese transgender women: A cross-sectional study. *PLoS One*, *10*(9), e0136975. [111, 112, 113].

Yang, X., Zhao, L., Wang, L., Hao, C., Gu, Y., Song, W., … Wang, X. (2016b). Quality of life of transgender women from China and associated factors: A cross-sectional study. *The Journal of Sexual Medicine*, *13*(6), 977–987. [114, 115, 116].