



The Crisis in Content Validity Among Existing Measures of Transphobia

Thomas J Billard¹

Received: 11 October 2017 / Accepted: 6 March 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

In their excellent review of existing transphobia measures, Morrison et al. (2017) found that no single measure demonstrated scale reliability, factor structure, and each form of validity—content, criterion, and construct—using gold-standard psychometric practices. Rather, even the most thoroughly validated measures faltered on at least one psychometric property, while the vast majority faltered on several. While this is not entirely surprising (gold standards are rarely met, even in the most prestigious journals), it should concern our field that we do not yet have *any* transgender attitudes scale of sufficient rigor. It is therefore apropos that Morrison et al. ended their study with a call for the development of measurements according to the best practices they outlined.

While the majority of Morrison et al.'s discussion focused on statistical issues with existing measures of transphobia, such as lack of criterion-related validity and internal consistency, they also noted the dearth of content validation among measures. Drawing on Yaghmaie (as cited in Morrison et al., 2017), they suggested sufficient content validity could be established by deriving items from three sources: (1) reviews of pertinent literature, (2) input from relevant stakeholders, and (3) consultation with subject-area and psychometric experts. Against these criteria, only three of the 83 scales they reviewed demonstrated sufficient content validity. Yet, despite thoroughly assessing the extent to which measures met (or, more accurately, failed to meet) these three criteria, nowhere did Morrison et al. assess the content of the scales' actual items. Even a cursory review would have revealed that many content validity issues stem not necessarily from the methodological practices by which they were generated but rather from how the researchers conceptualized both transphobia and transgender people.

As a salient example, let us take the Genderism and Transphobia Scale (GTS; Hill & Willoughby, 2005), the most widely used

measure of transphobia. As Morrison et al. noted, the GTS was developed based on a review of literature focusing on “anti-trans sentiments and the difficulties trans persons have on a day-to-day basis” (Hill & Willoughby, 2005, p. 8), but they did not further validate the content of their scale items through input from relevant stakeholders (i.e., transgender people and/or cisgender respondents) or through consultation with psychometric experts. Thus, the development of the GTS clearly failed to meet Morrison et al.'s criteria for content validity. However, examination of the GTS's items illustrates that the methods of development are of secondary concern compared to concerns over their text content. Consider the following items from the GTS:

1. “I have beat up men who act like sissies.” Immediately, we can note that this is a behavioral item and, as Walch, Ngamake, Francisco, Stitt, and Shingler (2012) noted, measures of transphobia (like all attitudinal scales) are most effective when they do not contain a mix of attitudinal and behavioral items. More troublingly, however, the necessary association of transphobia with acts of physical violence disregards the common shades of ambivalent prejudice that, in a more robust conceptualization of transphobia, we might rather focus on.
2. “If a friend wanted to have his penis removed in order to become a woman, I would openly support him.” This item, even more concerning than the last, constitutes a gross misconstrual of transgender identity. Not only does it misgender the hypothetical subject to whom it refers, it presents an erroneous characterization of genital reconstruction surgery and reifies the incorrect assumption transgender people require surgical intervention to realize their gender identities.
3. “Men who cross-dress for sexual pleasure disgust me.” This final example further illustrates how the GTS fails to appropriately conceptualize transgender people, toward whom it purports to measure attitudes. By including items measuring attitudes toward cross-dressers, the scale conflates distinct identities (in this case, transgender women and male cross-dressers) that should be measured separately. “Transgender” is a broad umbrella category, of

✉ Thomas J Billard
tbillard@usc.edu

¹ Annenberg School for Communication and Journalism,
University of Southern California, Los Angeles, CA 90089,
USA

course, but one cannot reasonably hope to measure attitudes toward all identities under that umbrella in one scale.

It should, therefore, be clear that even independent of Morrison et al.'s three criteria for content validation, existing measures of transphobia lack substantively valid content. Even in instances where best practices are used, such as Kanamori, Cornelius-White, Pegors, Daniel, and Hulgus's (2017) use of both literature review and expert consultation, content validity issues arise. Specifically, Kanamori et al. consulted a Christian theology expert and deliberately oversampled Evangelical Christians in the development phase of scale development, which they justified as important to tap "religious nuances" in the American public's attitudes. While technically meeting Morrison et al.'s criteria for expert consultation, Kanamori et al.'s choice of expert renders their scale unrepresentative of general population attitudes and overrepresentative of right-wing Christian beliefs about transgender identity.

The point of the content validity criteria laid out by Morrison et al. is to ensure that scale items accurately represent the construct of interest. In theory, these practices protect against the inclusion of items, like those mentioned above, that misrepresent the construct or elide significant conceptual differences. And, indeed, it is unfortunate that existing transphobia scales have not employed these practices. However, the problem with existing scales is not their methodological shortcomings, but their conceptual shortcomings. While Morrison et al. pointed out content validation issues and recommended the development of new measures because of them, the bulk of their analysis focused on the statistical rigor of these measures; content validation issues were not assessed textually but in terms of checking methodological boxes.

What might be suggested by Morrison et al.'s review of existing transphobia scales is that we should offer further methodological validation of the scales we already have. However,

I would argue that, due to the particular crisis in content validity, this is not the answer. 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. As such, what is needed is not the further statistical validation of existing measures, but the generation of entirely new measures developed on solid foundations of qualitative data assessing both the breadth and depth of current public attitudes (as I have done in the development and validation of the Attitudes Toward Transgender Men and Women scale [ATTMW]; Billard, 2018). And while Morrison et al.'s content validity criteria may provide methodological resources to lead us to greater content validity, these suggestions are not boxes to tick to say we have validity, not even necessarily the best or only route to valid scale items. But one thing is certain: Among all the issues Morrison et al. noted in their review, none is so grave as the crisis in content validity.

References

- Billard, T. J. (2018). Attitudes Toward Transgender Men and Women: Development and validation of a new measure. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2018.00387>.
- Hill, D. B., & Willoughby, B. L. B. (2005). The development and validation of the Genderism and Transphobia Scale. *Sex Roles*, *53*, 531–544. <https://doi.org/10.1007/s11199-005-7140-x>.
- Kanamori, Y., Cornelius-White, J. H. D., Pegors, T. K., Daniel, T., & Hulgus, J. (2017). Development and validation of the Transgender Attitudes and Beliefs Scale. *Archives of Sexual Behavior*, *46*, 1503–1515. <https://doi.org/10.1007/s10508016-0840-1>.
- Morrison, M. A., Bishop, C. J., Gazzola, S. B., McCutcheon, J. M., Paker, K., & Morrison, T. G. (2017). Systematic review of the psychometric properties of transphobia scales. *International Journal of Transgenderism*, *18*, 395–410. <https://doi.org/10.1080/15532739.1332535>.
- Walch, S. E., Ngamake, S. T., Francisco, J., Stitt, R. L., & Shingler, K. A. (2012). The Attitudes Toward Transgendered Individuals Scale: Psychometric properties. *Archives of Sexual Behavior*, *41*, 1283–1291. <https://doi.org/10.1007/s10508-012-9995-6>.