



COLLABORATIVE OF
**GENDER & SEXUAL
HEALTH EQUITY**

RESEARCH EDUCATION RESOURCES

Gender & Sex in Methods & Measurement

Research Equity Toolkit

*Tool #7: Developing & Working with
Validated Scales*



THE UNIVERSITY
OF BRITISH COLUMBIA

Introduction

A validated scale is a structured instrument or set of questions used in research to collect data or information on a specific topic. A *scale* is a single measure, or a series of measures, that respondents answer to indicate the level or intensity of the construct being assessed. For example, a single-item scale might ask someone to rate their overall pain from one to ten, whereas a multi-item pain scale might ask someone about worst, least, and usual pain, current pain, and self-assessments of pain across time. In various health-related fields, these scales might also be called *inventories*, *indexes*, *screening tools*, *diagnostic tools*, or *assessment tools*. For the purposes of this resource, we will use the term scales. No matter their name, these are often measures that are used alone, or in combination, to assess the presence and severity of specific symptoms or traits associated with a particular condition or construct. There are also scales to measure attitudes, beliefs, and values; assess personality traits, motivating factors, individual abilities, preferences, and behaviours; and quantify well-being, and satisfaction.

When a scale is referred to as *validated*, this means that it has undergone a systematic process to ensure that it is reliable (that it produces consistent and stable results) and valid (that it measures what it is designed to measure). In this tool, we use “validation study” to specifically refer to studies that aim to determine the accuracy, dependability, and consistency of a scale, where scales are tested in certain settings, in specific languages, or with specific populations. In some fields (esp. psychology and education), these studies are referred to in relation to a discipline of psychometrics, the scientific development of measurement models for constructs that may be difficult to observe or categorize (e.g., depression, knowledge sets). There are many practical recommendation guidelines for the development, validation, adaptation, modification and translation of scales, and a variety of methods that are used to rigorously generate, examine, and evaluate scales.

This *Gender & Sex in Methods & Measurement* tool invites researchers to critically consider the questions, “how are scales validated, and with whom?” The answer to these questions may elucidate and expose the limits and shortcomings of validated scales and demonstrate how the validity of any scale is subject to interpretation, including how it is used in specific research contexts.





This tool:

1. Provides guidance for ensuring that gender and sex are accurately, precisely, and inclusively mobilized in the development and validation of new scales. Gender and sex may be embedded into a scale by way of the language used in the scale name or within individual measures, or where gender and sex norms, stereotypes, and societal expectations are less obviously present within the measures. Gender and sex are also relevant when it comes to the people with whom a scale is validated; diversity and representations are key when working towards a highly sensitive scale with a large degree of generalizability across populations.
2. Considers whether, when and how to adapt existing validated scales, recognizing that many existing validated scales may be problematic, flawed or otherwise limited in some ways when it comes to gender and sex. This includes their applicability to populations who are marginalized and minoritized based on their genders, sexes and sexualities.

Developing & validating scales

Ensuring the accurate, precise, and inclusive use of gender and sex concepts in the development and validation of a new scale is crucial both for ethical and scientific reasons.

Ethical significance

1. **Avoiding Harm** – the misrepresentation and misuse of gender and sex concepts has the potential to cause harm. Inaccurate or insensitive measures can perpetuate stereotypes, and mismeasurement can contribute to the erasure and exclusion of marginalized and minoritized participants and communities.
2. **Dignity, Autonomy and Self-Determination** – inclusivity, accuracy, and precision in the mobilization of gender and sex concepts in our research tools is a way of signaling respect for the dignity, autonomy, and self-determination of individual respondents. Scales have the potential of acknowledging and affirming the diverse identities, experiences, and embodiments of participants.
3. **Equity and Justice** – scales can be used to recognize and address unique challenges and disparities, as experienced by various individuals and communities. For example, scales can be used to test the efficacy of an intervention or to understand the magnitude of an issue. As such, accuracy, precision, and inclusion with regards to gender and sex, in the development and validation of new scales can work to ensure that social justice efforts, healthcare policy and practice, social service delivery, etc. are attentive to the needs of all people, across genders, sexes and sexualities.

Scientific significance

1. **Validity and Reliability** – scientifically rigorous scales are essential to ensuring that the research findings are valid and reliable. Inaccurate, imprecise, or exclusionary measurements can lead to misleading and flawed results, undermining the credibility of the research.
2. **Transparency for Generalizability and Adaptation** – it is only by exposing the ways that we mobilized gender and sex concepts in the development and validation of new scales, that researchers can comment on the generalizability of those scales and make informed decisions about whether scale adaptation is needed.



Scale development: Conceptual considerations

Scenario

Drs. Heyam and Al-Ma'mun are developing a new scale to assess different people's values and attitudes about sexual behaviour. They want to understand how gendered and sexed societal norms shape respondents' attitudes towards premarital sex, extramarital sex, contraceptive use, and consent. They are aware of existing "Sexual Double Standard" scales, which present a series of measures in two forms – one relating to women and one relating to men. For example, in a Sexual Double Standard scale, respondents are first asked to rate their agreement on a scale of one-to-five to the statement, "sex before marriage for women is wrong," and then later asked to indicate their agreement with a parallel statement that reads, "sex before marriage for men is wrong."

Consider

Drs. Heyam and Al-Ma'mun want to develop a gender-inclusive scale, one that builds on this idea of sexual double standards. They want to assess whether, and to what extent, respondents' values and attitudes change, when asked about cisgender men and women, and about people of various other gender identities and modalities. They therefore develop a scale that has six versions of each statement. For example, they ask people to rate their agreement with the following six statements: 1) sex before marriage for cisgender women is wrong; 2) sex before marriage for cisgender men is wrong; 3) sex before marriage for transgender women is wrong; 4) sex before marriage for transgender men is wrong; 5) sex before marriage for nonbinary people assigned male at birth is wrong; and 6) sex before marriage for nonbinary people assigned female at birth is wrong.

They are interested in assessing whether respondents have different values about the permissibility of sex before marriage for transgender men and women, as compared to cisgender men and women. As such, they decide to differentiate between cis and trans men and women in statements one thru four. They further reason that statements five and six are important, because they want to assess the extent to which sexist values tied to sex assignment and associated presumed reproductive capacity impact respondents' attitudes towards the sexuality of nonbinary people of different sex assignments.

For example, whether people are more likely to indicate that sex before marriage for nonbinary people assigned female is more wrong than for nonbinary people assigned male, which may expose the extent to which attitudes towards sexual behaviour are entangled with gender and sex-related double standards.

Remember

When designing a new scale where gender, sex and sexuality are important constructs within that scale, it is important to keep the following factors in mind.

Begin with clear understandings of gender & sex as distinct concepts

Drs. Heyam and Al-Ma'mun are interested in how both gender and sex-related values impact attitudes towards sexual behaviour. A scale that only asks about gender (men v. women) and finds differences in reported permissibility of premarital sex is not necessarily sensitive enough to ascertain whether those differences are due to factors relating to sexed embodiment.

Be specific when using gendered language

Drs. Heyam and Al-Ma'mun are aware of binary gendered sexual double standards scales, and they are certain that, despite their use of the words “men” and “women”, these scales are ultimately attempting to assess values and attitudes about the sexuality of cis men and women, specifically.

Importantly, the words men and women are, by default, trans-inclusive. For example, the word woman refers to all women, which includes women who were assigned male at birth. Therefore, it may be important to remind respondents that this gendered language is meant to be interpreted as cis- and trans-inclusive. If you include a statement in your scale, such as, “sex before marriage for women is wrong,” and your intention is for respondents to think about all women, tell them as much, as thinking inclusively about all women may impact their degree of agreement. If, however, your intention is for respondents to be thinking only about the statement in relation to cisgender women, then your values statement will need to be amended so that that more narrow focus is made clear.

Theoretical & explanatory roots

By considering relevant theories, we can determine how different facets of gender, sex and sexuality are – or ought to be – embedded into our measures and scales. Drs. Heyam and Al-Ma'mun are aware of various gendered and sexed societal norms that might inform people's perspectives on sex before marriage, and they try to develop a scale that takes those factors into account. For example, they consider available evidence that suggests that attitudes about premarital sex will be informed by respondents' values and beliefs about pregnancy capacity. As such, they anticipate that respondents will rate the statement, "sex before marriage for cisgender men is wrong," differently than, "sex before marriage for transgender men is wrong." They are concerned that this nuance may be lost or rendered invisible, if the more general statement of, "sex before marriage for men is wrong," is all that is offered. Drs. Heyam and Al-Ma'mun reason that it is also necessary and justified to include two statements about nonbinary people, differentiated by sex assignment, where pregnancy capacity is presumed of those assigned female. Further, they hypothesize that people may moralize and pathologize the sexuality of transgender people more so than the sexuality of cisgender people and will thus be more likely to say that sex before marriage is wrong for all trans people, regardless of gender. They deem it necessary and justified to have values statements in their scale that tease apart different gender modalities, to expose the extent to which that is happening. If differentiation based on sex assignment, gender identity, gender modality, sexuality or other factors occurs within a scale, these differentiations ought to be justified and supported by available evidence. Where this differentiation does not occur, this decision also needs to be well considered, since important nuance may be lost if measures are unjustifiably binary, reductive, too narrow, too broad or otherwise exclusionary.

Engage community

Having community members, advisors, sensitivity readers (someone who reads for bias, misrepresentations, stereotypes or otherwise inappropriate or unclear content), and various stakeholders involved in the scale development process can be helpful for ensuring that the scale is appropriate for use in a variety of contexts, and with diverse respondents. A diversity of engagement and consultation strategies will be more beneficial than a singular conversation with one person; however, more intense engagement and consultation activities may be prohibitively expensive or time consuming if you are working within a tight budget or strict timeline.

Scale validation: Participant considerations

Scenario

Lorimer Hay is a Postdoctoral Research Fellow, and he is working to develop and validate a new inventory of measures for dignity in healthcare settings. He needs to validate the scale for adaptability of care, effective communication, privacy adherence, rapport, choice constraints, etc. Hay decides to use cognitive interviews as one aspect of his validation study. He recruits 50 participants and asks each participant to answer the survey measures in relation to their own recent experience of healthcare. He then uses a series of probes to assess their interpretation of the measure, whether the measure was clear, whether they could easily rate their experience using the provided scale, and whether they had any emotional reactions to the measures that would impact their ability or willingness to answer. Following a series of successful validation tests with the sample of 50 participants, he produces a manuscript to share his newly validated scale.

During the peer-review process, one anonymous reviewer notes that too few details about the validation study participants were included. The reviewer queries whether the participants were all endosex, cisgender and heterosexual, or whether intersex, trans, nonbinary, Two-Spirit, lesbian, gay, bisexual, and queer participants were also included in the validation process. The reviewer knows that more careful delineation of the validation populations is needed, especially in so far as evidence suggests that sexual and gender minoritized populations may have different experiences of healthcare indignity in light of their specific experiences of oppression and marginalization. The reviewer is concerned that a scale validated only with cisgender, heterosexual people will not be sufficiently sensitive to the experiences of other populations, for whom dignity in healthcare is also vital.

Consider

A study that is validated in consultation with only endosex, cisgender and heterosexual people is likely only sensitive to their specific experiences and needs. The resulting validated scale may not be able to be reliably used with people with whom it was not validated. Hay realizes that he never asked his validation study participants about markers of gender, sex, and sexuality. Hay has at least three choices: 1) He can report this oversight as a limitation of his validation study; 2) he can get ethics permission to contact his 50 study participants with a short survey to capture their sociodemographic information; or 3) he can recruit additional participants.

Remember

When validating a new scale, it is important to report clearly on the populations (languages, contexts) with whom the scale was validated. A scale validated only with endosex, cisgender or heterosexual people is not *wrong* – it is simply that this scale may need further testing, to assess its efficacy with intersex, trans, or queer populations. It is only by providing details about the validation process, that other researchers can ascertain the extent to which further adaptation and revalidation is required. Hay decides to focus on *transparency* and *accountability* for this manuscript. He adds what he knows about the validation study participants, namely their ages, that they all lived in a single urban centre, that most were university students, that the process took place in English. He also details everything he does not know about the participants, including their genders, sexes, and sexualities. He is clear in his revised and resubmitted manuscript *with whom* the scale has been validated, includes a discussion about whether the scale may be generalizable and sensitive to the experiences of participants who were not part of the validation study, and he commits to recruiting and reporting on the identities of diverse participants during the next iteration of his postdoc study.

Using & adapting pre-existing validated scales

Sometimes researchers make use of pre-existing validated scales. These scales might clearly detail with whom they were validated and the decision as to whether the scale will be sensitive for use with a new population may be relatively straightforward. A scale developed in English will likely need to be revalidated upon translation to ensure that it is psychometrically sound, efficient, and effective in another language (Fatima et al., 2022; Tsang et al., 2017). A scale developed to measure the magnitude of stressful life events or mental health challenges among people who have never experienced incarceration, will likely need to be adapted and validated anew before it can be reliably used among incarcerated people (Hart, 1997; Jones et al., 2019).

However, the decision whether to use or adapt an existing validated scale is not always so straightforward. Consider the following examples.

Example

A published validation study reports that equal numbers of ‘males and females’ were recruited, and as such, that the scale has been validated across gender.

Question

If the validation process involved the misuse of gender and sex concepts, can we be certain with whom the resulting scale is valid and reliable?

.....

Example

A published validation study may not clearly explain with whom the scale was validated in terms of gender, sex, and sexuality.

Question

If it’s not clear whether intersex, trans or queer people were included in the validation, can we be certain that it will be an effective, valid, and reliable scale when we use it to measure a construct among a sample that includes intersex, trans, and queer people?



Example

A published validation study may describe a scale that is itself gendered or sexed – where the name of the scale, as well as the measures within it, used gendered or sexed language.

Question

If the validated scale measures beliefs about women in leadership, can we be certain that it will be valid and reliable when used to assess beliefs about trans women in leadership? Will it need to be adapted and retested for efficacy, if we want to measure beliefs about nonbinary people in leadership or can we simply replace the word ‘women’ in the measures with ‘nonbinary’ and trust that the scale will continue to work?

.....

Example

A published validation study may describe how the numerical score associated with a scale is gendered – where a score of five for women, but four for men, can be reliably used to assess and diagnose a particular health condition.

Question

If the validated scale has a gendered cut-off, can we be certain that those cut-off scores apply to all men and women, including trans men and women? Or, are they cis-specific, and further testing is needed to ascertain the appropriate cut-off scores for trans and nonbinary people?

The Alcohol Use Disorders Identification Test

Consider


The Alcohol Use Disorders Identification Test (AUDIT) is a ten-item measure used to screen for harmful alcohol use. It is often used in clinical settings to ascertain whether further diagnostic evaluation for alcohol dependence is warranted. It is often used in research settings as an eligibility criterion for alcohol intervention research (where a certain score serves as a threshold for participation), and to measure intervention effectiveness (where changes in score following an intervention are used as evidence that the intervention worked).

It was validated over two decades and was found to provide an “accurate measure of risk across gender, age and cultures” (Babor et al., 2001). These validation studies, however, are replete with many of the common methodological and measurement errors described throughout this toolkit, including the conflation of gender and sex, and the erasure of intersex, trans and participants who are neither endosex nor cisgender. Saunders et al. (1993) describes how the project of developing the AUDIT scale involved six research centres in six countries. However, these authors also describe participants in their own reliability study as being “of both sexes,” who are thereafter named as men and women. It is therefore reasonable to assume that AUDIT was likely not validated specifically with intersex people, binary-identified trans people and nonbinary people.

Further, AUDIT uses gendered scoring, where a score of 13 or more in women and 15 or more in men is likely to indicate alcohol dependence. Other alcohol-related screening tools and measures are similarly gendered. For example, the Centers for Disease Control and Prevention (CDCP) indicate that four or more drinks for women, and five or more drinks for men, constitute binge or heavy episode drinking. The National Institute on Alcohol Abuse and Alcoholism (NIAAA) uses single-item method as indicative of at-risk drinking, where women are asked if they have had four or more drinks in one day in the last year, and men are asked if they have had five or more drinks in one day in the last year.

Question

AUDIT cut-off scores, gendered definitions of heavy episodic drinking and binge drinking, as well as gendered low risk drinking guidelines reflect both numerous, and complex biological mechanisms involved in the metabolism of alcohol, which are problematically reduced to a matter of binary gender and purportedly corresponding binary sex – the presumption being that all women are female, and all men are male.



Gendered dynamics involved in alcohol use are frequently unconsidered, for example where men’s drinking needing to reach a higher threshold than women’s before it is considered harmful even though there is a higher propensity for men to engage in risk-taking behaviour when intoxicated as compared to women (Lowik et al., 2020).

We need to question whether these cut off scores, measures, and screening tools are appropriate for people who are intersex, trans, or Two-Spirit – while also questioning their efficacy among cisgender women and men! If we assume that all men are male-assigned people who are tall, heavy, have large livers, have low estrogen, have high water content in their tissues, and participate in risk-taking when intoxicated, then perhaps a disparate cut off score of 15 as compared to women’s 13 makes sense. However, does this cut-off score sufficiently take into account the myriad of variables (age, body mass, health status, cultural factors, race, genetics) that impact alcohol, in ways that justify a higher score for men as compared to women?

Further, considering the conflation of gender and sex concepts during the validation process, can we reasonably call AUDIT *validated* if it were to be applied to intersex and trans people of all genders? Which cut off score would be appropriate for nonbinary people?

Flentje et al. (2020) sought to provide guidance regarding how to assess harmful alcohol use among gender minority people. They compared items from AUDIT, with both “four or more drinks on one occasion” and “five or more drinks on one occasion” to ascertain which measure performed better in predicting alcohol dependence symptoms and consequences among a group of 1,892 transgender people. They found that screening for five or more drinks on one occasion within the last year performed better than other measures, for all their participants, which included trans women, trans men, and nonbinary people.

Remember

The initial validation process for AUDIT had not included trans people. When questions were raised about which cut-off scores in AUDIT and which “drinks per occasion” metrics from CDCP and NIAAA would be appropriate for trans and nonbinary people, Flentje et al. (2020) needed to retest the scales and measures and determine whether, when and how to use them in assessing trans people for harmful alcohol use. Their finding not only provides a much-needed harmful alcohol use screening methods for trans people; their single, gender-inclusive measure also calls into question the validity of gendered cut-off scores for cisgender people.

The Community Attitudes Abortion Scale

Consider

The Community Attitudes Abortion Scale (CAAS) is a seven-item measure of stigma and attitudes towards women who have had abortions. It was originally validated among a population of 1500 women without a history of abortion, who were seeking abortions in six states across the United States. Most of the seven measures include gendered language and pronouns. For example, “women who have had abortions are bad people,” “women who have had abortions have done something wrong,” and, “if a friend of mine had an abortion, I would not judge her.” The two exceptions are about abortion legality, which ask participants to indicate agreement to the statements, “abortion should be legal and available,” and, “do you think abortion should be legal in all cases or illegal in all cases?” Scores are summed and averaged and given a dichotomous classification of “high stigma” for scores of at least three, and “low stigma” for scores below three.

Question

Considering that this scale includes gendered language in five of the seven measures, can the gendered language be replaced with gender-neutral and inclusive alternative language, without compromising its validity and reliability? If we replace “women” with “people” and “her” with “them” throughout, and use the adapted scale to measure abortion stigma, could we continue to call the scale validated? Or would we need to create a CAAS-Neutral version, and reassess its efficacy? Would it be worthwhile to keep the gendered measures, and add trans-specific ones, to ascertain differences in values and attitudes towards people of different genders who have had abortions?

Remember

Sometimes we can reasonably assume that shifts in language will not impact scale validity and reliability. For example, the “Mothers’ Autonomy in Decision-Making” (MADM) scale was developed to assess women’s experiences of maternity care and has the word “mother” in its name. However, the team at UBC’s Birth Place Lab has been using it, as is, to assess people of all genders’ ability to lead decision-making over the course of their pregnancy-related care. The name of the scale is not shown to participations, and none of the measures use gendered language. Although the scale’s psychometric testing only involved women, the available literature focused on trans pregnant people’s experiences of care suggests that autonomy in decision-making can be similarly measured. As such, the Birth Place Lab research team determined that it could reliably be used across genders. The Birth Includes Us subproject, focused on pregnancy care among LGBTQ/2S families, uses the MADM scale in their survey (Altman et al., 2023).

Sometimes, however, we cannot make use of an existing scale as is, and will need to adapt the scale for use, or retest the existing scale to re-determine and newly establish its effectiveness.

Scale use, adaptation & revalidation: Considerations & recommendations

When in doubt, retest and revalidate – while this is not always feasible since validation studies are resource intensive, this is often recommended. When retesting and revalidation is not possible, careful consideration will need to be given to whether, when and how to use the pre-existing scale, and how to transparently report on any interpretation or analysis assumptions or limitations resulting from its use as is.

Measure twice, report twice

If you are concerned about using a scale that uses gendered language, as is, but cannot devote the time and resources to retesting a gender-neutral version, ask your participants to fill out the scale twice. Then, you can report on the outcomes of the gendered version first, and the gender-neutral version second, and use any deviations in response to reflect on the validity of the gender-neutral adaptation. You can also do your analyses in ways that are typical and expected in your discipline, and then in a transgressive way, to demonstrate the utility and efficacy of the latter approach, all the while satisfying reviewers and readers. Publishing an adapted scale or an unexpected use of an existing scale can allow others to see what you've done, learn from it, and perhaps model their research after your approach – leading to the replication of that approach, which lends itself to validation.

Find the root

Often gendered and/or sexed scoring on scales is a proxy for something else (body weight, height, hormone levels, certain behaviours). If we can determine the true determinant of an outcome, we stand a better chance of developing a rigorous, accurate, precise, and inclusive measure or scale. If changing gendered language within an existing scale does not alter the underlying meaning of the measures, then a gender-neutral revision without further validation may be acceptable (e.g., changing husband/wife to partner, or gay and lesbian to LGBTQ). If, however, the root of the gendered language has to do with some anatomical or physiological process, then revising it to be gender-neutral may change the overall effectiveness of the

measure, and revalidation may be needed (e.g., changing men who have sex with men, to people with penises who have sex with people with penises, which will mean that the new measure is not about gender identity, but about people of all genders who have particular anatomy).

Check out Canan et al., (2023) who evaluated whether gender-inclusive language affected the psychometric properties of a heavily gendered rape myth acceptance scale. Recognizing that the existing validated scale presumed that rape myths were disproportionately more biased against women than men and recognizing that sexual and gender minority communities may assign different meaning to concepts related to rape myths, this study involved efforts to revalidate a gender-inclusive modified form of the scale. The authors reasoned that the heavily gendered measures within the scale would affect its generalizability, and rather than assuming that the scale would be reliable once adapted, they reasoned that a modified version would be different enough from the validated original to necessitate revalidation.

Anticipate & prepare for resistance

You may get push back by editors, reviewers, and readers if you use an existing validated scale with a new population, adapt a scale without retesting, or develop a new scale that addresses concerns you had with something that already exists. Sharing the logics that informed your decisions is helpful, in that it will allow everyone who reads your work to assess the choices that you made. Transparency allows for scientific communication to happen, which can be an important piece of advancing knowledge and understanding. See [Tool 6 Working with Pre-existing Secondary & Older Data](#) for a discussion of strategies for raising concerns about existing data and analyses.

A note on decolonizing validation

It is important to consider the extent to which Indigenous people, including Two-Spirit and Indigiqueer people, are involved in the development, validation, and adaptation of scales, as their absence from those processes can have profound impacts on the validity of scales when used with Indigenous respondents. As Cushman (2016) notes, “legacies of imperialist thought permeate understandings and uses of validity” (p. 1), and the process of developing and validating scales is no exception.

Hill et al. (2010) used a mixed-method approach to investigate the cultural validity of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2), a validated, standardized psychometric test of adult personality and psychopathology scale. They found that the MMPI-2 may pathologize Indigenous worldviews, knowledges, beliefs, and behaviours. Rather than accurately assessing psychopathy, they found that the MMPI-2 appears to judge Indigenous cultural knowledge and practices with regards to physical, emotional, and spiritual symptoms as illegitimate, while privileging Western standards and norms. Indigenous people will have unique and specific experiences of social and emotional wellbeing (Thomas et al., 2010), which merit consideration on their own terms. Further, Two-Spirit and Indigiqueer people, in particular, have unique experiences of inequities such that measuring community resilience using scales validated with non-Indigenous sexual and gender minoritized populations may be insufficient (Parmenter & Galliher, 2022).

Additional reading

This list includes both additional recommended readings and items that were cited in this tool.

Achenbach, T. M., & Ivanova, M. Y. (Undated). Gender to gender-inclusive assessment using the Achenbach System of Empirically Based Assessment (ASEBA): An integrated system of multi-informant assessment. <https://store.aseba.org/mobile/Guide-to-Gender-Inclusive-Assessment-Using-the-ASEBA/productinfo/020/>

Altman, M. R., Cragg, K., van Winkle, T., Julian, Z., Obedin-Maliver, J., Tarasoff, L. A., Eagen-Torkko, M. K., Ferrell, B. L., Rubashkin, N. A., Lusero, I., & Vedam, S. (2023). Birth includes us: Development of a community-led survey to capture experiences of pregnancy care among LGBTQ2S+ families. *Birth*, 50(1), 109-119.

American Psychological Association. (2020, Dec). Psychological and neuropsychological assessment with transgender and gender nonbinary adults. <https://www.apa.org/pi/lgbt/resources/transgender-gender-nonbinary>

Babor, T. F., Higgins-Biddle, J., Saunders, J. B., & Monteiro, M. G. (2001). AUDIT – The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Health Care, Second Edition. *World Health Organization, Department of Mental Health and Substance Dependence*.

Canan, S. N., Cozzolino, L., Myers, J. L., & Jozkowski, K. N. (2023). Does gender inclusive language affect psychometric properties of the Illinois rape myth acceptance scale-short form? A two-sample validation study. *Journal of Interpersonal Violence*, 38(3-4), 3373-3394.

Cowie, L. J., Greaves, L. M., & Sibley, C. G. (2019). Sexuality and sexism: Differences in ambivalent sexism across gender and sexual identity. *Personality and Individual Differences*, 148, 85-89.

Cross, E. J., Muise, A., & Hammond, M. D. (2021). Do scales measuring sexist attitudes have equivalent meaning for sexual minorities and majorities? *Sex Roles*, 85, 707-720.

Cushman, E. (2016). Decolonizing validity. *Journal of Writing Assessment*, 9(1), 1-6.

Fatima, A., Jami, H., & Muzzaffer, K. (2022). Translation and validation of gender minority stress and resilience scale. *Pakistan Journal of Social Science*, 4(3), 530-540.

- Flentje, A., Barger, B. T., Capriotti, M. R., Lubensky, M. E., Tierney, M., Obedin-Maliver, J. & Lunn, M. R. (2020). Screening gender minority people for harmful alcohol use. *PLoS One*, *15* (4), e0231022.
- Hart, C. B. (1997). Adaptation of the social readjustment rating scale for a prison population. *Journal of Offender Rehabilitation*, *25*(3-4), 105-121.
- Hill, J. S., Pace, T. M., & Robbins, R. R. (2010). Decolonizing personality assessments and honoring Indigenous voices: A critical examination of the MMPI-2. *Cultural Diversity and Ethnic Minority Psychology*, *16*(1), 16.
- Jones, R. M., Patel, K., Moscovici, M., McMaster, R., Glancy, G., & Simpson, A. I. F. (2019). Adaptation of the Clinical Global Impression for use in correctional settings: The CGI-C. *Frontiers in Psychiatry*, *10*, 687.
- Kattari, S. K., O'Connor, A. A., & Kattari, L. (2018). Development and validation of the transgender inclusive behavior scale (TIBS). *Journal of Homosexuality*, *65*(2), 181-196.
- Keo-Meier, C. L., & Fitzgerald, K. M. (2017). Affirming psychological testing and neurocognitive assessment with transgender adults. *Psychiatric Clinics of North America*, *40*(1), 51-64.
- Logie, C. H., & Earnshaw, V. (2015). Adapting and validating a scale to measure sexual stigma among lesbian, bisexual and queer women. *PLoS One*, *10*(2), e0116198.
- Lowik, A., Hoong, P., & Knight, K. (2020). Where's the science? A critical interrogation of how sex and gender are used to inform low-risk alcohol use guidelines. *Journal of Addiction Medicine*, *14*(5), 357-359.
- Parmenter, J. G., & Galliher, R. V. (2022). Development and initial validation of the LGBTQ+ community resilience and inequity scale. *Psychology of Sexual Orientation and Gender Diversity*, online ahead of print.
- Saunders, J. B., Aasland, O. G., Amudsen, A., & Grant, M. (1993). Alcohol consumption and related problems among primary health care patients: WHO Collaborative Project on Early Detection with Harmful Alcohol Consumption. *Addiction* *88*, 349-362.
- Schudson, Z. C., & van Anders, S. M. (2021). Gender/sex diversity beliefs: Scale construction, validation, and links to prejudice. *Group Processes & Intergroup Relations*, *25*(4), 1011-1036.

- Stephenson, R., Todd, K., Gamarel, K. E., Bonar, E. E., & Peitzmeier, S. (2020). Development and validation of a scale to measure intimate partner violence amongst transgender and gender diverse populations: Protocols for a linear three-phase study (Project Empower). *JMIR Research Protocols*, 9(11), e23819.
- Thomas, A., Cairney, S., Gunthorpe, W., Paradies, Y., & Sayers, S. (2010). Strong souls: Development and validation of a culturally appropriate tool for assessment of social and emotional well-being in Indigenous youth. *Australian and New Zealand Journal of Psychiatry*, 44(1), 40-48.
- Tsang, S., Royle, C. F., & Sulieman Terkawi, A. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi Journal of Anaesthesia*, 11(1), S80-90.
- Urban, R. E., & Pyland, C. P. (2022). Development and preliminary validation of the gender inclusive rape myth acceptance scale. *Journal of Interpersonal Violence*, 37(21-22), NP20630-NP20652.



Co-Authors/Advisory Team

Dr. A.J. Lowik is the lead author. Co-authors include Dr. Jessica Cameron, Jessy Dame, Dr. Jae Ford, Dr. Lex Pulice-Farrow, Dr. Travis Salway, Dr. Sari van Anders, Dr. Kate Shannon.

Funding

The Project Lead is Dr. A.J. Lowik, CGSHE Postdoctoral Fellow.

This work is funded by a CIHR Sex and Gender Science Chair in Gender-Transformative Sexual Health (PI: Dr. Kate Shannon).

Design & Development

Layout & design by Shivangi Sikri & Altaira Northe, CGSHE-UBC Comms Team.

Advisory & resource support by Majka Hahn & Esther Ibu, CGSHE Research Coordinators for Sexual Health.

Suggested Citation (APA)

Lowik, A., Cameron, J. J., Dame, J., Ford, J., Pulice-Farrow, L., Salway, T., van Anders, S., & Shannon, K. (2024). *Gender and Sex in Methods & Measurement - Research Equity Toolkit. "Tool #7: Developing and Working with Validated Scales,"* Collaborative of Gender and Sexual Health Equity Research Education Resources ("CGSHE"), University of British Columbia.

Note: The advisory authors are listed alphabetically, each having contributed equally to this collaborative project; the exceptions to the alphabetical ordering are the first author (who is the project lead and primary author) and the last author (who is the senior author/PI and CIHR Chair that funded this project).