

# **AN ETPG WHITE PAPER MODERN GENDER CONCEPTUALIZATION AND PSYCHOLOGICAL ASSESSMENT**

**Authors: Dragos Iliescu, University of Bucharest<sup>1</sup>; Andreea Butucescu, University of Bucharest and Raluca Dutu<sup>2</sup>, Stellenbosch University**

This white paper discusses the ways in which the current developments in the conceptualization of gender may influence psychological testing. This intersection between gender identity and psychological testing represents a neglected domain where more discussion and insight are needed (Rider et al., 2019), in order for assessment to continue to address key desiderata: soundness of interpretation, fairness, and equal opportunities for each test taker.

We acknowledge that this topic is ideologically loaded and may incite strong reactions. At the same time, we believe that the discussion is timely as this topic has not been discussed at large, let alone settled by the scientific community. The present paper is intended mostly as a preliminary "food for thought" approach, pinpointing some future directions for research and practice – and hopefully stirring up reactions and contributions from the general public, practitioner networks, researchers, and test developers, and of course test publishers.

## **GENERAL CONTEXT AND IMPORTANCE**

The understanding of gender has evolved greatly in recent years, moving away from the traditional binary perspective of biological sex (i.e., female vs. male), to a much more complex, fluid perspective. This shift in the societal paradigm influences the way in which individuals perceive themselves and are perceived and approached by others in various social, educational, and professional contexts. Some of these interactions may include psychological assessment, and it is therefore important that the testing community discusses the impact of gender conceptualizations in assessment practices.

Reconceptualizing our understanding of gender has brought with it some specific challenges to specific industries, but also some common challenges. In terms of general challenges and changes the most visible refers to the implementation of indiscriminative and inclusive policies (Connell, 2006). Several commercial and educational organizations have publicly declared their strategic egalitarian objectives, which include such key actions as limiting the collection of gender-related information, updating IT systems to allow users to select their preferred pronouns, allowing usernames and email addresses that reflect people's gender identity, making workplace dress codes more flexible so as not to interfere with a person's expression of their gender identity, using gender-neutral language to avoid assumptions about employees' gender identity or sexual preferences. Some of these actions may also be reflected in the practices of test publishers and test users and may show in redesigned online or paper-and-pencil forms, at least in terms of demographic variables, or mode of address.

In terms of more specific challenges, so far, there have not been many signs that the testing and assessment industry is concerned with or knows how to position itself on this front. We advance here a series of questions which are related to some of these specific challenges.

## **CONSIDERATIONS OF SUBJECTIVELY FAIR EXPERIENCES BY TEST TAKERS**

New conceptualizations of gender have a significant impact on many some key elements of test development. Based on these considerations, test providers may need to:

- (a) Review assessment tests from a theoretical standpoint, to ensure that the measure is grounded in the latest available theory and is sensitive to the entire range of possible gender experiences; this involves integrating the latest gender research and theory into the structure and assumptions of the tests.
- (b) Review the test language and ensure both the test questions and descriptions are gender neutral, in order to avoid excluding or offending groups of individuals with diverse gender identities; this also involves avoiding terms and scenarios that are based on gender assumptions.
- (c) Include response options that acknowledge the diversity of gender experiences, allowing test takers to identify themselves in a way that reflects their gender identity; this might mean providing more options than the classical “male/female.”
- (d) Carry out factor analyses for different gender groups, to ensure that the test consistently measures variables that are psychologically constant across groups. Modern approaches for the comparison of factor structures and establishment of equivalence include Multiple Group Confirmatory Factor Analysis. This helps calibrate the test towards accurately and fairly reflecting all gender identities. We acknowledge that this may be impossible for some gender groups that may rarely appear, even in carefully selected representative samples; factor analysis is a data-hungry procedure.
- (e) Address Differential Item Functioning (DIF): DIF occurs when groups (in this case, individuals of different gender identities) have different likelihoods of answering items correctly not because of ability, but because of the inherent nature of the test questions. Test developers habitually screen items for gender-related DIF to ensure that the test is fair and accurate for all participants; this procedure may be expanded beyond the male/female dichotomy.
- (f) Pilot tests with individuals of diverse gender identities and request feedback to understand how questions and answers are perceived and interpreted; this can help to calibrate and improve the test towards being as inclusive as possible.
- (g) Redesign norming. Norms are sometimes provided on gender groups (e.g., for most personality inventories). An assessment of the ways in which diverse group diverge from each other or from the classical male/female norms may be appropriate – and in case such differences are manifest, a re-design of normative tables. If such differences are not manifest (or can be reasonably argued to not be manifest, based on robust tests of construct, method, or item bias), the test could be considered to be fair without the need for several gender norm groups (and consequent gender norm tables).
- (h) Modify and adapt the interpretation guidelines. The way test results are interpreted may also require changes. This includes understanding the ways in which different aspects of gender identity can influence test results and providing guidelines that account for these influences.

(i) Provide training for psychologists and other professionals who will use the test, explaining the importance and approach to gender diversity in assessment, if this topic is potentially important for the respective test.

The above are a number of measures that could be implemented and which, in theory, may help test takers with non-binary gender self-identification to develop a subjectively fair experience regarding the assessment process.

## **CONSIDERATIONS OF OBJECTIVE FAIRNESS**

Beyond the subjective experience, it is critical that test takers are not discriminated against, i.e., that their responses are scored and interpreted in such a way as to reflect their true and unbiased focal trait. This process is highly dependent on norming, i.e., on the way in which the raw scores of a test taker are compared to normative samples that are correctly representative of the focal test taker. Normative data can be aggregated or stratified by demographic factors such as age, education, race, and gender (Bartram & van de Vijver, 2016). Some of these factors have received more attention than others.

For example, for decades, how to aggregate scores across culturally diverse groups (i.e., samples from different countries) has been a significant concern in the norming research and practice (Bartram & van de Vijver, 2016). The reason points to the consequences that derive from such aggregations. Significantly less emphasis has been placed on gender, despite the topic being equally critical. One question is now more prominent than ever: How do we aggregate across genders?

Current conceptualizations make a clear distinction between biological sex and gender. Sex refers to a set of biological characteristics of men and women, while gender refers to the socially constructed features of men and women (Griffin et al., 2021), encompassing roles, behaviors, activities, expectations, and societal norms that cultures may attach to men or women, which are also subject to personal identification and experience. In light of these changes, gender has become a more fluid construct, adding an element of personal choice (Diamond, 2020). The issue we are addressing is rooted in the traditional conflation of 'sex' and 'gender'. Historically, in psychological assessment as in many other areas, 'sex' and 'gender' were conflated, i.e., these two terms were used interchangeably. When applying norms, the reported behaviors of a test takers are compared with the typical behaviors of their gender group. This presumes a direct and consistent alignment of sex and gender, because norm groups are based on biological sex, while behaviors lie in the territory of gender.

Gender is sometimes suboptimally measured and integrated into test scoring and interpretation. Much of the available work still conceptualizes gender as a binary construct (Bartram & van de Vijver, 2016). Unless the topic is a concept specifically designed to measure sexual minorities, most instruments provide norms that apply only to men and women. This topic is in and of itself more relevant in the clinical field, where gender identity may actually be a legitimate measurement purpose. This happens rarely or never in the work and organizational, or the educational domain.

An example is probably welcome here. Let us assume that a test-taker obtains a certain score on a scale measuring the personality trait of Dominance. People who score high on Dominance tend to like to lead, organize and control others, who tend to be outspoken with their ideas and to have good social skills in their interaction with other people; they tend to be assertive, have good social poise, leave a good impression and feel at ease in the presence of others. Dominance has a significant gender difference: men tend in general to be more dominant than women. Because of this, most personality inventories that measure this trait offer gender-specific norms. A score that would be below average for a man may be above average for a woman. If our test taker were a woman, she would be characterized as enterprising, fast, inventive, outspoken, stubborn - and if he were a man he would be characterized as uncertain, tolerant, amiable, polite, discrete, and submissive. Gender-specific norms are critical to understand how our test taker behaves, i.e., more dominant than other women (gender-specific), or more dominant than any other person (non-specific)?

Using conventional norms with non-binary populations may have important implications. For many transgender people, for example, having their test scored with what they consider the wrong norm means denying their identity (Lewitzke, 2018). Using sex-based norms can perpetuate stereotypes and overlook the nuances of gendered behavior. When gender-based norms are unavailable, minority populations tend to be treated unequally, as behaviors that are often measured in psychological tests (such as aggression, empathy, communication styles, etc.) can be heavily influenced by gender norms and roles. This may be especially the case in clinical high-stake contexts, in which such populations are at risk of being overly pathologized (Keo-Meier & Fitzgerald, 2017): some of the scales that are commonly used in these contexts tend to show elevated scores with non-binary populations. These elevations most often reflect the psychological and emotional changes people face in their transition process (Keo-Meier & Fitzgerald, 2017). For example, depression and anxiety scales repeatedly showed elevated scores, which can be due to increased gender minority stress or gender dysphoria (Webb et al., 2016). Elevations on masculinity/femininity scales are also common, depending on whether there is an alignment between the adopted gender and its public manifestation (Tomita et al., 2019). Elevations on the paranoia scale may not always signal a pathology, as they can be caused by a prolonged feeling of being misunderstood, rejected, or even persecuted (Duckworth & Anderson, 2019).

## **FUTURE WORK**

Some short-term solutions to this problem are already used by psychologists in their practice. Such short-term solutions all emphasize professional judgment of psychologists in deciding what criteria should be used for scoring. Ideally, scores would be transformed based on both sets of gender-specific norms and also on the combined (general population) norms (Keo-Meier & Fitzgerald, 2017). However, these solutions are only temporary, and neither will solve the problem. The development of gender-based may be one solution but given the complexity of current gender-related self-identifications, robust norming for all these modes may be unrealistic. Best practices for assessing nonbinary clients are necessary, as they would guide the technical efforts in consistent directions, but they are not available yet (Webb et al., 2016). Fortunately, the level of awareness concerning this topic has started to increase.

We posit that this challenge requires a collective effort from the general public, practitioners (test users), test developers and test publishers. These key stakeholders will ideally to assume an active role: psychologists and other practitioners in charge of test administration and interpretation would have to stay updated such that they can make the optimal decisions consistent with current recommendations, the general public (especially those test takers that are directly affected) would need to be well informed, and test developers/researchers would have to be mindful when designing assessment materials by paying more attention to item wording, construction, reporting and, of course, norming (Cameron & Stinson, 2019).

## **REFERENCES**

- American Psychological Association. (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American Psychologist*, 70(9), 832–864.
- Bartram, D., & van de Vijver, F. J. (2016). Norming. In F. T. L. Leong, D. Bartram, F. M. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment* (pp. 438–448). Oxford University Press <https://doi.org/10.1093/med:psych/9780199356942.003.0030>.
- Cameron, J. J., & Stinson, D. A. (2019). Gender (mis) measurement: Guidelines for respecting gender diversity in psychological research. *Social and Personality Psychology Compass*, 13(11), e12506.
- Connell, R. (2006). The Experience of Gender Change in Public Sector Organizations. *Gender, Work & Organization*, 13(5), 435–452. <https://doi.org/10.1111/j.1468-0432.2006.00316.x>

- Diamond, L. M. (2020). Gender Fluidity and Nonbinary Gender Identities Among Children and Adolescents. *Child Development Perspectives*, 14(2), 110–115. <https://doi.org/10.1111/cdep.12366>
- Duckworth, J. C., & Anderson, W. P. (2019). *MMPI and MMPI-2: Interpretation manual for counselors and clinicians*. Taylor & Francis.
- Griffin, L., Clyde, K., Byng, R., & Bewley, S. (2021). Sex, gender, and gender identity: A re-evaluation of the evidence. *BJPsych Bulletin*, 45(5), 291–299. <https://doi.org/10.1192/bjb.2020.73>
- Keo-Meier, C. L., & Fitzgerald, K. M. (2017). Affirmative Psychological Testing and Neurocognitive Assessment with Transgender Adults. *Psychiatric Clinics of North America*, 40(1), 51–64. <https://doi.org/10.1016/j.psc.2016.10.011>
- Lewitzke, K. (2018). Working with Transgender Clients: Considerations for Psychological Testing and Assessment. *Graduate Student Journal of Psychology*, 17, 78–87.
- McLemore, K. A. (2018). A minority stress perspective on transgender individuals' experiences with misgendering. *Stigma and Health*, 3(1), 53–64. <https://doi.org/10.1037/sah0000070>
- Rider, G. N., Berg, D., Pardo, S., Olson, J., Sharp, C., Tran, K., Calvetti, S., & St. Amand, C. (2019). Using the Child Behavior Checklist (CBCL) with transgender/gender nonconforming children and adolescents. *Clinical Practice in Pediatric Psychology*, 7, 291–301. <https://doi.org/10.1037/cpp0000296>
- Tomita, K. K., Testa, R. J., & Balsam, K. F. (2019). Gender-affirming medical interventions and mental health in transgender adults. *Psychology of Sexual Orientation and Gender Diversity*, 6(2), 182.
- Webb, A., Heyne, G., Holmes, H., & Peta, J. (2016). Which box to check: Assessment norms for gender and the implications for transgender and nonbinary populations. *Society for the Psychology of Sexual Orientation and Gender Diversity*. <https://www.apadivisions.org/division-44/publications/newsletters/division/2016/04/nonbinary-populations>.

## ABOUT THE ETPG

The European Test Publishers Group (ETPG) is a group of psychological test publishers, all based in Europe. It was formed in 1991 to: *'help create an intellectual and commercial environment which values scientific measurement in psychological testing. It supports its members and works with test customers, users, developers, researchers, academics, and policy makers to achieve this.'*

ETPG's annual conference focuses on new developments in psychological testing, looking at ways our industry can meet changing needs. Our white papers on key testing topics contribute to scientific and professional conferences as well as relevant policy proposals. We welcome discussions on ideas that will improve testing's impact on European society.

ETPG works with other associations (the European Federation of Psychological Associations [EFPA], the International Test Commission [ITC], the Association of Test Publishers [ATP] among others) to continually improve testing practices. Its annual conference is a major forum to discuss new test developments and the needs of European test users. ETPG gathers together individuals and companies with huge experience and knowledge in all aspects of psychological testing.

The ETPG's mission is to improve European Testing by:

- raising standards in test design and use;
- progressing creative developments in testing;
- promoting professional use of tests internationally;
- developing fair systems of copyright protection to the benefit of all parties;
- increasing and promoting the benefits of applied psychology to European citizens;
- developing links with test industries outside Europe to raise overall standards.

The group's values focus on:

- Innovating
- Social Engagement
- Being evidence-based
- Using scientific methods to develop products and services.
- Being open and transparent
- Adhering to professional standards

### **Author Note**

Correspondence concerning this article should be addressed to Dragos Iliescu. Email: [dragos.iliescu@fpse.unibuc.ro](mailto:dragos.iliescu@fpse.unibuc.ro)