Central Lancashire Online Knowledge (CLoK)

Title	Gender identity assessment with trans individuals – findings of a systematic literature review of assessment instruments and ethical considerations
Type	Article
URL	https://clok.uclan.ac.uk/32577/
DOI	##doi##
Date	2020
Citation	Henrich, Soren orcid iconORCID: 0000-0002-4792-8725 (2020) Gender identity assessment with trans individuals – findings of a systematic literature review of assessment instruments and ethical considerations. Journal of Criminological Research, Policy and Practice . ISSN 2056-3841
Creators	Henrich, Soren

It is advisable to refer to the publisher's version if you intend to cite from the work. ##doi##

For information about Research at UCLan please go to http://www.uclan.ac.uk/research/

Gender identity assessment with trans individuals – Findings of a systematic literature review of assessment instruments and ethical considerations

Purpose: In several Western legislations, trans individuals must frequently undergo some form of gender identity assessment, for example, to receive legal recognition of their gender or to access therapeutic interventions. Thus, a standardised and empirically supported assessment approach becomes necessary. This article will critically reflect on the current international guidelines for assessments by the World Professional Association for Transgender Health, which will be compared to standards in secure forensic settings, illustrated by British prison policies.

Design: Findings of a systematic literature review following PRISMA standards is presented, summarising the current state of research pertaining to gender identity assessment instruments. Studies were included, when they presented empirical details pertaining assessment approaches and passed the quality appraisal, but were excluded when they did not utilise a trans sample or presented clinical assessments not linked to gender identity.

Findings: The 21 included English articles, which mostly have been published in the United States in the last 20 years, propose a total of 10 different assessment approaches. The most studies support the use of the Minnesota Multiphasic Personality Inventory-2, the Bem Sex-Role Inventory, Body Image Scale for Transsexuals, and the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults. The instruments are briefly summarised.

Practical Implications: It becomes apparent that this field is severely understudied and that there is no consensus regarding the best assessment approach. Hence, any recommendations are only preliminary and are contextualised with further ethical considerations and suggestions for future research.

Originality: This is the first systematic literature review pertaining the (semi-)structured assessment of gender identity.

With current changes in policies, it becomes important to review the state of health services provided to the trans community. Generally, trans, a shortened term for 'transgender' or 'transsexual', refers to individuals that are a different gender than their assigned sex at birth (e.g., Coleman et al. 2012). In the simplest terms, 'sex' is defined as biological features such as hormones or genitalia, while 'gender' is a rather social construct, including society's expectations for gender-specific behaviour and attitudes. The trans community—as any other group in society—depend on health services. However, several studies indicate that trans individuals, have negative experiences when trying to access care, often resulting in treatments being postponed or cancelled, and individuals seeking out unsafe alternatives (James et al. 2016). Similar findings were yielded in a European survey (European Union Agency for Fundamental Rights 2014). These issues are not only limited to the quality of care in the community, but also in secure settings (i.e., prisons or forensic hospitals). There is currently no research available on trans men, but trans women are often sent to male prisons, where they are placed on the lowest level of a masculine-dominated prison

hierarchy (e.g., Glezer, McNiel and Binder 2013), putting them at increased risk of harassment and discrimination (e.g., Brown 2009), physical and sexual victimization (e.g., Jenness, Sexton and Sumner 2009), and developing symptoms of depression and/or commiting suicide (e.g., Cole et al. 1997).

One way to advance care for the trans community is via the improvement of the initial assessment. This is due to the fact that throughout several Western legislations, trans individuals are required to undergo assessments to access appropriate care. The World Professional Association for Transgender Health (WPATH)—an interdisciplinary association promoting empirically supported education, training, and practice—published standards of care (SOC) and assessment of trans and gender non-conforming individuals in the 7th version now (Coleman et al. 2012) that have been incorporated in several policies in various legislations around the globe (Wylie et al. 2016). First and foremost, the standards emphasise that gender non-conformity is not pathological (Coleman et al. 2012), mirroring the de-classification of trans identity as a mental disorder by the World Health Organisation (WHO; 2018), triggering a paradigm shift in the field (Motmans, Nieder, and Bouman 2019). The guidelines specify that any form of assessment should be conducted by professionals with appropriate training in the area; this includes, amongst other requirements, continuous participation in education about the assessment of gender dysphoria, (Coleman et al. 2012). According to the SOC, the diagnosis is one of the central criteria to access any transitional intervention (Coleman et al. 2012). 'Gender dysphoria' is part of the Diagnostic and statistical manual of mental disorders (DSM-5) and subsumes an individual's persistent distress or impairment of daily functioning regarding the incongruence between their assigned gender and their actual gender (American Psychiatric Association [APA], 2013). Here, the DSM-5 primarily focuses on the symptom's consequences (i.e., anxiety, sorrow, or pain related to the incongruence), but does not clarify the syndrome's origin; an aspect that results in criticism about the diagnosis' legitimacy (e.g., Schulz 2018). Hence, trans individuals might be diagnosed with gender dysphoria, even if their distress is not stemming from the incongruence itself, but from discrimination and harassment by others as a response to their gender (Schulz 2018). In addition to the diagnosis, other criteria listed in the SOC that are considered depending on the level of the treatment's irreversibility are, for example, reasonably well controlled medical and/or mental issues or living in the acquired gender for the previous 12 months (Coleman et al. 2012).

While the SOC are general guidelines, some standards—especially in the forensic context— require the assessor to scrutinise the assessed individual even more. An example is the UK prison policies that explicitly list counterevidence (National Offender Management Service 2016). Indicators that a trans individual might not be genuine are: no history of previously living in the acquired gender, insincere motivation for self-disclosure due to personality disorder related issues, and/or ulterior motives to either undermine prison policies, gain access to victims, or for other reasons. Note that while these guidelines are more concrete than the WPAHT's guidance, the policy document does not clarify what aspects of their recommendations are based on empirical evidence. Besides this inclusion of counterevidence and other minor changes, the UK prison guidelines are reasonably comparable to the WPATH recommendations. Some of the differences can be attributed to the fact that the community-based focus is profoundly different from the forensic outlook, because the latter must also account for society's safety in addition to individual's rights.

However, due to the reasonable level of similarity amongst the sets of recommendations, the body of research reviewing the SOC is expected to also be applicable to the UK context.

Overall, scholars view the latest version of the WPATH criteria as a considerable improvement, because the SOC does not require anymore partaking in psychotherapy prior to the review, instead framing the therapeutic alliance as an opportunity to exchange information (Wylie et al. 2016). Furthermore, the guidelines expanded their definition to gender identities beyond the binary system (i.e., beyond just male and female), effectively promoting to normalise all gender expressions (Fraser 2015). However, others advocate the exclusion of the gender dysphoria diagnosis as a criterion (e.g., Lev 2009; Dewey and Gesbeck 2017). While Wylie et al. (2016) regard the diagnostic process as 'straightforward' (p. 5), Lev (2009) states that the diagnosis is one of the most complex, because it predominantly relies on self-disclosure. Beyond the discussed practicality of this diagnosis, Lev (2009) also raises the issue about its implications, reiterating previous observations by Meyer and Reter (1979): It appears when clinicians label trans individuals with such a diagnosis, they are less confident that these individuals can make their own informed decisions. Additionally, research shows that not all trans individuals experience distress due to their gender (Schulz 2018). This can lead to trans assessees feeling pressured into presenting themselves as mentally stable, while also aiming to meet the threshold for a diagnosis, as Dewey and Gesbeck (2017) found when interviewing mental health professionals and trans patients. Further, Jones et al. (2017) found that trans individuals might feel the need to socially transition (i.e., disclose the gender to others and/or present in their gender by dressing a certain way or use another name) for their assessment to be successful; something that is explicitly required by UK prison policies termed 'Actual Life' (National Offender Management Service 2016: 16), and implied by the general community SOC recommending a year of living in the acquired gender (Coleman et al. 2012). However, scholars like Lester (2018) emphasise that living openly trans before any form of medical transition can be for some individuals dangerous and prolongs their distress, deeming the requirement unjust. Further criticism pertains to the arbitrary time scope requirements that are not substantiated by any empirical evidence (e.g., Jones et al. 2017), the vagueness of the guidelines described as 'cryptic and confusing' (Lev 2009: 76) lacking guidance for further psychological treatment, and the fact that multiple assessments are required, can become a financial burden for the assessees, depending on the insurance policies in their given country (Schulz 2018).

Nevertheless, Lev (2009) also recognises that accessing transition-related medical intervention is not the only context in which a psychosocial assessment is required prior to a treatment, referring to other irreversible procedures. Hence, in many legislations, clinicians are required to conduct initial assessments for safeguarding purposes. Furthermore, Lev (2009) emphasises that this also serves as an opportunity to identify possible mental health issues. However, the assessor should not become a 'psychological detective seeking reasons to disqualify the client' (Lev 2009: 92). Instead the community-based assessment should shift to an informed consent approach, where trans clients and clinicians work collaboratively on identifying best treatment options. This is reflective of recent developments in the field, in which scholars advocate to move away from the established gatekeeping approach in care (e.g., Denny, Green, and Cole 2007; Jones et al. 2017; Lester 2018; Schulz 2018). Historically, this approach regards clinicians holding the executive power pertaining the access to trans-related treatments (Hale 2007), often centring around the diagnosis of gender dysphoria,

which runs the risk of pathologising trans individuals (e.g., Schulz 2018). Opposed to this is the informed consent approach, in which trans individuals autonomously decide whether they are eligible for certain treatments, while the role of the practitioner is the one of providing guidance (e.g., Schulz 2018). While there are currently initiatives of this sort successfully conducted, for example in Boston, US (Reisner et al. 2015), it appears that the current SOC does not support such diagnostic changes (Dewey and Gesbeck 2017). This is only in regards of community-based practice; further research is required for care models in secure forensic settings.

The previous outline makes it obvious that a structured guidance for clinical judgement is necessary beyond a list of required criteria. Similar to psychological assessment instruments, the field is in need for a transparent and efficient procedure that is comparable across contexts (Jones et al. 2017). Furthermore, it must encompass a non-pathologising perspective to reflect the recent changes. Wanta and Unger (2017) demonstrate in their review that in both aspects research and overviews are severely lacking, making the current study necessary. Hence, to inform best practice for the gender identity assessment, this study aims to review the available empirical evidence regarding all structured or semi-structured approaches pertaining this topic discussed in the literature. A systematic literature review is conducted. In the following sections, the methodology is presented, providing an overview of the literature in the field—with special focus on clinical implications and quality of the research—and concluding with a discussion that contextualises the studies' findings with further ethical considerations.

Methodology. A systematic literature review was conducted, following the standards of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA; Moher et al., 2009).

Data search. The inclusion criteria for the literature search were as follows: (1) The paper had to include some form of structured gender identity assessment; (2) it had to relate to the trans community specifically; (3) the publication had to provide empirical support for the presented assessment approach, and (4) the empirical methodology arriving on those findings had to be deemed good. Hence, papers were excluded, if they only presented general guidelines, provided no empirical details, did not utilise a trans sample, and/or related to any other aspect of clinical assessment regarding trans individuals (e.g., sexual orientation or fetishism). The following search string was used, considering only English papers published till April 2019.: "transgender" AND "assessment" NOT "legal" OR "legislation" OR "law" OR "politics" OR "HIV" OR "cancer" OR "fertility" OR "pregnancy". The search was conducted using the following databases: PsycInfo, PsycArticles, MEDLine, Criminal Justice Abstracts, SocINDEX. Articles.

Quality appraisal. To establish whether studies met the last inclusion criterion, for each included study a quality appraisal was completed using an amended 12-item version of the Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies checklist developed by the Agency for Healthcare Research and Quality (AHRQ). Amongst others, this included whether studies utilised inferential statistics. The quality of the paper was then summarised as 'good', 'fair', or 'poor'. Both, the search and the appraisal, was reviewed by a second rater. There were minor discrepancies on item level, however, these were resolved through discussion.

Results.

Literature search. The initial literature search yielded a total of 816 articles. After removing duplicates, overviews, opinion pieces, and reviews, 294 titles were screened for relevance. The step was repeated on an abstract-level with 97 articles. This resulted in a total of 56 articles, for which full-text copies were obtained for further screening. For 5 publications, no copies were accessible, because no contact to the authors could be established. An in-depth review of the remaining articles regarding the previously described inclusion criteria led to the removal of 15 further texts. However, 8 articles were added based on the hand search of the full text references. Lastly, before the quality appraisal, all case studies were removed, a total of 7. From the resulting 37 publications, 13 were labelled 'Fair', and 3 were labelled as 'Poor', meaning that the final review consists of 21 articles labelled 'Good' (Figure 1).

[Insert Figure 1 here.]

Notably, the studies that were viewed having poor or fair quality, none of them utilised advanced statistical analyses and only 7 of them employed control groups. Out of the 3 studies classified as 'Poor', only one defined their independent variables, and 2 either did not define or did not report their outcome variables.

Characteristics of included studies. Out of the 21 remaining studies, all were conducted in Western countries—with 10 out of 21 stemming from the United States, and 15 were published during or after 2000 (see Table 1). Note that 'inpatient' usually referred to trans individuals being in transition-related treatment. Some of the studies also contrasted their findings with cis-gendered participants, referring to individuals who experience a congruence between the gender they got assigned with at birth and their current gender identity, as opposed to trans individuals. Furthermore, note that some studies refer to their participants as being diagnosed with 'gender identity disorder'. This is an outdated syndrome description used in previous iterations of the DSM (e.g., 4th ed., text rev.; *DSM–IV–TR*; APA 2000); other conceptualisations used in the listed publications, that would not be used today, are 'cross-dressers' and 'transvestites', both terms commonly referring to cis men who wear female clothes.

[Insert Table 1 here.]

Overview of assessment instruments. Overall, 21 studies propose 10 different approaches to assess gender identity, or aspects linked to this. Table 2 is providing an overview of these assessment instruments, with their respective studies supporting or contradicting their utility listed next to them. Due to the scope of this article, only the four assessment instruments with the most published English articles are described in more detail. Furthermore, it should be noted that some studies utilised several assessment instruments to contrast their main approach (e.g., to establish congruent validity). Only the main instruments are described in the subsequent presentation, because quality indexes for contrasting instruments were often not reported.

[Insert Table 2 here.]

Minnesota Multiphasic Personality Inventory-2 MMPI-2 (Butcher et al. 1989). In the current review, this instrument has received the most research, with 7 out of 21 publications. Briefly outlined, the MMPI-2 is a standardised, well validated (e.g., Ægisdóttir *et al.* 2006) psychometric test in its second version, with 567 items, including several scales and sub-scales, that provide information about an individual's personality and psychopathology.

Early research pertaining the MMPI concluded that trans individuals demonstrate high ratings related to their gender on the MF-scale, which represents culturally stereotypical male and female behaviour (Langevin, Paitich, and Steiner 1977; Roback *et al.* 1976; Leavitt and Berger 1990; Caron and Archer 1997). The scores on this scale were often comparable to the trans participants' cis-gendered equivalents, meaning that trans women on average expressed similar levels of femininity as cis women and vice versa (Langevin, Paitich, and Steiner 1977; Roback *et al.* 1976; Caron and Archer 1997). These findings were later replicated for the MMPI-2 (Michetl *et al.* 2002; Keo-Meier *et al.* 2015). Gómez-Gil, Vidal-Hagemeijer, and Salamero (2008) did not directly compare trans to cis participants, but concluded that on average the trans patient's scores were within the norm range.

However, there are conflicting conclusions regarding trans individuals' psychopathology. Earlier studies found evidence of psychopathology with trans patients, when compared to cis patients. Roback *et al. (*1976) concluded based on the validity scales (that they did not report in the study) that trans patients were generally more self-critical than the cis control group, but found no pronounced neurotic or psychotic tendencies. The authors suspected that latter was due to trans patients attempting to present themselves more favourable during the assessment. Caron and Archer (1997) also found slight elevations for trans individuals on some validity scales (e.g., K-scale pertaining to corrections), leading the authors to theorise that the absence of any psychopathology in their sample could be due to impression management or in fact higher levels of resilience.

Langevin, Paitich, and Steiner's (1977) results indicated antisocial tendencies that they attributed to a general negative attitude towards society due to the trans patients' previous experiences. Leavitt and Berger (1990) found the most pronounced clinically relevant profile out of all the reviewed articles, describing trans women who were sexually inactive or received pleasure from their penis during sex as experiencing distress due to their body perception, expressing anger or struggle towards society, mistrust, and schizophrenic tendencies, especially compared to trans women who avoided their penis during sexual intercourse. Recent studies only found sub-clinical levels of distress or social introversion that were comparable to the general public and which mostly subsided after the studied trans patients received transitional treatment of any kind (Gómez-Gil, Vidal-Hagemeijer, and Salamero 2008; Keo-Meier *et al.* 2015). Michetl *et al.* (2002) linked this decrease in distress positively to the degree to which trans women in their study expressed their femininity.

Regarding the use of the MMPI-2 for assessment with trans individuals, it is recommended that the assessor takes into account whether the assessee has undergone transition to not over-emphasise mild psychopathology before the start of a treatment (Keo-Meier *et al.* 2015). Furthermore, Gómez-Gil, Vidal-Hagemeijer, and Salamero (2008) and Keo-Meier *et al.* (2015) explicitly emphasise that assessors should use the MMPI-2 interpretation template for the MF-scale that is in line with the trans individuals current gender, as opposed to the gender they were assigned to at birth; for other scales, it can be useful to interpret the

scores in accordance with both templates, due to a currently unclear level of evidence in research.

Bem Sex-Role Inventory (Bem 1981). This assessment instrument received 4 out of 21 studies. The inventory is predominantly a research methodology to explore masculine and feminine self-identification. Individuals rate a total of 60 items on a 7-point Likert-scale indicating how much the gender-stereotypical traits relate to themselves. The scores for femininity and masculinity are compiled separately and then compared to norms, subsequently indicating whether an individual is identifying as masculine, feminine, androgynous (i.e., as both), or undifferentiated (i.e., none of the two).

Herman-Jeglińska, Grabowska, and Dulko (2002) showed that overall trans women score comparably to cis women on the inventory, and trans men are comparable in their results to cis men. Factor and Rothblum's (2017) conclusions were similar, demonstrating in their study that trans women and trans men on average scored equally on the inventory as compared to their respective cis sisters and brothers, while being significantly different to each other. When only focusing on a sample of trans women, Leavitt and Berger (1990) found the highest rate for femininity scores in participants who either avoided their penis or received pleasure via their penis during sex. The sexually inactive sample was scored on average equally as feminine and androgynous. Nevertheless, all three groups were higher on the feminine scores as the cis male norms that are part of the inventory. However, Gómez-Gil *et al.* (2012) only found the 20 items relating to femininity were able to reliably distinguish trans and cis men from trans and cis women, in a Spanish sample.

Body Image Scale for Transsexuals BIS (Lindgren and Pauly 1975). The 30-item self-rating questionnaire explores an individual's satisfaction with their own body on a 5-point Likert-scale. Originally, it was designed by Lindgren and Pauly (1975) specifically with the trans community in mind, surveying 16 trans women and 16 trans men. The items are available in two versions, depending on the individual's assigned gender at birth. Van de Griff *et al.* (2016a) confirmed six sub-scales referring to sources of dissatisfaction in an international sample before their medical treatment: social and hair (i.e., features that might be perceived as dissatisfying by the trans individual in social situations, e.g., general appearance or voice), head and neck, muscularity and posture, hip region, breasts, and genitalia. Overall, the authors showed that all participants experienced the highest degree of dissatisfaction with their primary sex characteristics (e.g., genitalia) and secondary sex characteristics (e.g., pubic hair, breasts). Furthermore, trans women appeared to be more dissatisfied with their body than trans men. Two subsequent studies by Van de Griff *et al.* (2016b, 2016c) revealed similar patterns. While genitalia remained a predominant source of dissatisfaction, trans women reported consistently higher ratings for the 'social and hair' sub-scale and trans men seemed predominantly dissatisfied with their muscularity and posture.

Van de Griff *et al.* (2016a) attributed this to trans men socially transitioning more often before the medical transition, meaning that they more visibly live in the social male role. This could facilitate perception of congruence between body and identity, lowering dissatisfaction. Van de Griff *et al.* (2016a) see the reason for less social transition in the trans women sample in the fact that male roles are generally more accepted in society than female. This could also have an impact on how clinicians assess their trans patients more or less favourably (Van de Griff *et al.* 2016a).

Overall, the BIS appears to have good congruent validity, in comparison with several other assessment instruments (e.g., self-esteem measurements; Van de Griff 2016b), and it appears to reliably depict treatment successes (Van de Griff 2016c). Limitations of these studies are that no collateral information was available to support self-reports, that only correlational analyses were conducted (limiting the interpretation of causality), and that only body (dis)satisfaction is measured (e.g., Van de Griff *et al.* 2016b).

Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults (GIDYQ-AA, Deogracias et al. 2007). This assessment instrument also received 4 out of 21 studies. The GIDYQ-AA is a 27-item questionnaire in male- and female-version, in which an individual indicates on a 5-point Likert-scale how often they have experienced subjective, social, somatic, and sociolegal indicators of gender dysphoria over the past 12 months. According to Deogracias *et al.* (2007) the instrument is based on previous unpublished considerations by the North American Task Force on Intersexuality Research Protocol Working Group, as well as studies by Cohen-Kettenis and Van Goozen (1997) and Docter and Fleming (2001). Latter publication, as well as their preceding study (Docter and Fleming 1993), are part of this review, but are not presented in detail, because they merely influenced the wording of items in the final version of the GIDYQ-AA. The suggested 70-item questionnaire by Docter and Fleming (1993, 2001) appears to not be in use anymore.

Deogracias *et al.* (2007) found a one-factor solution in their comparison of trans female patients and hetero- and homosexual cis women, as well as between trans male patients and hetero- and homosexual cis men. This means that the total score of the GIDYQ-AA successfully distinguished between trans participants and cis participants, with reported sensitivity being 90.4% and specificity being 99.7%. Similar results were found, when the GIDYQ-AA was conducted with trans adolescents (specificity = 100%, sensitivity = 93.3%), and when trans adults and trans adolescents were contrasted with a clinical control group representing a heterogeneous variety of mental disorders who self-identified on the questionnaire as cis-gendered (specificity = 100%, sensitivity = 87.3%; Singh *et al.* 2010). The authors from both studies (Deogracias *et al.* 2007; Singh *et al.* 2010) recommended that the instrument could be used to explore aspects of the gender dysphoria diagnosis and could successfully distinguish trans patients from other psychiatric patients.

Conclusion. The review highlights the great variety of different approaches to assess gender identity. But it seems that a critical reflection on the instruments is lacking, with studies often only presenting one instrument at a time, and only Langevin, Paitich, and Steiner (1977) and Gómez-Gil *et al.* (2012) presenting counterevidence for the utility of certain assessments. It reiterates Wanta and Unger's (2017) conclusion about the state of trans research, demonstrating that only a limited amount of good-quality publications exists on this topic. Two main areas emerged: The assessment of gender identity and general psychopathology (addressed by the MMPI-2 and the Bem Sex-Role Inventory), and the specific assessment of gender identity dysphoria (addressed by the BIS and the GIDYQ-AA).

The MMPI-2 is rather lengthy, but offers an explorative insight into an individual's personality and psychopathology. In contrast, the Bem-Sex Role Inventory is limited in its scope, but can be completed in a timely manner. Both demonstrated that trans individuals exhibit similar scores than their respective cis counterparts. It appears that both instruments

need further research to establish valid interpretation guidelines, with authors voicing uncertainty whether some scores are due to impression management or expressions of resilience. Either way, most publications emphasise that these instruments are not designed to measure deception regarding gender identity and therefore, should not be used as such. Furthermore, the instruments explicitly assess male and female stereotypes, deeming them unsuitable for individuals whose identity does not fit into the binary system.

When specifically addressing gender dysphoria, the BIS and the GIDYAQ-AA seem more appropriate. The BIS is explicitly designed to prioritise areas of medical intervention for trans individuals who seek transition based on their predominant source of dissatisfaction and to track treatment successes. This makes it suitable to be used in a collaborative therapeutic approach. The only instrument that claims to be able to distinguish trans individuals experiencing gender dysphoria from the general clinical population is the GIDYQ-AA. While it could seem plausible to utilise such instruments for settings like UK prisons, where the exploration of counterevidence is required, more research is needed. It is unclear how the studies established ground truth regarding the control group's gender identities and none of the studies addressed ulterior motivations. Furthermore, it is currently unclear to what extend attempts of deception regarding gender identity in forensic settings is actually a problem. However, at minimum the GIDYQ-AA could provide a more structured, transparent, and empirical approach to this assessment context. While the instruments do not rely on the exploration of stereotypical gender expressions, they are based on the construct of gender dysphoria, which is not without criticism.

Ethical considerations for application. It must be emphasised that these recommendations are only preliminary. At no point the impression should arise that the mere awareness or knowledge about the presented assessment instruments is sufficient. Authors like Keo-Meier *et al.* (2015), who endorse the use of certain assessments, are stressing that clinicians should be extremely cautious in their interpretation of results. Assessors must be aware of potential biases regarding personal beliefs and liability (e.g., Hale 2007; Dewey and Gesbeck 2017). Clinicians should also not only balance risks and benefits, but enable clients to be autonomous (Hale 2007). Hale (2007) states that this is not achieved via the gate-keeping approach promoted by the SOC, but through the informed-consent approach. As a perquisite competence for own decisions should be assumed and further assessments should only be indicated, if questions arise.

Limitation. This review is naturally limited due to the scope of this article. Not all assessment instruments discussed in the current literature could be presented. That should not negate the fact that there are other valid considerations and new approaches that could be worthwhile for the assessment of gender identity. Furthermore, this article only summarised English publications. In other countries, there are other well-established instruments (e.g., FBeK), but they are published in other languages and therefore could not be included in this review.

Future research. The field is in need for more good quality research. It is apparent that there is not one superior assessment approach to explore gender identity. However, despite vague international guidelines and a lack of consensus regarding best practice, trans individuals are often legally obliged take part in assessments to receive certain treatment or access. Future research must clarify how assessors currently make decisions, what criteria they utilise, and

in which point of the process the here listed assessment instruments can be efficiently and reliably utilised. It must be identified what aspects of the process are working well and which aspects are problematic. As part of this, future research must include trans perspectives more, as well as addressing practitioner's needs for protection and risk reduction, thus they can competently safeguard vulnerable individuals. Only when both perspectives are included, a more appropriate assessment approach can be developed.

References

Ægisdóttir, S., White, M.J., Spengler, P.M., Maugherman, A.S., Anderson, L.A., Cook, R.S., Nichols, C.N., Lampropoulos, G.K., Walker, B.S., Cohen, G. and Rush, J.D., 2006. The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. The Counseling Psychologist, 34(3), pp.341-382.

American Psychiatric Association, 2000, *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.), Washington, DC: Author

American Psychiatric Association, 2013, *Diagnostic and statistical manual of mental disorders* (5th ed.), Washington, DC: Author.

Becker, I., Auer, M., Barkmann, C., Fuss, J., Möller, B., Nieder, T.O., Fahrenkrug, S., Hildebrandt, T. and Richter-Appelt, H., 2018. A cross-sectional multicenter study of multidimensional body image in adolescents and adults with gender dysphoria before and after transition-related medical interventions. *Archives of sexual behavior*, *47*(8), pp.2335-2347.

Bem, S.L., 1981. Bem sex-role inventory. *Professional Manual*.

Brown, G.R. 2009, Recommended revisions to the World Professional Association for Transgender Health's Standards of Care section on medical care for incarcerated persons with gender identity disorder. *International Journal of Transgenderism*, *11*(2), pp.133-139.

Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A. and Kaemmer, B., 1989. Manual for the administration and scoring of the MMPI-2.

Caron, G.R. and Archer, R.P., 1997. MMPI and Rorschach characteristics of individuals approved for gender reassignment surgery. *Assessment*, *4*(3), pp.229-241.

Cohen-Kettenis, P.T. and Van Goozen, S.H., 1997, Sex reassignment of adolescent transsexuals: a follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(2), pp.263-271.

Cole, C.M., O'Boyle, M., Emory, L.E. and Meyer III, W.J. 1997, Comorbidity of gender dysphoria and other major psychiatric diagnoses. *Archives of sexual behavior*, *26*(1), pp.13-26.

Coleman, E., Bockting, W., Botzer, M., Cohen-Kettenis, P., DeCuypere, G., Feldman, J., Fraser, L., Green, J., Knudson, G., Meyer, W.J. and Monstrey, S. 2012, Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *International journal of transgenderism*, *13*(4), pp.165-232.

Denny, D., Green, J. and Cole, S. 2007, Gender variability: Transsexuals, crossdressers, and others. *Sexual health*, *4*, pp.153-187.

Deogracias, J.J., Johnson, L.L., Meyer-Bahlburg, H.F., Kessler, S.J., Schober, J.M. and Zucker, K.J., 2007. The gender identity/gender dysphoria questionnaire for adolescents and adults. *Journal of sex research*, *44*(4), pp.370-379.

Dewey, J.M. and Gesbeck, M.M. 2017, (Dys) Functional diagnosing: Mental health diagnosis, medicalization, and the making of transgender patients. *Humanity & Society*, *41*(1), pp.37-72.

Docter, R.F. and Fleming, J.S., 1993. Dimensions of transvestism and transsexualism: The validation and factorial structure of the Cross-Gender Questionnaire. *Journal of Psychology & Human Sexuality*, *5*(4), pp.15-38.

Docter, R.F. and Fleming, J.S., 2001, Measures of transgender behavior. *Archives of Sexual Behavior*, *30*(3), pp.255-271.

European Union Agency for Fundamental Rights 2014, *Being Trans in the European Union Comparative analysis of EU LGBT survey data*. Publications Office of the European Union.

Factor, R.J. and Rothblum, E.D., 2017. A comparison of trans women, trans men, genderqueer individuals, and cisgender brothers and sisters on the Bem sex-role inventory: Ratings by self and siblings. *Journal of homosexuality*, *64*(13), pp.1872-1889.

Fraser, L., 201,. Standards of care, transgender health. *The international encyclopedia of human sexuality*, pp.1115-1354.

Glezer, A., McNiel, D.E. and Binder, R.L. 2013, Transgendered and incarcerated: A review of the literature, current policies and laws, and ethics. *Journal of the American Academy of Psychiatry and the Law Online*, *41*(4), pp.551-559.

Gómez-Gil, E., Gómez, A., Cañizares, S., Guillamón, A., Rametti, G., Esteva, I., Vázquez, A. and Salamero-Baró, M., 2012. Clinical utility of the Bem Sex Role Inventory (BSRI) in the Spanish transsexual and nontranssexual population. *Journal of Personality Assessment*, *94*(3), pp.304-309.

Gómez-Gil, E., Vidal-Hagemeijer, A. and Salamero, M., 2008. MMPI–2 characteristics of transsexuals requesting sex reassignment: Comparison of patients in prehormonal and presurgical phases. *Journal of personality assessment*, *90*(4), pp.368-374.

Hale, C.J. 2007, Ethical problems with the mental health evaluation standards of care for adult gender variant prospective patients. *Perspectives in biology and medicine*, *50*(4), pp.491-505.

Herman-Jeglińska, A., Grabowska, A. and Dulko, S., 2002. Masculinity, femininity, and transsexualism. *Archives of Sexual Behavior*, *31*(6), pp.527-534.

James, S., Herman, J., Rankin, S., Keisling, M., Mottet, L. and Anafi, M.A. 2016. The report of the 2015 US transgender survey.

Jenness, V., Sexton, L. and Sumner, J. 2009, Transgender inmates in California's prisons: An empirical study of a vulnerable population. In *The California Department of Corrections and Rehabilitation Wardens' Meeting. Irvine, CA: University of California, Irvine, Center for Evidence-Based Corrections, Department of Criminology, Law and Society*.

Jones, B.A., Brewin, N., Richards, C., Van Eijk, M., Stephenson-Allen, A. and Arcelus, J. 2017, Investigating the outcome of the initial assessment at a national transgender health service: Time to review the process?. *International Journal of Transgenderism*, *18*(4), pp.427-432.

Keo-Meier, C.L., Herman, L.I., Reisner, S.L., Pardo, S.T., Sharp, C. and Babcock, J.C., 2015. Testosterone treatment and MMPI–2 improvement in transgender men: A prospective controlled study. *Journal of consulting and clinical psychology*, *83*(1), p.143.

Langevin, R., Paitich, D. and Steiner, B., 1977. The clinical profile of male transsexuals living as females vs. those living as males. *Archives of Sexual Behavior*, *6*(2), pp.143-154.

Leavitt, F. and Berger, J.C. (1990), "Clinical patterns among male transsexual candidates with erotic interest in males", Archives of Sexual Behavior, Vol. 19 No. 5, pp. 491-505.

Lester, C.N. 2018, *Trans Like Me: Conversations for All of Us*. Seal Press.

Lev, A.I. 2009, The ten tasks of the mental health provider: Recommendations for revision of the World Professional Association for Transgender Health's Standards of Care. *International Journal of Transgenderism*, *11*(2), pp.74-99.

Lindgren, T.W. and Pauly, I.B., 1975. A body image scale for evaluating transsexuals. *Archives of sexual behavior*, *4*(6), pp.639-656.

Meyer, J.K. and Reter, D.J. 1979, Sex reassignment: Follow-up. *Archives of General Psychiatry*, *36*(9), pp.1010-1015.

Michel, A., Legros, J.J., Cornet, J.P., Ansseau, M., Pitchot, W. and Mormont, C., 2002. Comparisons of two groups of sex-change applicants based on the MMPI. *Psychological reports*, *91*(1), pp.233-240.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. (2009), "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement", Annals of Internal Medicine, Vol. 151 No. 4, pp. 264-269.

Motmans, J., Nieder, T.O. and Bouman, W.P. 2019, Transforming the paradigm of nonbinary transgender health: A field in transition.

National Offender Management Service 2016, *The Care and Management of Transgender Offenders*, accessed 25 November 2019, https://www.justice.gov.uk/downloads/offenders/psipso/psi-2016/PSI-17-2016-PI-16-2016-AI-13-2016-The-Care-and-Management-of-Transgender-Offenders.docx.

Reisner, S.L., Bradford, J., Hopwood, R., Gonzalez, A., Makadon, H., Todisco, D., Cavanaugh, T., VanDerwarker, R., Grasso, C., Zaslow, S. and Boswell, S.L. 2015, Comprehensive transgender healthcare: the gender affirming clinical and public health model of Fenway Health. *Journal of Urban Health*, *92*(3), pp.584-592.

Roback, H.B., McKee, E., Webb, W., Abramowitz, C.V. and Abramowitz, S.I., 1976. Psychopathology in female sex-change applicants and two help-seeking controls. *Journal of abnormal psychology*, *85*(4), p.430.

Schulz, S.L. 2018, The informed consent model of transgender care: An alternative to the diagnosis of gender dysphoria. *Journal of humanistic psychology*, *58*(1), pp.72-92.

Singh, D., Deogracias, J.J., Johnson, L.L., Bradley, S.J., Kibblewhite, S.J., Owen-Anderson, A., Peterson-Badali, M., Meyer-Bahlburg, H.F. and Zucker, K.J., 2010. The gender identity/gender dysphoria questionnaire for adolescents and adults: Further validity evidence. *Journal of sex research*, *47*(1), pp.49-58.

Sjoberg, M.D., Walch, S.E. and Stanny, C.J., 2006. Development and initial psychometric evaluation of the Transgender Adaptation and Integration Measure (TG AIM). *International Journal of Transgenderism*, *9*(2), pp.35-45.

van de Grift, T.C., Cohen-Kettenis, P.T., Steensma, T.D., De Cuypere, G., Richter-Appelt, H., Haraldsen, I.R., Dikmans, R.E., Cerwenka, S.C. and Kreukels, B.P., 2016a, Body satisfaction and physical appearance in gender dysphoria. *Archives of sexual behavior*, *45*(3), pp.575-585.

van de Grift, T.C., Cohen-Kettenis, P.T., Elaut, E., De Cuypere, G.R.E.T.A., Richter-Appelt, H., Haraldsen, I.R. and Kreukels, B.P., 2016b, A network analysis of body satisfaction of people with gender dysphoria. *Body image*, *17*, pp.184-190.

van de Grift, T.C., Kreukels, B.P., Elfering, L., Özer, M., Bouman, M.B., Buncamper, M.E., Smit, J.M. and Mullender, M.G., 2016. Body image in transmen: Multidimensional measurement and the effects of mastectomy. *The journal of sexual medicine*, *13*(11), pp.1778-1786.

Wanta, J. W., & Unger, C. A. 2017, Review of the transgender literature: where do we go from here?. *Transgender health*, *2*(1), 119-128.

World Health Organization (WHO) 2018, *ICD-11: Classifying disease to map the way we live and die,* accessed 5 November 2019, https://www.who.int/news-room/spotlight/international-classification-of-diseases.

Wylie, S.A., Corliss, H.L., Boulanger, V., Prokop, L.A. and Austin, S.B. (2010), "Socially assigned gender nonconformity: a brief measure for use in surveillance and investigation of health disparities", Sex Roles, Vol. 63 Nos 3/4, pp. 264-276.

Wylie, K., Knudson, G., Khan, S.I., Bonierbale, M., Watanyusakul, S. and Baral, S (2016), "Serving transgender people: clinical care considerations and service delivery models in transgender health", The Lancet, Vol. 388 No. 10042, pp. 401-411.

Appendix

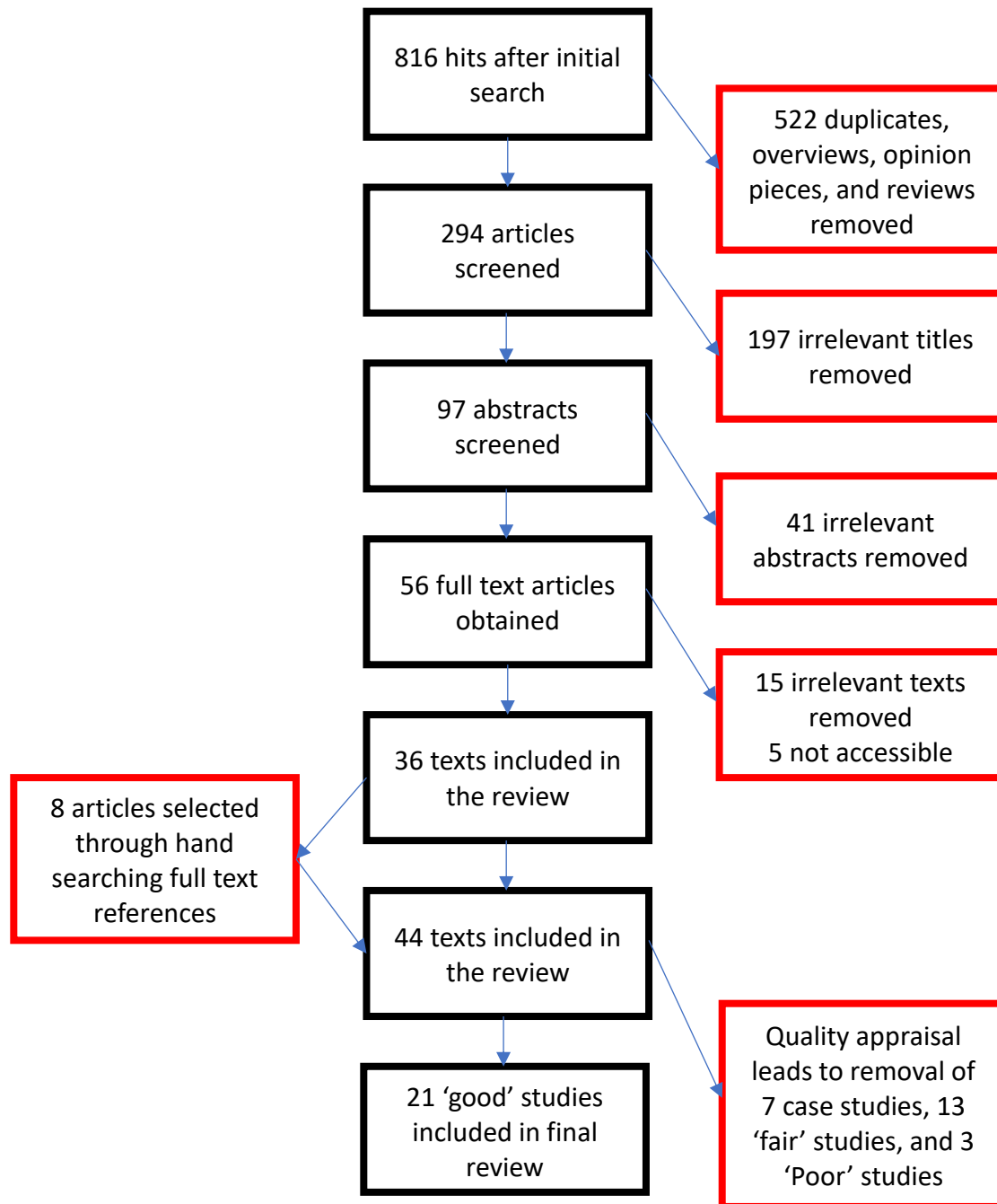**Figure 1. Flowchart depicting the consecutive steps of the systematic literature review following the PRISMA guidelines.**

Table 1. Study characteristics of all reviewed English 'good'-quality publications.

Reference	Country	Study design	Participants demographic
Becker *et al.* 2018	Germany	cross-sectional at different points of transition	N_1 = 82 trans adolescents N_2 = 120 trans adults
Caron & Archer 1997	United States	comparison with matched control-group	N_1 = 112 trans inpatient N_2 = 122 cis inpatient N_3 = 122 cis members from community
Deogracias *et al.* 2007	Canada	comparison with control-group	N_1 = 73 adult & adolescent gender identity (GID) patients N_2 = 389 university-based participants
Docter & Fleming 1993	United States	cross-validation, randomised into 2 validation groups	N_1 = 518 cis male crossdressers N_2 = 78 periodical cis male crossdressers N_3 = 86 trans women
Docter & Fleming 2001	United States & Canada	cross-sectional	N_1 = 454 cis male transvestites N_2 = 62 trans women
Factor & Rothblum 2017	United States	comparison with sibling control-group	N_1 = 295 gender non-conforming adults N_2 = 129 siblings of N_1
Gómez-Gil *et al.* 2012	Spain	comparison with control-group	N_1 = 121 trans patients N_2 = 156 cis members of community
Gómez-Gil, Vidal-Hagemeijer, & Salamero. 2008	Spain	cross-sectional	N_1 = 107 trans female patients N_2 = 56 trans male patients
Herman-Jeglińska, Grabowska, & Dulko 2002	Poland	comparison with control group	N_1 = 103 trans male patients N_2 = 29 trans female patients N_3 = 135 cis men N_4 = 303 cis women
Keo-Meier *et al.* 2015	United States	longitudinal comparison with matched control-group	N_1 = 48 trans men N_2 = 53 cis men N_3 = 62 cis female
Langevin, Paitich, & Steiner 1977	Canada	comparison with control-groups	N_1 = 25 trans female patients N_2 = 19 crossdressing male patients N_3 = 20 homosexual cis male patients N_4 = 19 homosexual cis men N_5 = 24 heterosexual cis men
Leavitt & Berger 1990	United States	cross-sectional	N_1 = 36 sexually inactive trans women N_2 = 30 trans women

			receiving pleasure from their penis N_3 = 15 trans women avoiding their penis during sex
Lindgren & Pauly 1975	United States	cross-sectional	N_1 = 16 trans women N_2 = 16 trans men
Michel *et al.* 2002	Belgium	comparison with control-group	N_1 = 16 trans female patients N_2 = 13 cis male patients
Roback *et al.* 1976	United States	comparison with control-groups	N_1 = 10 trans men N_2 = 10 cis women applying for bypass surgery N_3 = 10 cis female outpatients
Singh *et al.* 2010	Canada	comparison with control-groups	Study 1 N_1 = 44 adolescents with GID N_2 = 98 adolescents without GID Study 2 N_1 = 41 adults with GID N_2 = 94 adults without GID
Sjoberg, Walch, & Stanny 2006	United States	cross-sectional	N = 108 trans women
Van de Grift *et al.* 2016a	The Netherlands Belgium Germany Norway	cross-sectional	N_1 = 374 trans women N_2 = 286 trans men
Van de Grift *et al.* 2016b	The Netherlands Belgium Germany Norway	cross-sectional	N_1 = 308 trans women N_2 = 77 trans men
van de Grift *et al.* 2016c	The Netherlands	longitudinal comparison with control-group	N_1 = 33 trans women N_2 = 99 cis adults
Wylie *et al.* 2010	United States	cross-sectional	N = 82 gender non-conforming young adults (18 – 30 years)

Table 2. Overview of assessment instruments supported in the literature.

Assessment instrument	Studies supporting instrument	Studies contradicting instrument
Minnesota Multiphasic Personality Inventory-2 or previous version	• Caron & Archer 1997 • Gómez-Gil, Vidal-Hagemeijer, & Salamero 2008 • Keo-Meier *et al.* 2015	/

	• Langevin, Paitich, Steiner 1977	
	• Leavitt & Berger 1990	
	• Michel *et al.* 2002	
	• Roback *et al.* 1976	
Bem Sex-Role Inventory	• Factor & Rothblum 2017	• Gómez-Gil *et al.* 2012
	• Herman-Jeglińska, Grabowska, & Dulko 2002	
	• Leavitt & Berger 1990	
Body Image Scale for Transsexuals	• Lindgren & Pauly 1975	/
	• van de Grift *et al.* 2016a	
	• van de Grift *et al.* 2016b	
	• van de Grift *et al.* 2016c	
Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults	• Deogracias *et al.* 2007	/
	• (Docter & Fleming 1992)	
	• (Docter & Fleming 2001)	
	• Singh *et al.* 2010	
16-Personality Factors Test	/	Langevin, Paitich, Steiner 1977
Body Image Assessment Questionnaire	• Becker *et al.* 2018	/
Cognitive Interview	• Wylie *et al.* 2010	/
Draw-A-Person Test	• Leavitt & Berger 1990	/
Rorschach Inkblot Test	• Caron & Archer 1997	/
Transgender Adaptation and Integration Measure	• Sjoberg, Walch, & Stanny 2006	/

Note. The four assessment instruments with the most support of 'good' studies are listed first, ranked following the amount of studies. The remaining assessment instruments are listed alphabetically.