

Measuring Gender Dysphoria: A Multicenter Examination and Comparison of the Utrecht Gender Dysphoria Scale and the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults

Catharina Schneider¹ · Susanne Cerwenka¹ · Timo O. Nieder¹ · Peer Briken¹ · Peggy T. Cohen-Kettenis² · Griet De Cuypere³ · Ira R. Haraldsen⁴ · Baudewijntje P. C. Kreukels² · Hertha Richter-Appelt¹

Received: 4 May 2013 / Revised: 19 January 2016 / Accepted: 21 January 2016 / Published online: 16 February 2016
© Springer Science+Business Media New York 2016

Abstract This study examined two instruments measuring gender dysphoria within the multicenter study of the European Network for the Investigation of Gender Incongruence (ENIGI). The Utrecht Gender Dysphoria Scale (UGDS) and the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults (GIDYQ-AA) were examined for their definitions of gender dysphoria and their psychometric properties, and evaluated for their congruence in assessing the construct. The sample of 318 participants consisted of 178 male-to-females (MtF) and 140 female-to-males (FtM) who were recruited from the four ENIGI gender clinics. Both instruments were significantly correlated in the group of MtFs. For the FtM group, there was a trend in the same direction but smaller. Gender dysphoria was found to be defined differently in the two instruments, which led to slightly different findings regarding the subgroups. The UGDS detected a difference between the subgroups of early and late onset of gender identity disorder in the group of MtFs, whereas the GIDYQ-AA did not. For the FtM group, no significant effect of age of onset was found. Therefore, both instruments seem to capture not only similar but also different aspects of gender dysphoria. The UGDS

focusses on bodily aspects, gender identity, and gender role, while the GIDYQ-AA addresses subjective, somatic, social, and sociological aspects. For future research, consistency in theory and definition of gender dysphoria is needed and should be in line with the DSM-5 diagnosis of gender dysphoria in adolescents and adults.

Keywords Gender dysphoria · Transsexualism · Utrecht Gender Dysphoria Scale · Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults · DSM-5

Introduction

In the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), gender dysphoria in adolescents and adults has been chosen as the new name of the former diagnosis gender identity disorder (GID) (or transsexualism in ICD-10). It refers to “distress that may accompany the incongruence between one’s experienced or expressed gender and one’s assigned gender” (American Psychiatric Association, 2013, p. 451). It is a more descriptive term than the previously used gender identity disorder of DSM-IV-TR (American Psychiatric Association, 2000) and focuses explicitly on dysphoria as the clinical problem.

This development was accompanied by changes in the health-care system, which are reflected in the latest version of the Standards of Care of the World Professional Association for Transgender Health (WPATH) with a shift from identity-based to distress-based healthcare (Coleman et al., 2011). Formerly, practitioners had to treat people with the diagnosis gender identity disorder and were thus in the uncomfortable position to pathologize one’s identity although they were predominantly treating the body. Now treatment more generally focusses on distress caused by the discrepancy between one’s identity and one’s

✉ Hertha Richter-Appelt
hrichter@uke.de

¹ Department of Sex Research and Forensic Psychiatry, Center for Psychosocial Medicine, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, 20246 Hamburg, Germany

² Department of Medical Psychology, VU University Medical Center Amsterdam, Amsterdam, The Netherlands

³ Department of Sexology and Gender Problems, University Hospital Ghent, Ghent, Belgium

⁴ Department of Neuropsychiatry and Psychosomatic Medicine, Rikshospitalet Oslo, Oslo, Norway

bodily aspects and the goal of a multimodal treatment lies in the reduction of gender dysphoria (Nieder & Richter-Appelt, 2012). But how is gender dysphoria defined within the instruments that are used to measure it, which component does it consist of, and how is it operationalized?

This study aimed at describing the conceptual issues and at examining the applicability of two of the most frequently used questionnaires for gender dysphoria within the European multicenter study of the European Network for the Investigation of Gender Incongruence, (ENIGI) (see Kreukels et al., 2012). The two measures of gender dysphoria, the Utrecht Gender Dysphoria Scale (UGDS) (Cohen-Kettenis & van Goozen, 1997) and the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults (GIDYQ-AA) (Deogracias et al., 2007), were compared in three steps. Firstly, the two instruments were compared in respect of their definitions and conceptions of gender dysphoria, secondly, concerning their psychometric properties, and thirdly, with regard to the question of whether the two scales produced similar patterns of group differences within a sample of applicants to the four European clinics within the ENIGI project.

Gender dysphoria was introduced by Fisk (1973) for descriptive and communicational purposes. Gender dysphoria (syndrome) was conceptualized more broadly than transsexualism and it aimed to describe in a dimensional manner the dissatisfaction, distress, anxiety, or discomfort with one's gender, ranging from a non-pathological pole to transsexualism at the other end, which was described as the most extreme form of gender dysphoria (Fisk, 1974).

In subsequent years, the term gender dysphoria became more and more accepted in the clinical context and questionnaires were developed for its measurement. Some researchers in the field followed Fisk's point of view (see, for example, Kuiper & Cohen-Kettenis, 1988) and developed the UGDS (Cohen-Kettenis & van Goozen, 1997): "*Gender Dysphoria* refers to distress caused by discrepancy between sense of self (gender identity) and the aspects of the body, [which are] associated with sex/gender, other people's misidentification of one's gender, and the social roles associated with gender" (de Vries, Cohen-Kettenis, & Delemarre-van de Waal, 2006, p. 83).

Others stated that gender dysphoria was subjective distress with one's gender identity and described it as a continuum with two poles, namely (unproblematic) gender identity and gender dysphoria based on a bi-gender system. Against this background, Deogracias et al. (2007) developed the GIDYQ-AA: "We conceptualized gender identity/gender dysphoria as a bipolar continuum with a male pole and a female pole and varying degrees of gender dysphoria, gender uncertainty, and gender identity transitions between the poles" (Deogracias et al., 2007, p. 371).

It should be stressed that these two definitions, although very similar, are not the same since they are based on different underlying continua. Questions on the UGDS focus on dissat-

isfaction with bodily aspects, gender identity, and gender roles (e.g., *I feel unhappy because I have to behave like a girl* [female to male (FtM)] or *Every time someone treats me like a boy I feel hurt* [male to female (MtF)]). In the GIDYQ-AA, gender dysphoria seems to be the problematic or pathological counterpart of gender identity itself, and an effort was made to capture subjective, somatic, social, and sociolegal aspects (e.g., *In the past 12 months, have you felt uncertain about your gender, that is, feeling somewhere in between a woman and a man?* or *In the past 12 months, at home, have you dressed and acted as a man/woman?*). Therefore, although both questionnaires are designed to measure the degree of gender dysphoria a person is struggling with, they probably will do so in a slightly different manner since each instrument captures only some aspects of the construct.

After comparing the definitions underlying the two questionnaires, the main difference seems to be the continua under which gender dysphoria is measured dimensionally. For the UGDS, the poles range from *not dysphoric* to *dysphoric* (see, for example, the follow-up study of Smith, van Goozen, Kuiper, & Cohen-Kettenis, 2005). For the GIDYQ-AA, they range from an unproblematic male/female gender identity to gender dysphoria (Deogracias et al., 2007). Both instruments also differ in terms of number of items (12 within UGDS; 27 within GIDYQ-AA). Moreover, the GIDYQ-AA items are phrased parallel for MtF and FtM and narrowed to a 12-month period, while the UGDS items are not. The fact that the GIDYQ-AA uses a time frame while the UGDS does not could lead to different outcomes for certain subgroups.

Traditionally, gender dysphoric people were classified as MtF and FtM as well as homosexual versus non-homosexual. In recent times, another differentiation was made between persons with an early onset (EO) of gender dysphoria, which means an onset in childhood, and individuals with a late onset (LO), which stands for an onset in or after puberty (Nieder et al., 2011). Using EO and LO as subgroups, one would predict different outcomes when using a time frame of lifetime than when using a 12-month time period. More precisely, one could expect that persons with a late onset would score lower on gender dysphoria than persons with an early onset in the lifetime frame, because the former do not look back on a lifelong experience of gender dysphoria, while people with an early onset do. On the other hand, persons with LO are usually older at first application (Cerwenka, Nieder, & Richter-Appelt, 2012; Nieder et al., 2011) and with age it is assumed that coping strategies change (Aldwin, Sutton, Chiara, & Spiro, 1996; Amirkhan & Auyeung, 2007; Diel, Coyle, & Labouvie-Vief, 1996), which could be a reason why the distress might not be as strong for persons with LO than for persons with EO (distress represented by gender dysphoria).

Cohen-Kettenis and Pfäfflin (2010) described psychometric properties of the UGDS and the GIDYQ-AA. For the latter, a Cronbach's alpha of .97 was found. For the UGDS, it was .66–.80 in one sample and .78–.92 in another. Since lower alphas were only found within control subjects, Cohen-Kettenis and

Pfäfflin assumed that they may be related to the lower variability of gender dysphoria within this group. They also reported good discriminant validity, comparing adolescents and adults with and without a GID diagnosis.

After comparing definitions and psychometric properties, both instruments were statistically analyzed and comparisons between the subgroups were made. General research questions are whether the two scales produce similar patterns of group differences and whether they show similar correlation patterns across subgroups. To investigate these questions, the two scales were correlated and compared in terms of the outcomes they produced. Since both instruments are supposed to measure the same construct, they should correlate to a high degree. Note that the UGDS uses a sum score with high scores representing strong gender dysphoria, while the GIDYQ-AA uses a mean score with low scores representing gender dysphoria; therefore, the UGDS has been converted into a mean score, while the GIDYQ-AA was reversed scored in order to produce consistent comparability.

Method

Participants

All applicants presented themselves between January 2007 and October 2011 in one of the four clinics of the ENIGI project ($N = 776$). The ENIGI initiative is a collaborative study of the gender identity clinics of Amsterdam, Ghent, Hamburg, and Oslo. All clinics use the same diagnostic protocol and a standard assessment battery (see Kreukels et al., 2012).

Individuals who met the diagnostic criteria for GID according to DSM-IV-TR, who could be classified as either EO or LO (Nieder et al., 2011), and who did not have any medical gender-confirming treatment (e.g., hormones, gender reassignment surgery) yet were included in the present study ($N = 380$). In the

sampling process, two persons (0.03 %) were under 17 years of age, 127 (16.37 %) were excluded since they had not undergone the entire diagnostic procedure, 86 (11.08 %) could not be diagnosed as GID, 87 (11.21 %) could not be classified as either EO or LO, and 94 (12.11 %) had already undergone some kind of medical treatment. Due to missing values in the biographical information (10 missings) and/or the UGDS and GIDYQ-AA (52 missings), the final sample size consisted of 318 individuals (for more details, see Table 1).

The sample consisted of 178 MtFs and 140 FtMs, a ratio of 1.27:1 (MtF:FtM). Of these participants, 25.2 % ($n = 80$) were from Ghent, 18.6 % ($n = 59$) from Hamburg, 46.5 % ($n = 148$) from Amsterdam, and 9.7 % ($n = 31$) from Oslo. While the overall sex ratio was almost equal, they differed considerably between the countries, $\chi^2(3) = 30.5$, $p < .001$ (see Table 2).

MtFs had a more balanced proportion of EO and LO individuals, while FtMs mostly had an EO GID, $\chi^2(1) = 38.55$, $p < .001$. At first clinical presentation, applicants were on average 29 years old with a range from 17 to 70 years. Age at clinical presentation differed significantly between the sexes. Since the Levene-test was significant ($p < .001$), a t test for heterogeneous variances was calculated, $t(314.09) = 5.99$, $p < .001$, revealing that FtMs ($M = 26.75$, $SD = 9.06$) were younger than MtFs ($M = 33.88$, $SD = 12.11$). MtFs (25.6 %, $n = 45$) more often had a higher educational level compared to FtMs (13.7 %, $n = 19$), $\chi^2(2) = 6.98$, $p = .031$. Sexual orientation was distributed differently between the groups. FtMs were more often attracted to women than MtFs were to men, $\chi^2(1) = 72.80$, $p < .001$ (for more details, see Table 2).

Measures

For information on whether applicants met the criteria of GID according to the DSM-IV-TR criteria, standardized evaluation sheets, constructed within the ENIGI initiative, were completed

Table 1 Sample attrition

| | Total | % | FtM | % | MtF | % |
|--------------------------------------|-------|-------|-----|-------|-----|-------|
| Total | 776 | 100 | 280 | 36 | 496 | 64 |
| Under 17 years | 2 | 0.03 | 2 | 0.07 | 0 | 0 |
| Evaluation missing ^a | 127 | 16.37 | 41 | 14.64 | 86 | 17.34 |
| No GID Diagnosis | 86 | 11.08 | 15 | 5.36 | 71 | 14.32 |
| Residual group/not classified | 87 | 11.21 | 25 | 8.93 | 62 | 12.50 |
| Medical treatment (hormones/surgery) | 94 | 12.11 | 25 | 8.93 | 69 | 13.91 |
| Total | 380 | 100 | 172 | 45.3 | 208 | 54.7 |
| Missing Information and/or | 30 | 7.89 | | | | |
| Missing UGDS/GIDYQ-AA | 52 | 13.68 | | | | |
| Total | 318 | 100 | 140 | 44 | 178 | 56 |

FtM female to male, *MtF* male to female, *GID* gender identity disorder, *UGDS* Utrecht Gender Dysphoria Scale, *GIDYQ-AA* Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults

^a Participants who had not undergone the entire diagnostic process

Table 2 Sample characteristics

| Variable | | FtM (%) | MtF (%) | Total (%) |
|---|-----------------|-------------------|-------------------|-------------------|
| Country | Belgium | 26 (18.6) | 54 (30.3) | 80 (25.2) |
| | Germany | 33 (23.6) | 26 (14.6) | 59 (18.6) |
| | The Netherlands | 55 (39.3) | 93 (52.2) | 148 (46.5) |
| | Norway | 26 (18.6) | 5 (2.8) | 31 (9.7) |
| Age (in years) | <i>M</i> | 26.75 | 33.88 | 30.73 |
| | <i>SD</i> | 90.63 | 12.11 | 11.419 |
| | Median | 24 | 32 | 29 |
| | Range | 17–60 | 17–70 | 17–70 |
| Missings | | 0 | 1 | 1 |
| Diagnosis | <i>N</i> | (<i>n</i> = 140) | (<i>n</i> = 178) | (<i>n</i> = 318) |
| Early onset | | 120 (85.7) | 94 (52.8) | 214 (67.3) |
| Late onset | | 20 (14.3) | 84 (47.2) | 104 (32.7) |
| Education | <i>n</i> | (<i>n</i> = 139) | (<i>n</i> = 176) | (<i>n</i> = 315) |
| Low ^a | | 31 (22.3) | 37 (21.0) | 68 (21.0) |
| Middle ^b | | 89 (64.0) | 94 (53.4) | 183 (58.1) |
| High ^c | | 19 (13.7) | 45 (25.6) | 64 (20.3) |
| Sexual orientation | <i>n</i> | (<i>n</i> = 139) | (<i>n</i> = 176) | (<i>n</i> = 315) |
| Gynephilic ^d /androphilic ^e | | 115 (82.7) | 61 (34.7) | 176 (55.3) |
| Notgynephilic/androphilic | | 24 (17.3) | 115 (65.3) | 139 (44.7) |

FtM female to male, *MtF* male to female

^a Low = lower education, lower vocational

^b Middle = secondary education, secondary vocational and high school

^c High = higher vocational, bachelor, master & PhD

^d Gynephilic = refers to FtM attracted exclusively to women

^e Androphilic = refers to MtF attracted exclusively to men

by the clinicians at the end of the diagnostic phase (Kreukels et al., 2012; Paap et al., 2011). Analogous to the approach by Nieder et al. (2011), individuals were classified as either having an EO or LO GID by retrospectively evaluating DSM-IV-TR criteria for GID in childhood. Using the Kinsey Homosexual–Heterosexual Rating Scale (Kinsey, Pomeroy, & Martin, 1948), participants were also asked to rate their sexual preference for men or women on a 7-point scale from exclusively heterosexual (Kinsey 0) to exclusively homosexual (Kinsey 6). In addition, they had the opportunity to state little or no sexual attraction at all (Kinsey X). In order to facilitate comparisons with other studies, the approach of Lawrence (2005) was adopted and all participants were assigned to one of four categories, indicating whether they felt sexually attracted to females (gynephilic), males (androphilic), both (bisexual) or neither (asexual).

Demographic information was collected at the point of first clinical presentation via an adaption of the Dutch Biographic Questionnaire on Transsexualism (Doorn, Poortinga, & Verschoor, 1994), constructed for the ENIGI Initiative.

To measure the degree of gender dysphoria, two different questionnaires were used. Firstly, the UGDS (Cohen-Kettenis & van Goozen, 1997) which consists of 12 items, to be answered on

a 1–5 point scale, resulting in a sum score between 12 and 60. The higher the sum score, the stronger the gender dysphoria. Items were developed for FtM and MtF separately, resulting in, e.g., *I wish I had been born as a boy* or *I hate having breasts* for FtM and *Only as a girl my life would be worth living* or *I hate having erections* for MtF (see also de Vries et al., 2006).

Cohen-Kettenis and van Goozen (1997) reported Cronbach's alpha to be .92 for male and .78 for female applicants to their program. Discriminant validity was reported to be excellent between transsexual and non-transsexual individuals ($p < .001$) as well as between applicants for gender-confirming surgery who were or were not referred for treatment ($p < .001$). In a recent validation study by Steensma et al. (2013), Cronbach's alphas of .98 were reported for both the male and female version of the instrument. Using a cut-point of 40, they reported a sensitivity of 88.3 % for clinically referred participants and a specificity of 99.5 % for controls for the male version. For the female version, sensitivity was reported to be 98.5 % and specificity 97.9 %.

The GIDYQ-AA (Deogracias et al., 2007) measures the degree to which an individual struggles with his/her gender identity. It contains 27 items in analogous versions for MtF and FtM.

Making an effort to capture different indicators of gender dysphoria items are, for example, *In the past 12 months, have you felt unhappy about being a woman/man?* (subjective indicator), *In the past 12 months, have you felt pressured by others to be a man/woman, although you don't really feel like one?* (social indicator), *In the past 12 months, have you wished to have hormone treatment to change your body into a man's/woman's?* (somatic indicator), or *In the past 12 months, have you felt bothered by seeing yourself identified as male/female or having to check the box "M" for male/"F" for female on official forms (e.g. employment applications, driver's licence, passport)?* (sociological indicator). Mean item scores between 1 and 5 are calculated. The cut-off is 3. Scores below 3 have been used to signify case-ness for gender dysphoria. Sensitivity was reported to be 90.4 % for gender identity patients and specificity was 99.7 % for controls (Singh et al., 2010).

Standardized instruments (questionnaires, interviews, and evaluation sheets) and written informed consent was obtained from all participants. The study was approved by the ethics committees of all four clinics. Applicants 17 years of age and older were asked to participate in the study. Individuals with insufficient command of the local language or with an acute psychotic disorder were excluded.

Statistical Analysis

For statistical analysis, SPSS 19.0 was used. Group differences and relationships between variables on a nominal data level were investigated via χ^2 tests. In case the χ^2 tests did not fit the data, Fisher's exact probability test was used. *t* tests and analysis of variance (ANOVA) were performed to compare means. After recoding the GIDYQ-AA, so that higher scores signify more gender dysphoria (comparable to the UGDS), the UGDS was transformed into mean scores, so both scales show comparable values (absolute range 1–5). Missing values were replaced by mean values for participants who answered more than 85 % of all items.

Results

Table 3 shows the mean score on the UGDS and GIDYQ-AA as a function of group (MtF vs. FtM participants). A 2 (Group: MtF vs. FtM) \times 2 (Scale: UGDS vs. GIDYQ-AA) ANOVA with Scale as a within-subjects factor yielded significant main effects for Group, $F(1, 316) = 66.50, p < .001, \eta^2 = .17$, and Scale $F(1, 316) = 1300.10, p < .001, \eta^2 = .80$, as well as a significant Group \times Scale interaction $F(1, 316) = 42.54, p < .001, \eta^2 = .12$. On both measures, the FtMs were more gender dysphoric than the MtFs. The UGDS yielded higher gender dysphoric scores than the GIDYQ-AA. The interaction was primarily accounted for by the FtMs being more gender dysphoric ($M = 4.74, SD = 0.25$) than the MtFs ($M = 4.32,$

$SD = 0.50$) on the UGDS, $t(316) = -8.99, p < .001, d = 1.03$, than on the GIDYQ-AA (FtM: $M = 3.76, SD = 0.27$; MtF: $M = 3.64, SD = 0.30$), $t(316) = -3.67, p < .001, d = 0.42$.

We next conducted a 2 (Group: MtF vs. FtM) \times 2 (Age of Onset: Early vs. Late) \times 2 (Scale: UGDS vs. GIDYQ-AA) ANOVA with Scale as within-subjects factor, which showed significant main effects for Group, $F(1, 314) = 48.78, p < .001, \eta^2 = .13$, and Scale, $F(1, 314) = 817.66, p < .001, \eta^2 = .72$, but not for Age of Onset, $F(1, 314) = 1.70, p = .19, \eta^2 = .01$. The interaction between Group and Age of Onset was significant $F(1, 314) = 4.78, p < .05, \eta^2 = .02$, but the interactions between Scale and Age of Onset, $F(1, 314) = 1.59, p =$

Table 3 Means and SDs of UGDS (mean) and GIDYQ-AA (reversed), and combined scores

| Questionnaire | Groups | N | M | SD |
|------------------------------|--------|-----|------|------|
| UGDS | MtF | 178 | 4.32 | 0.50 |
| | FtM | 140 | 4.74 | 0.25 |
| | EO | 214 | 4.59 | 0.42 |
| | LO | 104 | 4.32 | 0.49 |
| | MtF | | | |
| | EO | 94 | 4.42 | 0.51 |
| | LO | 84 | 4.21 | 0.47 |
| | FtM | | | |
| | EO | 120 | 4.73 | 0.26 |
| | LO | 20 | 4.76 | 0.22 |
| GIDYQ-AA | MtF | 178 | 3.64 | 0.30 |
| | FtM | 140 | 3.76 | 0.27 |
| | EO | 214 | 3.72 | 0.29 |
| | LO | 104 | 3.34 | 0.29 |
| | MtF | | | |
| | EO | 94 | 3.68 | 0.30 |
| | LO | 84 | 3.60 | 0.27 |
| | FtM | | | |
| | EO | 120 | 3.76 | 0.28 |
| | LO | 20 | 3.80 | 0.22 |
| Mean of UGDS and GIDYQ-AA | MtF | 178 | 3.98 | 0.34 |
| | FtM | 140 | 4.25 | 0.21 |
| | EO | 214 | 4.16 | 0.30 |
| | LO | 104 | 3.98 | 0.33 |
| | MtF | | | |
| | EO | 94 | 4.05 | 0.35 |
| | LO | 84 | 3.91 | 0.32 |
| | FtM | | | |
| | EO | 120 | 4.25 | 0.21 |
| | LO | 20 | 4.28 | 0.17 |

FtM female to male, *MtF* male to female, *EO* early onset, *LO* late onset, *UGDS* Utrecht Gender Dysphoria Scale (absolute range 1–5, higher scores indicate stronger dysphoria), *GIDYQ-AA* Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults (absolute range 1–5; higher scores indicate stronger dysphoria)

.20, $\eta^2 = .01$, and Group, Scale, and Age of Onset were not, $F(1, 314) = 0.96, p = .33, \eta^2 = .00$. The Group \times Age of Onset interaction was decomposed by a series of *t* tests. Both EO and LO FtMs ($M = 4.25, SD = 0.21$) were significantly more gender dysphoric than the EO and LO MtFs ($M = 3.98, SD = 0.34$), $t(316) = -8.16, p < .001, d = 0.95$. The EO MtFs ($M = 4.05, SD = 0.35$) were more gender dysphoric than the LO MtFs ($M = 3.91, SD = 0.32$), $t(176) = 2.81, p = .005, d = -0.42$, but the difference between the EO and LO FtMs was not significant, $t(138) = -.73, p = .47, d = 0.19$.

A significant correlation between UGDS and GIDYQ-AA was found in the group of MtFs, $r(176) = .44, p < .001$, indicating that a stronger rejection of one's body and cross-gender identification (UGDS) goes along with stronger problems with one's gender identity (GIDYQ-AA). For FtMs, there was a trend in the same direction as for MtFs, but smaller, $r(138) = .22, p < .05$.

Discussion

This study set out to explore two commonly used measures for gender dysphoria in a multicenter study and compare their outcomes in terms of group differences between MtF and FtM as well as EO and LO. After comparing the theoretical constructs underlying the two questionnaires, we examined the outcomes the two measures produced and the correlations between them.

The first ANOVA revealed two main effects, for Group (MtF vs. FtM) and Scale (UGDS vs. GIDYQ-AA), as well as an interaction. The most prominent difference was the one between the two scales, revealing that the UGDS showed generally higher scores than the GIDYQ-AA. This is hard to interpret since the interaction between both factors was also significant. As has been described, both scales differed in item number and time frame. Maybe the time frame has an influence. Individuals could describe their feelings on gender dysphoria as being more severe when not being reminded of a longer period of time that they should relate their feelings to. The items themselves can also have an influence. Also, the difference between FtM and MtF can have a part in this result. Since the UGDS had ceiling effects within the female version, with almost one-third reaching the highest possible score of 60, this also could influence the overall difference between the two scales.

The second finding was the main effect on Group. It was found that FtMs reported stronger gender dysphoria than MtFs which is in line with the literature (Cohen-Kettenis & van Goozen, 1997; Deogracias et al., 2007; Singh et al., 2010). FtMs therefore seem to report a stronger antipathy toward their own bodies, their gender identity, and their gender role (UGDS) and more distress concerning their gender identity (GIDYQ-AA). Considering the significant interaction between Group and Age of Onset, it could also be that the general difference between the groups of FtM and MtF could be due to the difference in gender

dysphoria within the groups of EO and LO. MtF with EO reported significantly higher gender dysphoria than MtF with LO, supporting a statement by de Vries et al. (2006) that MtFs with EO are more dysphoric than MtFs with LOGID. Supposing that this also is the case for FtMs, and the majority of this group has an EO (85.7%), it could be that age of onset rather than direction of transition might be responsible for the overall group difference. Since we cannot test this hypothesis for an insufficient group size of LO within FtMs ($n = 20$), this should be taken into consideration for future research.

The difference between FtMs and MtFs might also be due to ceiling effects within the female version of the UGDS. Since nearly one-third of the group of FtMs reached the highest possible score, while within the group of MtFs only 10.7% reached a score of 60, it can be concluded that ceiling effects in one version might be responsible for different outcomes within both groups. This would also lead to problems interpreting the correlation between both instruments within the FtM group, since ceiling effects can artificially induce lower correlations.

Although ceiling effects have to be taken into account, it has to be noted that even in the group of MtF the two instruments correlated only moderately. Therefore, it might be possible that the construct of gender dysphoria consists of more components than measured by the instruments. Although it could be an artifact of the homogeneity within the group of FtMs as well, the questionnaires seem to capture somewhat different aspects of gender dysphoria, as has been shown by comparing the underlying definitions of the concept.

Another point one has to keep in mind is that the two UGDS versions used for MtFs and FtMs are not exactly the same. For instance, items referring to physical aspects differ between the scales. If they are not captured similarly within the MtF and FtM version of the UGDS, this too could be responsible for different outcomes for the two groups.

Furthermore, all scores were in the dysphoric range of 50–60 (using the cut-point set by Steensma et al., 2013). This could be seen as a sign of quality, since both scales are diagnostic assessment tools designed to make categorical distinctions between persons with clinically relevant gender dysphoria and those who do not have clinically significant gender dysphoria. Their power to distinguish between inter-individual differences between subgroups, especially within homogeneous samples like this one (group of FtM), might not be very high, showing that the theoretical relevance of this question is probably higher than its clinical relevance. This again, could contribute to the small, rather than high correlations between the variables.

One more theoretical implication has to be taken into consideration. As stated before, both instruments differ in terms of the time frame used in them. While the UGDS does not use any time-related words and therefore refer to the actual moment of filling out the questionnaire, the GIDYQ-AA uses a time frame of 1 year. Assuming that a time frame has an impact on the outcome and considering our data, it can be argued that persons

with an LO of GID score lower than persons with EO under no time frame, since their gender dysphoria may be more familiar to them and they may have already developed coping strategies, while under a 12-month time frame groups do not differ at all.

In summary, the two instruments (UGDS and GIDYQ-AA) differ in measuring gender dysphoria. Although both fulfill their purpose to distinguish between clinical and subclinical groups of gender dysphoric individuals, they could do a better job in capturing the construct more similarly. Therefore, it might enhance them to exclude items with little discriminatory power and thus to establish short versions. In terms of a time frame they could be adapted, too, so they could be used interchangeably.

Several limitations of this study have to be considered. Firstly, since this study is a clinic-based study, the sample size is relatively large but not necessarily representative for all gender incongruent or gender dysphoric persons. Therefore, no generalization can be drawn for gender variant persons who do not seek medical service or who seek it apart from official clinics (as in the black market or abroad). Findings from this sample can be generalized only for a group of individuals that apply at gender clinics, who have had no medical interventions and who can be diagnosed with GID and classified either EO or LO. Moreover, although the sample as a whole may be large, some subgroups (e.g., FtM-LO) are not, diminishing informative value for these groups. Secondly, biases in sample selection could have had an impact on the results. Also, differences between countries could not be considered within the present study, so that specifics in clinical procedures could not be analyzed. Thirdly, since they were excluded from the sample no conclusions can be drawn for applicants who did not fulfill the criteria for GID, who had already undergone any medical treatment and who could not be classified as either EO or LO.

Furthermore, intending to compare the two scales UGDS and GIDYQ-AA in their ability to capture gender dysphoria, one major aspect is missing. Since no non-dysphoric control group existed, there was no option to check whether both instruments fulfill their purpose to discriminate between persons with gender dysphoria and non-dysphoric persons to the same degree.

Moreover, while in MtFs early and late onset showed several distinctions, the same was not true for FtMs. One reason for this is certainly the small number of late onsets in the group of FtMs. Whether or not this subgroup holds specific characteristics and needs could be an interesting future field of study. Despite different levels of heterogeneity within the two groups, it might also be due to the times when some of the scales were constructed that the evaluated instruments provided more information about MtFs than FtMs. Typologies about MtFs are longer established (Benjamin, 1966; Blanchard, 1989; Blanchard, Clemmensen, & Steiner, 1987) than for FtMs, for whom, only a few years ago, a classification was thought to be redundant (see Lawrence, 2010). It might be as well that this reflects a societal bias that leads to different visibility, rules of passing, and accepted behavior for MtFs and FtMs which again results in differences

between the two groups. Thus, it would be interesting, to study transgender issues in account to social, cultural, and historical factors associated with gender dysphoria.

Acknowledgments We are most grateful to all persons who made this research possible, especially the participants of this study. We would also like to thank the collaborating medical and psychosocial staff and colleagues of all four clinics in Belgium, in the Netherlands, in Norway, and in Hamburg. Further, we would like to especially thank the Editor, for his thoughtful comments and his input regarding the statistical analyses. The first author is especially grateful for his patience and support in the process.

Appendix

Utrecht Gender Dysphoria Scale

Female-to-Male Version

Response categories are agree completely, agree somewhat, neutral, disagree somewhat, disagree completely. Items 1, 2, 4–6, and 10–12 are scored from 5 to 1; Items 3 and 7–9 are scored from 1 to 5.

- (1) I prefer to behave like a boy.
- (2) Every time someone treats me like a girl I feel hurt.
- (3) I love to live as a girl.
- (4) I continuously want to be treated like a boy.
- (5) A boy's life is more attractive for me than a girl's life.
- (6) I feel unhappy because I have to behave like a girl.
- (7) Living as a girl is something positive for me.
- (8) I enjoy seeing my naked body in the mirror.
- (9) I like to behave sexually as a girl.
- (10) I hate menstruating because it makes me feel like a girl.
- (11) I hate having breasts.
- (12) I wish I had been born as a boy.

Male-to-Female Version

Response categories are agree completely, agree somewhat, neutral, disagree somewhat, disagree completely. Items are all scored from 5 to 1.

- (1) My life would be meaningless if I would have to live as a boy.
- (2) Every time someone treats me like a boy I feel hurt.
- (3) I feel unhappy if someone calls me a boy.
- (4) I feel unhappy because I have a male body.
- (5) The idea that I will always be a boy gives me a sinking feeling.
- (6) I hate myself because I'm a boy.
- (7) I feel uncomfortable behaving like a boy, always and everywhere.
- (8) Only as a girl my life would be worth living.
- (9) I dislike urinating in a standing position.
- (10) I am dissatisfied with my beard growth because it makes me look like a boy.

- (11) I dislike having erections.
 (12) It would be better not to live than to live as a boy.

Scoring and Evaluation

To calculate the total score, all Items of the Male version and Items 1, 2, 4–6, and 10–12 of the female version have to be reversed scored.

References

- Aldwin, C. M., Sutton, K. J., Chiara, G., & Spiro, A. (1996). Age differences in stress, coping, and appraisal: Findings from the normative aging study. *Journal of Gerontology: Psychological Science B*, *51*, 179–188.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- Amirkhan, J., & Auyeung, B. (2007). Coping with stress across the lifespan: Absolute vs. relative changes in strategy. *Journal of Applied Developmental Psychology*, *28*, 298–317. doi:10.1016/j.appdev.2007.04.002.
- Benjamin, H. (1966). *The transsexual phenomenon: A scientific report on transsexualism and sex conversion in the human male and female*. New York: Julian.
- Blanchard, R. (1989). The classification and labeling of nonhomosexual gender dysphorias. *Archives of Sexual Behavior*, *18*, 315–334.
- Blanchard, R., Clemmensen, L. H., & Steiner, B. W. (1987). Heterosexual and homosexual gender dysphoria. *Archives of Sexual Behavior*, *16*, 139–152.
- Cerwenka, S., Nieder, T. O., & Richter-Appelt, H. (2012). Sexuelle Orientierung und Partnerwahl transsexueller Frauen und Männer vor körpermedizinischen geschlechtsanpassenden Maßnahmen [Sexual orientation and partner-choice of transsexual women and men before gender-confirming interventions]. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *62*, 214–222. doi:10.1055/s-0032-1309030.
- Cohen-Kettenis, P. T., & Pfäfflin, F. (2010). The DSM diagnostic criteria for gender identity disorder in adolescents and adults. *Archives of Sexual Behavior*, *39*, 499–513. doi:10.1007/s10508-009-9562-y.
- Cohen-Kettenis, P. T., & van Goozen, S. H. M. (1997). Sex reassignment of adolescent transsexuals: A follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*, 263–271. doi:10.1097/00004583-199702000-00017.
- Coleman, E., Adler, R., Bockting, W., Botzer, M., Brown, G., Cohen-Kettenis, P., ... Zucker, K. (2011). Standards of Care for the Health of Transsexual, Transgender, and Gender-Nonconforming People, Version 7. *International Journal of Transgenderism*, *13*, 165–232.
- de Vries, A. L. C., Cohen-Kettenis, P. T., & Delemarre-van de Waal, H. (2006). Clinical management of gender dysphoria in adolescents. *International Journal of Transgenderism*, *9*(3–4), 83–94. doi:10.1300/J485v09n03_04.
- Deogracias, J. J., Johnson, L. L., Meyer-Bahlburg, H. F. L., Kessler, S. J., Schober, J. M., & Zucker, K. J. (2007). The Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults. *Journal of Sex Research*, *44*, 370–379. doi:10.1080/00224490701586730.
- Diel, M., Coyle, N., & Labouvie-Vief, G. (1996). Age and sex differences in strategies of coping and defence across the life span. *Psychology and Aging*, *11*, 127–139.
- Doorn, C. D., Poortinga, J., & Verschoor, A. M. (1994). Cross-gender identity in transvestites and male transsexuals. *Archives of Sexual Behavior*, *23*, 185–201.
- Fisk, N. M. (1973). Gender dysphoria syndrome (The how, what, and why of a disease). In D. R. Laub & P. Gandy (Eds.), *Proceedings of the second interdisciplinary symposium on gender dysphoria syndrome* (pp. 7–14). Palo Alto, CA: Stanford University Press.
- Fisk, N. M. (1974). Gender dysphoria syndrome: The conceptualization that liberalizes indications for total gender reorientation and implies a broadly based multi-dimensional rehabilitative regimen. *Western Journal of Medicine*, *120*, 386–391.
- Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. (1948). *Sexual behavior in the human male*. Philadelphia, PA: Saunders.
- Kreukels, B. P., Haraldsen, I. R., De Cuypere, G., Richter-Appelt, H., Gijs, L., & Cohen-Kettenis, P. T. (2012). A European Network for the Investigation of Gender-Incongruence: The ENIGI Initiative. *European Psychiatry*, *27*, 445–450.
- Kuiper, A. J., & Cohen-Kettenis, P. T. (1988). Sex reassignment surgery: A study of 141 Dutch transsexuals. *Archives of Sexual Behavior*, *17*, 439–457. doi:10.1007/BF01542484.
- Lawrence, A. A. (2005). Sexuality before and after male-to-female sex reassignment surgery. *Archives of Sexual Behavior*, *34*, 147–166.
- Lawrence, A. A. (2010). Sexual orientation versus age of onset as bases for typologies (subtypes) for gender identity disorder in adolescents and adults. *Archives of Sexual Behavior*, *39*, 514–545. doi:10.1007/s10508-009-9594-3.
- Nieder, T. O., Herff, M., Cerwenka, S., Preuss, W. F., Cohen-Kettenis, P. T., De Cuypere, G., ... Richter-Appelt, H. (2011). Age of onset and sexual orientation in transsexual males and females. *Journal of Sexual Medicine*, *8*, 783–791. doi:10.1111/j.1743-6109.2010.02142.x.
- Nieder, T. O., & Richter-Appelt, H. (2012). Transsexualität und Geschlechtsdysphorie [Transsexualism and gender dysphoria]. *Gynäkologie und Gynäkologische Onkologie*, *8*, 60–71.
- Paap, M. C. S., Kreukels, B. P. C., Cohen-Kettenis, P. T., Richter-Appelt, H., de Cuypere, G., & Haraldsen, I. R. (2011). Assessing the utility of diagnostic criteria: A multisite study on gender identity disorder. *Journal of Sexual Medicine*, *8*, 180–190. doi:10.1111/j.1743-6109.2010.02066.x.
- Singh, D., Deogracias, J. J., Johnson, L. L., Bradley, S. J., Kibblewhite, S. J., Owen-Anderson, A., ... Zucker, K. J. (2010). The Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults: Further validity evidence. *Journal of Sex Research*, *47*, 49–58. doi:10.1080/00224490902898728.
- Smith, Y. L. S., van Goozen, S. H. M., Kuiper, A. J., & Cohen-Kettenis, P. T. (2005). Sex reassignment: Outcomes and predictors of treatment for adolescent and adult transsexuals. *Psychological Medicine*, *35*, 89–99. doi:10.1017/S0033291704002776.
- Steenma, T. D., Kreukels, B. P. C., Jürgensen, M., Thyen, U., de Vries, A. L. C., & Cohen-Kettenis, P. T. (2013). *The Utrecht Gender Dysphoria Scale: A validation study*. Manuscript submitted for publication.