Taylor & Francis
Taylor & Francis Group

Check for updates

Utrecht Gender Dysphoria Scale - Gender Spectrum (UGDS-GS): Construct validity among transgender, nonbinary, and LGBQ samples

Jenifer K. McGuire[a], Dianne Berg[a], Jory M. Catalpa[a], Quin J. Morrow[a], Jessica N. Fish[b], G. Nic Rider[a] (iD), Thomas Steensma[c], Peggy T. Cohen-Kettenis[c] and Katherine Spencer[a]

[a]Family Social Science, University of Minnesota, Minneapolis, MN, USA; [b]Family Science, University of Maryland, College Park, MD, USA; [c]VU University Medical Center, Medical Psychology, Center of Expertise on Gender Dysphoria, Free University Netherlands, Amsterdam, The Netherlands

ABSTRACT

Background: Researchers combined both versions of the original Utrecht Gender Dysphoria Scale (UGDS) to create a single gender spectrum version (UGDS-GS) which measures dissatisfaction with gender identity and expression over time as well as comfort with affirmed gender identity.

Aim: This study examined the construct validity of the newly revised, UGDS-GS.

Method: Tests of measurement invariance were conducted in stages to assess measurement invariance of the UGDS-GS across three groups: cisgender, binary transgender, and nonbinary/genderqueer.

Results: Findings indicate that the UGDS-GS functions acceptably in all three gender groups (configural and metric invariance). Also, across binary transgender and nonbinary/genderqueer groups, the measure functions very similarly with all four types of invariance. Item level findings highlight the specificity of the measure to distinguish experiences of binary transgender and nonbinary/genderqueer persons differently from cisgender LGBQ individuals.

Conclusions: The UGDS-GS demonstrates a large degree of invariance across binary transgender, nonbinary/genderqueer, and cisgender LGBQ subgroups; and therefore, findings indicate this revision to be a substantial improvement. This 18-item self-report, Likert-type scale measure is a) inclusive of all gender identities and expressions (e.g., transfeminine spectrum, transmasculine spectrum, genderqueer, nonbinary, cisgender); b) appropriate for use longitudinally from adolescence to adulthood; and c) administered at any point in the social or medical transition process, if applicable, or in community-based research focused on gender dysphoria that examines cisgender and transgender persons.

The Utrecht Gender Dysphoria Scale (UGDS) is a validated, 12-item screening measure for both adults and adolescents used extensively in gender clinics to assess gender dysphoria (Cohen-Kettenis & van Goozen, 1997; Steensma et al., 2013). The scale is offered as part of a standard battery of diagnostic questionnaires in a variety of international gender clinics, including in Amsterdam, Ghent, Hamburg, and Oslo, and is often used in longitudinal studies that investigate gender dysphoria and other psychological outcomes for transgender people. For example, the UGDS has been used to follow up on childhood gender dysphoria clinical referrals when they reach adolescence (e.g., Wallien & Cohen-Kettenis, 2008), and to track change in gender dysphoria after puberty suppression, hormone therapy, and gender-affirming surgeries (e.g., De Vries, McGuire, Steensma, Wagenaar, Doreleijers, & Cohen-Kettenis, 2014; De Vries, Steensma, Doreleijers, & Cohen-Kettenis, 2011; Khatchadourian, Amed, & Metzger, 2014; Smith, Van Goozen, Kuiper, & Cohen-Kettenis, 2005; van de Grift et al., 2017). Researchers studying intersex populations have also used the UGDS as either a control variable (van de Grift, Cohen-Kettenis, de Vries, & Kreukels, 2018) or as a dependent variable (Jürgensen et al., 2013) in studies investigating psychosexual development, gender dysphoria, and body image.

As originally written, the UGDS has two unique forms that are administered on the basis of sex assigned at birth (Cohen-Kettenis & van Goozen, 1997; Steensma et al., 2013). These dimorphic versions for people assigned male at birth and people assigned female at birth were originally factored and normed separately, and thus have few items in common. Further, the versions attend to distinct elements of dysphoria, with differing instrumental versus affective triggers (Cohen-Kettenis & van Goozen, 1997). For example, the assigned male version contains more emotional, stereotypically feminine-coded language: "My life would be meaningless if I would have to live as a boy/man." In contrast, the assigned female version uses more pragmatic, stereotypically masculine-coded language: "I prefer to behave like a boy/man." The assigned male version contains 11 items expressing dysphoria with a gender role traditionally ascribed to boys/men, and one item expressing desire for a gender role associated with girls/women (e.g., "Only as a girl/woman my life would be worth living"). None requires reverse scoring. In contrast, the assigned female version contains four items expressing dysphoria with a gender role attributed to girls/women: "I feel unhappy because I have to behave like a girl/woman." Four items express positive feelings about a female role and are reversed for clinical scoring: "I like to behave sexually as a girl/woman." The remaining four items express desire for a male role (e.g., "I wish I had been born as a boy/man").

The dimorphic standardization of the UGDS creates measurement difficulties and applicability challenges in both clinical and community settings. For example, there is no true way to assess dysphoria that continues after a gender role change, and longitudinal studies have been limited because of measurement error introduced when switching the instrument version between instances of data collection. Additionally, transmasculine participants consistently report higher gender dysphoria than transfeminine people (Olson, Schrager, Belzer, Simons, & Clark, 2015). Because the original UGDS is composed of two scales, it is impossible to determine if this is a real difference in gender dysphoria between groups or if this is an artifact of measurement error (Steensma et al., 2013).

Furthermore, nonbinary people may not be comfortable accurately responding to either version of the survey, as they may not reference themselves against the yardstick of either male or female social norms. To address these measurement and applicability limitations, the present study seeks to describe and validate a gender-neutral, single-version adaptation of the original UGDS that retains its measure structure when administered longitudinally.

This adaptation, called the Utrecht Gender Dysphoria Scale – Gender Spectrum (UGDS-GS), measures dissatisfaction with gender identity and expression over time as well as comfort with affirmed gender identity. Researchers combined both versions of the original UGDS to create an 18-item self-report, Likert-type scale measure that is: a) inclusive of all gender identities and expressions (e.g., transfeminine spectrum, transmasculine spectrum, genderqueer, nonbinary, cisgender); b) appropriate for use longitudinally from adolescence to adulthood; and c) administered at any point in the social or medical transition process, if applicable, or in community-based research focused on gender dysphoria that examines cisgender and transgender persons. Dysphoria can fluctuate over time regardless of birth-assigned sex or engagement with gender-affirming medical interventions. Researchers used the Gender Affirming Lifespan Approach (GALA; Berg et al., 2017; Rider et al., 2019), a multidimensional, interdisciplinary, and transaffirmative conceptual model, to inform the revision process when creating the UGDS-GS. The UGDS-GS operationalizes the capacity to conceptualize gender dysphoria and expression along a spectrum or continuum, allowing for gender-related developmental differences to fluctuate across the lifespan.

Methods

Procedures

Scale development

The original UGDS is a validated, 12-item interview assessment evaluating the degree to which participants endorsed dissatisfaction with gender identity or expression over time (Cohen-Kettenis & van Goozen, 1997; Steensma et al., 2013). The original 12-item measure of gender dysphoria

used two separate versions – one designed for participants assigned male at birth and one for participants assigned female at birth. Only two items across versions were almost identical (1. Every time someone treats me like a girl/woman [or boy/man] I feel hurt, and 2. I feel uncomfortable behaving like a boy/man, always and everywhere vs I feel unhappy because I have to behave like a girl/woman. Clinicians and researchers collaborated to revise the UGDS to be more appropriate for use across a spectrum of gender identities, and to be consistent over time so that the scale remained the same before, during, or after gender-affirming interventions.

For the revision, researchers combined both versions, changing the number of items from 12 per version (a total of 24 items) to a single 20-item measure. Wording changes were made to create more contemporary language and to refrain from assumptions about an individual's current identity based on their assigned sex. For example, "my life is meaningless" was shifted to "I feel hopeless," the verb 'misgender' was included on one item, and non-gender-specific language was used for puberty and body changes. For example, "I am dissatisfied with my beard growth because it makes me look like a boy/man", "I dislike urinating in a standing position", "I dislike having erections", "I hate having breasts", and "I hate menstruating because it makes me feel like a girl/woman" were merged into the following two questions: "The bodily functions of my assigned sex are distressing for me (i.e., erection, menstruation)" and "Physical sexual development was stressful." The collaboration team chose "assigned sex" to indicate sex assigned at birth and "affirmed gender" to indicate a person's current gender identity. To prevent confusion, in the revised survey instructions, participants were told that assigned sex represented the sex the participant was assigned at birth, and affirmed gender represented the gender with which participant currently identified, without reference to the words male/man, female/woman or any other gender.

Pilot testing

Several iterations of informal and formal pilot testing occurred. As part of a psychoeducational exercise in a group therapy session, adolescent transgender and genderqueer participants and their parents provided anonymous written feedback about the initial revision of the UGDS-GS measure. One of the clinicians in the group, who is also part of the research team, shared these written responses with the leader of this research team and a research assistant, who are both authors on this paper and have experience in measurement development and qualitative research specifically within the gender diverse community. Six clinicians at the researchers' home institutions, three who have been working in gender health care for over 10 years in both a research and clinical capacity and three who were in a postdoctoral clinical training program specific to gender health care also provided feedback on the proposed measure. In particular, the clinician-researchers when using the measure clinically had been paying attention to how clients were responding to the measure's items and actively asking follow-up questions about client's perceptions of the items. Over the course of development, the measure was presented twice at gender health conferences where feedback related to ongoing analyses and revisions was solicited from clinicians and researchers with gender expertise both during audience participation and in private conversations following the presentation.

All convenience samples were collected via Amazon Mechanical Turk (MTurk) (Buhrmaster, Kwang, & Gosling, 2011), which is an integrated internet crowdsourcing tool that enables researchers to coordinate workers to complete questionnaires called HITs (Human Intelligence Tasks). Workers were compensated between one and three U.S. dollars depending on the length of the overall survey they completed. For each iteration of testing, MTURK workers completed screening items to determine eligibility and were consented into the study, following IRB approval procedures for the lead author's home institution. After completion of the survey, and review by a research assistant, funds were released into the workers' accounts. Researchers formally pilot tested the revision of the scale online with an initial, small sample of participants recruited via MTurk, including transgender, nonbinary, genderqueer ($n = 142$), and cisgender LGBQ (i.e.,

lesbian, gay, bisexual, and queer; $n = 123$) participants. Around 60% of the pilot sample was between 18-44 years old, the rest were older, and 45% reported a nonwhite race/ethnicity, split fairly evenly among Asian, Latinx, African American and American Indian. In addition to the scale items, researchers asked these pilot test participants about their perceptions and experiences of taking the survey. LGBQ persons provide a natural validation subsample because many are likely to be familiar with language relevant to gender dysphoria, and some may experience gender variance or discomfort with gender roles themselves. However, by virtue of our sampling strategy in MTurk, the LGBQ sample indicated they were not transgender-, nonbinary-, or genderqueer-identified. Participants responded to questions about the language, inclusivity, and instructions of the survey using a Likert-type scale ranging from 1 (completely disagree) to 5 (completely agree). The mean for all evaluation questions was nearly 4, indicating that participants generally agreed that the instructions and questions used simple, clear language and were free of gender bias, worded appropriately, and gender inclusive. In written comments, 25 cisgender participants responded, and nine of these individuals reported uncertainty about questions referencing "affirmed gender." Among transgender, nonbinary, and genderqueer participants who made comments, 10 individuals thanked the researchers for asking the survey questions, another 10 individuals provided random comments, and one individual noted "affirmed gender" as confusing. An additional 20 transgender, nonbinary, and genderqueer participants indicated No or N/A to the question, "Do you have comments about the questions on the UGDS-GS?" Of note, some transgender, nonbinary, and genderqueer participants disclosed that even though the questions touched on sensitive or sad topics, they expressed a willingness to participate because they felt it was for a good cause. Data collection procedures for prior formal pilot testing and for the final confirmatory factor analysis (CFA) sample reported below were approved by the institutional review board of the lead author's home institution.

The pilot data were used for Exploratory Factor Analyses (EFA) and Principle Components Analyses (PCA) using SPSS 23 software to estimate variance extracted. EFA was used to evaluate the dimensionality of the items on the UGDS-GS by uncovering the least number of factors needed to explain the correlation among the items (Brown, 2006). Analyses began with PCA, which identified two likely factors (Affirmed Gender and Dysphoria). The Affirmed Gender subscale includes four positively valenced items that indicate complete agreement with the benefits of living in the affirmed gender. Dysphoria is measured with 14 items that indicate distress about one's physical characteristics, expected behaviors and sense of self in their assigned sex. Using pilot studies of binary transgender, nonbinary/genderqueer, and cisgender LGBQ samples, oblimin rotation with pairwise deletion of missing data estimated the likely factor loadings on each of the two strongest factors (McGuire & Catalpa, 2017).

Based on the EFA, two items did not meet the item-factor loading criteria of $> .40$ (Tabachnick, Fidell, & Osterlind, 2001): "Living in my assigned sex feels positive for me" and "I enjoy seeing my naked body in the mirror". There were many opportunities to receive feedback from respondents, clinicians, and researchers that informed the decision to drop these items. Respondents had opportunity for feedback in two separate groups after piloting the measures, as well as in written commentary after the online pilot administration. Clinician teams in both of the involved clinics participated in reviewing and providing feedback on items and inclusion or dropping of poorly performing items. Finally, the findings were reported at several conferences as they unfolded, where at least 200 transgender persons, researchers and clinicians have had opportunity to ask questions and provide input into the measurement revision. The first item was excluded given these low initial factor loadings as well as conceptual feedback from group participants and clinicians that the first item (i.e., "living in my assigned sex feels positive for me") was confusing for nonbinary or non-transgender persons. Feedback from youth, pilot study participants, and clinicians suggested that the item, "I enjoy

Table 1. Utrecht Gender Dysphoria Scale – Gender Spectrum (UGDS-GS) survey and scale.

Directions: For each question, select the response that best describes how much you agree with each statement. *Note:* **Assigned sex** means the sex you were assigned at birth and **affirmed gender** is the gender you currently identify with.

	Disagree completely	Disagree	Neither agree nor disagree	Agree	Agree completely
1. I prefer to behave like my affirmed gender.[GA]	1	2	3	4	5
2. Every time someone treats me like my assigned sex I feel hurt.	1	2	3	4	5
3. It feels good to live as my affirmed gender.[GA]	1	2	3	4	5
4. I always want to be treated like my affirmed gender.[GA]	1	2	3	4	5
5. A life in my affirmed gender is more attractive for me than a life in my assigned sex.[GA]	1	2	3	4	5
6. I feel unhappy when I have to behave like my assigned sex.	1	2	3	4	5
7. It is uncomfortable to be sexual in my assigned sex.	1	2	3	4	5
8. Puberty felt like a betrayal.	1	2	3	4	5
9. Physical sexual development was stressful.	1	2	3	4	5
10. I wish I had been born as my affirmed gender.	1	2	3	4	5
11. The bodily functions of my assigned sex are distressing for me (i.e. erection, menstruation).	1	2	3	4	5
12. My life would be meaningless if I would have to live as my assigned sex.	1	2	3	4	5
13. I feel hopeless if I have to stay in my assigned sex.	1	2	3	4	5
14. I feel unhappy when someone misgenders me.	1	2	3	4	5
15. I feel unhappy because I have the physical characteristics of my assigned sex.	1	2	3	4	5
16. I hate my birth assigned sex.	1	2	3	4	5
17. I feel uncomfortable behaving like my assigned sex.	1	2	3	4	5
18. It would be better not to live, than to live as my assigned sex.	1	2	3	4	5

Note. [GA] indicates items on the *Gender Affirmation* subscale, others indicate *Dysphoria*.
Suggested scale citation:
McGuire, J. K., Rider, G. N, Catalpa, J.M., Steensma, T. D, Cohen-Kettenis, P.T., & Berg, D. R., (2019). Utrecht Gender Dysphoria Scale – Gender Spectrum (UDGS-GS). In Milhausen, R., Sakaluk, J., Fisher, T., Davis, C., & Yarber, W. (Eds.), *Handbook of Sexuality-Related Measures*. New York: Routledge, https://doi.org/10.4324/9781315183169

seeing my naked body in the mirror" indicated that responses to this item could be related to body image in ways distinct from gender identity or dysphoria. Based on these comments, and the fact that the item did not load on either factor > .40, it was also eliminated.

Final proposed measure

The revised UGDS-GS now consists of one 18-item instrument with gender-neutral language. Researchers can use this updated measure with a person of any gender identity and expression. For the current study, this finalized measure was again sampled using MTurk to examine responses from 121 cisgender LGBQ, 295 binary transgender, and 587 nonbinary or genderqueer persons. The final tested measure can be seen in Table 1 (McGuire et al., 2019).

Current study participants

To complete CFA and measurement invariance testing on the UGDS-GS, researchers recruited three comparison samples (total N = 1005) from an international online population of adults over age 18. Overall, 13% were age 18-24, 64% were 25-34, 15% were 35-44, and 8% were over 45.

Specifically, we recruited cisgender LGBQ participants for one sample (n = 121), binary-identified transgender people (i.e., transgender men and transgender women) for the second sample (n = 297), and nonbinary/genderqueer spectrum transgender people for the third sample (n = 587). Across samples, participants tended to be well-educated, with the majority of participants receiving a bachelor's degree or higher, and median individual income was between 40,000-49,000 USD. Our survey included the revised UGDS-GS and several other gender- and health-focused questionnaires unrelated to the content of the present study.

Cisgender

A little under half of the cisgender LGBQ sample were assigned male at birth (n = 56, 46.7%), whereas 53.3% (n = 64) were assigned female at birth. The cisgender LGBQ participants identified their sexual orientations as lesbian (n = 27, 22.3%), gay (n = 28, 23.1%), bisexual (n = 50, 41.3%), queer (n = 1, 0.8%), mostly heterosexual (n = 1, 0.8%), asexual (n = 4, 3.3%), pansexual (n = 9, 7.4%) or other (n = 1, 0.8%). Cisgender LGBQ participants took the survey from a variety

of locations, with 27% of the sample living in a country other than the U.S., including India ($n = 16$, 13.%) or the European Union ($n = 17$, 14%), and the remainder from across the U.S. Cisgender LGBQ participants identified as White, non-Latinx ($n = 70$, 59.3%), Black ($n = 18$, 14.9%), Latinx ($n = 18$, 14.9%), Native American ($n = 5$, 4.1%), Asian/Pacific Islander ($n = 14$, 11.6%), and multiethnic/other ($n = 3$, 2.5%).

Binary transgender

A majority of the binary-identified transgender sample was assigned male at birth ($n = 209$, 70.4%), whereas 29.0% ($n = 86$) were assigned female at birth, and two (.5%) participants were intersex. Binary transgender participants identified their sexual orientations as lesbian ($n = 36$, 12.1%), gay ($n = 75$, 25.3%), bisexual ($n = 98$, 33.0%), queer ($n = 27$, 9.1%), mostly heterosexual ($n = 37$, 12.5%), asexual ($n = 6$, 2.0%), pansexual ($n = 11$, 3.7%) and other ($n = 7$, 2.4%). Binary transgender participants all lived in the U.S. Binary transgender participants identified as White, non-Latinx ($n = 171$, 57.6%), Black ($n = 51$, 17.2%), Latinx ($n = 69$, 23.2%), Native American ($n = 5$, 1.7%), Asian/Pacific Islander ($n = 14$, 4.7%), and multiethnic/other ($n = 3$, 1.0%).

Nonbinary/genderqueer

A little over half of the nonbinary/genderqueer sample were assigned male at birth ($n = 323$, 55.5%), whereas 40.9% ($n = 238$) were assigned female at birth, and 2.4% ($n = 14$) were intersex. Nonbinary/genderqueer participants identified their sexual orientations as lesbian ($n = 52$, 8.9%), gay ($n = 56$, 9.6%), bisexual ($n = 184$, 31.5%), queer ($n = 94$, 16.1%), mostly heterosexual ($n = 76$, 13.0%), asexual ($n = 42$, 7.2%), pansexual ($n = 69$, 11.8%) and other ($n = 12$, 2.1%). Nonbinary/genderqueer participants all lived in the U.S. Nonbinary/genderqueer participants identified as White, non-Latinx ($n = 311$, 53.0%), Black ($n = 88$, 15.0%), Latinx ($n = 151$, 25.7%), Native American ($n = 14$, 2.4%), Asian/Pacific Islander ($n = 38$, 6.5%) and multiethnic/other ($n = 17$, 3.0%).

Analytic approach

CFA and measurement invariance testing were conducted in Mplus 8.1 (Muthén & Muthén, 2004–2012). We tested whether the data demonstrated good model fit with the aforementioned two-factor structure CFA. Next, we tested configural, metric, scalar, and residual invariance across the binary transgender, nonbinary/genderqueer, and cisgender LGBQ groups. Configural invariance supports that the measurement model are equivalent for each group. Metric invariance suggests that the degree to which each measured item contributes to the latent factor(s) are statistically equal across groups (Putnick & Borntstein, 2016). If an item is more strongly associated with a given factor for one group, this reflects that this specific item is more closely related to the factor for that group compared to other groups. Scalar invariance supports the idea that differences or changes in a latent mean are accurately reflected in mean differences across items. Finally, residual invariance is the degree of error not explained by the latent factor. Although not a requirement for assessing mean-level differences, the residual invariance provides important diagnostic utility when assessing the sensitivity of items across groups.

Test of measurement invariance starts with configural invariance, where parameters (i.e., item loads, intercepts, and residuals) are freely estimated to assess whether the same number of factors and patterns of loadings are similar across groups. This is followed by a systematic sequence of equality constraints that first assess the equivalence of factor loadings for metric invariance, intercepts for scalar invariance, and residuals for residual invariance. If model fit statistically declines when equality constraints are imposed, this signals that constrained parameters are statistically different across groups.

Recent work in the measurement invariance literature point to robust metrics of change in model fit that are not as sensitive to sample size as the traditional χ^2 difference test (Chen, 2007; Cheung & Resnvold, 2002; see Putnick & Bornstein, 2016 for review). For the current study, we used a series of criteria for invariance including a .01 change in CFI (ΔCFI; Chueng &

Rensvold, 2002), a .015 change in ΔRMSEA, and .030 change in ΔSRMR for metric invariance or a .015 change in ΔSRMR for scalar invariance (Chen, 2007; Putnick & Bornstein, 2016). We also present χ^2 values and $\Delta\chi^2$ for comparison and sensitivity.

Tests of measurement invariance were conducted in stages in order to assess measurement invariance of the UGDS-GS across all three groups of interest. If models were metric, scalar, or residual non-invariant, we assessed item-level differences to determine items contributing to the decline in model fit. This approach was followed by tests of measurement invariance via one-to-one comparisons across groups – binary transgender compared to nonbinary/genderqueer, binary transgender compared to cisgender LGBQ, and nonbinary/genderqueer compared to cisgender LGBQ – to better assess the measurement sensitivity across groups.

Results

Pilot testing occurred with each iteration of the measure including the refinement of item wording and decisions about the inclusion or exclusion of items based on factor loadings and reliability analyses. In these early stages, we established that two factors emerged in exploratory analyses across a wide range of samples, with one factor representing dysphoria, and a second representing affirmation. The eigenvalue for the dysphoria vs. affirmation factor is higher across pilot samples (~8 vs. ~2.2, respectively). However, in PCA, the four affirmation items tended to load poorly or negatively on the dysphoria factor, and would ultimately have been dropped in order to meet reliability requirements for a single scale. Furthermore, we were interested in maintaining subscales with uni-dimensionality (e.g., the item variance within a subscale is the result of a single latent factor) for future analyses that could employ item response examinations to test differential item functioning (DIF) across groups. It is critical that each subscale in the measure meet this assumption of uni-dimensionality for these future assessments. Future item response examinations and DIF tests will allow for investigation of pivot items across groups, identifying items

that are especially sensitive or specific in one group (e.g., binary transgender) but not another (e.g., genderqueer).

Exploratory and confirmatory factor analysis

Using our final data set ($N = 1005$) which includes purposive subgroups who were not included in any prior analyses of binary transgender ($n = 297$), nonbinary/genderqueer ($n = 587$) and cisgender LGBQ persons ($n = 121$), we calculated correlations (reported in Table 2), PCA, and then EFA to assess the extent to which the current data source mirrored our previous pilot assessments and whether the minor refinements did not change the overall structure of the constructs. Similar to the pilot data, the EFA with oblimin rotation specified two factors sorted into the dysphoric (loadings from .60-.80) and affirming factors (loadings .76-.84), with the same items loading in similar ways for the combined sample and across subgroups. There is one notable exception. In the EFA, one item stands out as distinctive among the groups: "I wish I had been born in my affirmed gender." In the full sample (e.g., binary transgender, nonbinary/genderqueer, and cisgender LGBQ) analyses, this item loads strongly on the dysphoria scale. For the nonbinary/genderqueer persons, this item also loads better on dysphoria ($\lambda = .60$) than on affirmed ($\lambda = .49$). For cisgender LGBQ persons, this item loads similarly on both affirmed ($\lambda = .51$) and dysphoria ($\lambda = .55$). However, for binary transgender persons, this item is clearly a component of the affirmed construct ($\lambda = .62$) relative to dysphoria ($\lambda = .42$). The EFA results clearly suggests that this item factors differently across the gender identity subgroups, but was maintained in the dysphoria subscale because it demonstrated adequate fit for this factor in the full sample model, and two subgroups.

Next, we estimated model fit with CFA using the full combined sample. The two factor model demonstrated adequate fit (CFI = .934, TLI = .924, SRMR = .056, and RMSEA = .068), factor loadings for the affirmation subscale ranged from $\lambda = .662-.786$, and from $\lambda = .623-.791$ for the dysphoria subscale, with a correlation of $r = .388$, $p = .032$ between the two factors.

Table 2. UGDS-GS item correlation matrix for combined sample (N = 1005).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 Behave *	1																	
n	1000																	
2 Hurt	.178	1																
n	997	998																
3 Live *	.526	.167	1															
n	994	993	995															
4 Treated *	.530	.256	.618	1														
n	995	994	991	996														
5 Attractive *	.438	.319	.552	.530	1													
n	997	996	993	994	998													
6 Unhappy	.182	.545	.213	.253	.368	1												
n	996	995	992	993	996	997												
7 Sexual	.174	.470	.101	.157	.275	.501	1											
n	996	995	992	993	995	995	997											
8 Puberty	.145	.525	.153	.243	.310	.487	.533	1										
n	995	994	991	992	994	994	994	996										
9 Stressful	.147	.400	.185	.247	.322	.419	.482	.571	1									
n	996	995	992	993	995	995	995	994	997									
10 Wish	.315	.377	.317	.374	.421	.413	.430	.449	.408	1								
n	992	991	988	989	991	991	991	991	991	993								
11 Bodily	.141	.472	.139	.215	.310	.440	.508	.547	.513	.463	1							
n	997	996	993	994	997	997	996	995	996	992	998							
12 Life	.196	.574	.141	.226	.325	.511	.527	.561	.455	.444	.521	1						
n	997	996	993	994	997	997	996	995	996	992	998	998						
13 Hopeless	.226	.512	.186	.249	.337	.554	.535	.524	.451	.477	.492	.618	1					
n	997	996	993	994	996	995	995	994	995	991	996	996	999					
14 Misgender	.190	.513	.204	.261	.272	.478	.406	.487	.425	.375	.426	.499	.544	1				
n	997	996	993	994	996	995	995	994	995	991	996	996	997	999				
15 Physical	.123	.497	.138	.198	.281	.525	.531	.525	.479	.438	.508	.539	.629	.515	1			
n	996	995	992	993	995	994	994	993	994	990	995	995	996	996	998			
16 Hate	.113	.515	.076	.152	.231	.496	.476	.542	.420	.433	.498	.598	.628	.488	.599	1		
n	993	992	989	991	993	992	991	990	991	987	993	993	993	993	993	995		
17 Uncomfort.	.147	.482	.112	.196	.315	.573	.503	.507	.469	.446	.491	.556	.633	.526	.600	.589	1	
n	995	994	991	992	995	994	993	992	993	989	995	995	995	995	995	993	997	
18 Not live	.114	.546	.074	.147	.196	.455	.456	.492	.354	.354	.455	.630	.579	.509	.541	.628	.536	1
n	993	993	989	990	993	992	991	990	991	987	993	993	993	993	993	991	994	995
Mean	2.69	2.23	2.83	2.85	2.83	2.48	2.28	2.31	2.55	2.67	2.46	2.31	2.33	2.36	2.30	2.21	2.47	2.08
SD	1.07	1.25	1.05	1.03	1.05	1.25	1.29	1.29	1.20	1.14	1.24	1.29	1.29	1.20	1.27	1.37	1.26	1.42

Note: Pairwise correlations are reported. Mean values closer to 4 indicate greater gender dysphoria. For items with an * higher values indicate greater gender affirmation.

Table 3. Model fit and tests of measurement invariance across groups.

	χ	df	p	$\Delta\chi$	Δdf	p	CFI	ΔCFI	RMSEA	ΔRMSEA	SRMR	ΔSRMR	Decision
FULL MODEL COMPARISON													
Configural	1112.73	402	< .001				.914		.073		.064		
Metric	1176.26	434	< .001	63.53	32	.001	.911	.003	.072	.001	.071	-.007	Accept
Scalar	1273.31	466	< .001	97.05	32	< .001	.903	.008	.072	.000	.075	-.004	Accept
Residual	1449.82	502	< .001	176.51	36	< .001	.886	**.017**	.075	-.003	.080	-.005	Reject
BINARY TRANSGENDER VS. NONBINARY/ GENDERQUEER													
Configural	848.91	268	< .001				.917		.070		.056		
Metric	870.71	284	< .001	21.80	16	.150	.916	.001	.068	.002	.060	-.004	Accept
Scalar	903.39	300	< .001	32.69	16	.008	.914	.002	.068	.000	.062	-.002	Accept
Residual	911.90	318	< .001	8.51	18	.970	.915	-.001	.065	.003	.063	-.001	Accept
BINARY TRANSGENDER VS. CISGENDER LGBQ													
Configural	660.11	268	< .001				.887		.084		.080		
Metric	694.47	284	< .001	34.36	16	.005	.882	.005	.083	.001	.087	-.007	Accept
Scalar	751.06	300	< .001	56.59	16	< .001	.870	**.012**	.085	-.002	.092	-.005	Reject
Residual	864.09	318	< .001	113.03	18	< .001	.843	**.027**	.091	-.006	.100	-.008	Reject
NONBINARY/ GENDERQUEER VS. CISGENDER LGBQ													
Configural	716.45	268	< .001				.927		.069		.062		
Metric	759.31	284	< .001	42.85	16	< .001	.923	.004	.069	.000	.068	-.006	Accept
Scalar	825.27	300	< .001	65.96	16	< .001	.915	.008	.070	-.001	.073	-.005	Accept
Residual	864.09	318	< .001	38.82	18	.003	.843	**.072**	.091	**-.021**	.100	**-.027**	Reject

Tests of measurement invariance: three group comparison

Table 1 presents model fit across tests of measurement invariance, and Table 2 presents the item loadings, intercepts, and residual variances from the configural model. In line with our EFA and CFA results, measurement invariance testing across groups supported configural invariance for the two factor model across the three groups, χ^2 = 1112.73, df = 402, p < .001, CFI = .914, TLI = .902, SRMR = .064, and RMSEA = .073, 95% CI .068, .078)

Next, we tested metric and scalar invariance by constraining the factor loadings and intercepts, respectively, to be equal across groups. Although $\Delta\chi2$ indicated a statistical difference in model fit, ΔCFI, ΔRMSEA, and ΔSRMR all support metric and scalar invariance across groups (Table 3). Both metric and scalar models also demonstrated acceptable fit to the data. Findings from residual invariance testing indicated error variance differed across groups. Item-level assessments of metric and scalar models showed that equality constraints to item 5 ("A life in my affirmed gender is more attractive for me than a life in my assigned gender") were the largest contributors to change in χ^2. A metric model that allowed item 5 to be freely estimated for cisgender LGBQ persons showed no decrement in model fit on the basis of $\Delta\chi^2$ (configural vs. metric with item 5 loadings freely estimated for cisgender LGBQ group, $\Delta\chi^2$ = 42.34, Δdf = 31, p = .084).

Subscale correlations for each subgroup ranged based on level of constraints. Binary transgender subgroup recipients with higher gender dysphoria also reported more happiness with affirmed gender (range r = .520-.525, p < .001). For nonbinary/genderqueer participants, the relationship was still positive and significant (r = .443-.487, p < .001). For cisgender LGBQ participants, the relationship between dysphoria and affirmed gender was not significant (r = .040-.108, *n.s.*).

Tests of measurement invariance: one-to-one comparisons

We also assessed the configural, metric, scalar, and residual invariance of models in one-to-one comparisons between our three groups of interest to better assess the sensitivity of the UGDS-GS to measuring items across groups. Results showed that binary transgender and nonbinary/genderqueer subgroup model demonstrated adequate fit up to residual invariance, suggesting that the assessment operates equally well for both subgroups (see Table 3).

Comparisons between binary transgender and cisgender LGBQ subgroups were metrically invariant, but failed to meet the criteria for scalar

invariance. Item-level testing indicated that constraints to the intercept of item 5 ("A life in my affirmed gender is more attractive for me than a life in my assigned sex.") contributed to the decline in model fit. When the intercept for item 5 was freely estimated in an otherwise scalar invariant model, the ΔCFI dropped to .005, indicating partial scalar invariance. In order to achieve partial residual invariance between binary transgender and cisgender LGBQ groups, residuals for item 5, 15 ("I feel unhappy because I have the physical characteristics of my assigned sex."), and 18 ("It would be better not to live, than to live as my assigned sex.") had to be freely estimated, in addition to the intercept for item 5. The ΔCFI between the partially invariant scalar model to the partially invariant residual model reflect a ΔCFI of .009. Notably, ΔRMSEA and ΔSRMR were well within acceptable range for full and partial invariant models (see Table 4).

We found that nonbinary/genderqueer and cisgender LGBQ groups demonstrated configural, metric, and scalar invariance, but not residual invariance. Attempts to free specific residual variances across groups did not successfully improve model fit, suggesting that residual variances differ appreciably across these two groups and should be constrained to be equal. Fortunately, residual variance is not an expected requisite for the use of measures to compare group means, given that it is unlikely that random error is equivalent across groups (Little, 2013).

Discussion

In the current study, we aimed to validate revisions made to the Utrecht Gender Dysphoria Scale (UGDS). Based on these analyses, we conclude that the newly developed UGDS-Gender Spectrum (UGDS-GS) demonstrates a large degree of invariance across binary transgender, nonbinary/genderqueer, and cisgender LGBQ subgroups; and therefore find this revision to be a substantial improvement. The measure functions acceptably in all three gender identity groups (configural and metric invariance). Across binary transgender and nonbinary/genderqueer groups, the measure functions very similarly with all four types of invariance. However, some items

tend to have different meanings among some groups, which was reflected through scalar non-invariance between cisgender LGBQ and binary transgender and residual non-invariance between cisgender LGBQ and both binary transgender and nonbinary/genderqueer groups. This last point highlights the specificity of the measure to distinguish the experiences of binary transgender and nonbinary/genderqueer persons differently from cisgender LGBQ individuals. It is valuable for a measure to serve the purpose of delineating the experiences of multiple distinct groups. Binary and nonbinary subgroups will have ways that they are similar to and different from one another, and potentially differ in ways that they are similar to and different from cisgender subgroups. The power of this measure to potentially distinguish those differences will hopefully be useful for community members grappling with identity, dysphoria and affirmed gender.

It is also notable that the degree to which factors are correlated differed across groups. In other words, our results suggest that the constructs of dysphoria and affirmed gender experiences exist within all groups and can be measured in each of the three groups. However, what is meant by gender dysphoria and gender affirmation likely is distinct for different gender identity subgroups, and the concepts are related to each other somewhat differentially across groups. These differences could bias estimates in studies that attempt to assess group differences, particularly between binary transgender and cisgender LGBQ subgroups, because the measure overall has less error variance in the binary transgender subgroup. For example, examining predictors of dysphoria and affirmed gender will be statistically more successful among binary transgender and nonbinary/genderqueer subgroups than cisgender LGBQ subgroups in part because of the enhanced error variance in the cisgender LGBQ group. Fortunately, the items appear to operate similarly for binary transgender and nonbinary/genderqueer persons, which eliminates current shortcomings in the original UGDS.

Through the iterative processes, multiple samples were obtained to pilot and validate the measure using standard measurement techniques including EFA, CFA, and measurement

Table 4. Unconstrained (configural) loadings, intercepts, and residual variances for binary transgender, nonbinary/ genderqueer, and cisgender LGBQ subgroups.

	Binary transgender							Nonbinary/ genderqueer							Cisgender LGBQ						
	β	λ	(se)	τ	(se)	θ	(se)	β	λ	(se)	τ	(se)	θ	(se)	β	λ	(se)	τ	(se)	θ	(se)
Affirmation																					
ITEM 1	0.679	0.734	(.06)	2.814	(.06)	0.632	(.06)	0.661	0.692	(.04)	2.616	(.04)	0.618	(.04)	0.600	0.676	(.10)	2.717	(.10)	0.813	(.12)
ITEM 3	0.810	0.869	(.06)	2.876	(.06)	0.396	(.05)	0.773	0.789	(.04)	2.791	(.04)	0.420	(.04)	0.835	0.910	(.09)	2.883	(.10)	0.359	(.10)
ITEM 4	0.783	0.777	(.05)	2.999	(.06)	0.382	(.04)	0.758	0.773	(.04)	2.786	(.04)	0.442	(.04)	0.885	0.988	(.09)	2.767	(.10)	0.269	(.11)
ITEM 5	0.752	0.774	(.06)	2.947	(.06)	0.461	(.05)	0.748	0.747	(.04)	2.892	(.04)	0.440	(.04)	0.432	0.493	(.11)	2.275	(.10)	1.057	(.14)
Dysphoria																					
ITEM 2	0.592	0.676	(.06)	2.595	(.07)	0.845	(.07)	0.632	0.739	(.05)	2.279	(.05)	0.820	(.05)	0.830	0.997	(.09)	1.100	(.11)	0.450	(.07)
ITEM 6	0.669	0.783	(.06)	2.655	(.07)	0.755	(.07)	0.631	0.725	(.05)	2.607	(.04)	0.795	(.05)	0.745	1.003	(.11)	1.425	(.12)	0.806	(.11)
ITEM 7	0.627	0.767	(.07)	2.530	(.07)	0.910	(.08)	0.600	0.727	(.05)	2.402	(.05)	0.939	(.06)	0.812	0.990	(.09)	1.117	(.11)	0.506	(.07)
ITEM 8	0.644	0.752	(.06)	2.567	(.07)	0.799	(.07)	0.706	0.875	(.05)	2.394	(.05)	0.769	(.05)	0.700	0.881	(.10)	1.242	(.12)	0.807	(.11)
ITEM 9	0.538	0.598	(.06)	2.747	(.07)	0.879	(.08)	0.621	0.710	(.04)	2.629	(.05)	0.804	(.05)	0.539	0.705	(.11)	1.683	(.12)	1.218	(.16)
ITEM 10	0.465	0.487	(.06)	2.946	(.06)	0.858	(.07)	0.591	0.647	(.04)	2.695	(.05)	0.781	(.05)	0.516	0.618	(.10)	1.878	(.11)	1.055	(.14)
ITEM 11	0.520	0.565	(.06)	2.750	(.06)	0.862	(.07)	0.670	0.801	(.05)	2.500	(.05)	0.788	(.05)	0.728	0.994	(.11)	1.558	(.13)	0.875	(.12)
ITEM 12	0.694	0.791	(.06)	2.639	(.07)	0.673	(.06)	0.731	0.906	(.05)	2.380	(.05)	0.717	(.05)	0.812	1.031	(.10)	1.200	(.12)	0.547	(.08)
ITEM 13	0.739	0.844	(.06)	2.672	(.07)	0.590	(.06)	0.754	0.927	(.04)	2.401	(.05)	0.653	(.04)	0.801	1.039	(.10)	1.158	(.12)	0.604	(.09)
ITEM 14	0.648	0.742	(.06)	2.537	(.07)	0.759	(.07)	0.713	0.840	(.04)	2.400	(.05)	0.681	(.04)	0.488	0.612	(.11)	1.726	(.12)	1.202	(.16)
ITEM 15	0.604	0.684	(.06)	2.575	(.07)	0.813	(.07)	0.732	0.887	(.04)	2.405	(.05)	0.684	(.04)	0.862	1.020	(.09)	1.092	(.11)	0.360	(.06)
ITEM 16	0.630	0.737	(.06)	2.636	(.07)	0.828	(.07)	0.724	0.961	(.05)	2.260	(.06)	0.837	(.05)	0.825	1.067	(.10)	0.958	(.12)	0.534	(.08)
ITEM 17	0.694	0.793	(.06)	2.755	(.07)	0.677	(.06)	0.720	0.847	(.04)	2.576	(.05)	0.667	(.04)	0.742	0.957	(.10)	1.275	(.12)	0.749	(.10)
ITEM 18	0.607	0.780	(.07)	2.483	(.08)	1.041	(.09)	0.684	0.952	(.06)	2.127	(.06)	1.027	(.07)	0.791	0.916	(.09)	0.833	(.11)	0.501	(.07)

β = standardized loadings; λ = unstandardized loadings; τ = item intercepts; θ = item residual variance.
Configural model fit CFI = .914, TLI = .902, SRMR = .064, and RMSEA = .073, 95% CI .068, .078.
Correlation between affirmation and dysphoria r = .520, p < .001 for binary transgender; r = .443, p < .001 for nonbinary/ genderqueer; r = -.085, p = .404 for Cisgender LGBQ. The correlation between factors for Cisgender LGBQ statistically differed from binary transgender (Wald = 27.93, p < .001) and nonbinary/ genderqueer (Wald = 23.31, p < .001), but did not differ between binary transgender and nonbinary/ genderqueer subgroups (Wald = 1.38, p = .241).

invariance to explore and validate the UGDS-GS. The EFA from the pilot study revealed a two-factor structure in the UGDS-GS, one factor representing gender dysphoria and one factor representing gender affirmation. Comparatively, the original UGDS supported a one-factor structure (Cohen-Kettenis & van Goozen, 1997; Steensma et al., 2013). The EFA also showed that two proposed questions performed poorly with low factor loadings, thus these two questions were dropped from further analyses, resulting in an 18-item final measure. CFA confirmed that a two-factor solution represented a viable model for the revised UGDS-GS in binary transgender, nonbinary/genderqueer, and cisgender LGBQ persons. The findings suggested relatively minor measurement error on items containing the word, affirmed, suggesting that gender affirmation has slightly distinct meanings, particularly for cisgender LGBQ people. Results suggest that the newly revised UGDS-GS can be used for clinical and community measurement with a variety of gender identities and expressions, and meets analytic assumptions for repeated administration.

A major conceptual implication of this revision pertains to the question 'What does 'affirmed gender' mean to a cisgender person?" The act of gender transition – previously called gender reassignment or sex reassignment – assumed a binary nature of gender and that transgender persons would move from one gender to another essentially "opposite" gender. Our efforts to reconstruct this scale to allow for more fluid movement along a spectrum of gender will likely need further refinement after usage in clinical settings to fully capture the nature of dysphoria, and the anchor against which it is referenced. For instance, we should expect that any cisgender person will have some elements of dysphoria, even if relatively minor (e.g., "I don't feel feminine enough"). If someone shifts which gender they reference their gender against, they may still feel dysphoria in the new referenced gender (e.g., "I now identify as male, but may not always want to be treated as male"). For a cisgender person, what it means to be treated as their affirmed gender can be difficult to conceptualize, especially if they are a little gender nonconforming. (e.g., "I am cis-female, identify as a female/woman, feel a

bit dysphoric in that role. What is my 'affirmed gender'?"). The new measure assumes that individuals define affirmed gender for themselves and answer with reference to the gender identity and role they express and use (e.g., butch, male, femme, female, genderqueer, agender, etc.). Over time and with use, the conceptualization of gender affirmation will develop, and at that point, adjustment to the measure may be warranted.

There is one item with the word affirmed that was retained in the Dysphoria subscale. This item "I wish I had been born my affirmed gender?" represents dysphoria for the largest number of people, but represents affirmed gender for one specific population that is the focus of this study, binary transgender persons. This is clearly a very powerful item that contains significant information, and opportunity for reflection. It was a difficult choice to retain this conceptual item with dysphoria, however, given the analytic advantage it posed, and the value it created for nonbinary/genderqueer persons to leave it in the Dysphoria subscale, it was determined to be an important indicator in that scale. Continuing item response analyses with differential item functioning (DIF) will allow us to disentangle the specificity and sensitivity of this item. Considerations of this item as a pivot item used to engage individuals in the nature of binary and nonbinary identities and interventions could provide useful feedback about the relevance of this concept (birth gender) for binary transgender and nonbinary persons.

These revisions to create the UGDS-GS are necessary to increase the methodological and clinical utility of the measure as society's conceptualization of gender becomes more complex. For example, in the past, the function of the UGDS was as a diagnostic tool to predict who would likely go on to request medical transition services. Diamond, Pardo, and Butterworth (2011) posit that within this medical model of transition, genderqueer, nonbinary, and other gender nonconforming identities that claim both or neither male or female identities were often overlooked by the medical community because they were not seen as appropriate candidates for medical gender care or transition. As medical gatekeeping reduces, there is greater variability in who is likely to seek medical transition and there is

greater variability in the types of both medical and surgical interventions transgender persons seek (Beek, Kreukels, Cohen-Kettenis, & Steensma, 2015). In support of these developments in the field, the revised UDGS-GS serves as a tool for a wide variety of gender dysphoric individuals to find and receive useful medical and mental health treatment.

Limitations, future directions and implications

The present study provides evidence that the revised UGDS-GS is a valid measure of gender dysphoria for cisgender and transgender spectrum (including nonbinary/genderqueer) individuals; however, this research represents only initial testing and validation of the UGDS-GS. Amazon MTurk can be a useful platform for collecting large, relatively diverse samples quickly (Cassese, Huddy, Hartman, Mason, & Weber, 2013), which can be invaluable for researchers engaged in measurement design and validation as they need to test multiple renditions of a new scale. That being said, MTurk workers represent a specific subpopulation, and may not fully generalize to offline populations. Additionally, as in all self-report research designs, the quality of MTurk data relies upon participants being honest about their identities and experiences, and it can be more difficult for researchers to know exactly who is participating in an MTurk survey than in a survey specifically advertised at community organizations or on listservs catering to a given population, especially because MTurk participants are paid for their time to complete the survey. However, the pay for time on this survey is consistent with normal pay times in the Mturk platform (1-3$for 10-20 minutes) so there was no reason to think participants would take the survey simply for the money. Items were also cross-validated across the survey to ensure consistent identity responses and participants with inconsistent identity patterns were dropped. Finally, the cisgender LGBQ sample used in the present study was not initially collected with the intention of comparing responses to the two transgender samples. Although all the survey items used in the present study were consistent across the cisgender and transgender samples, the cisgender sample data was collected about 1 year earlier than the

transgender samples, and the MTurk call for participants differed for the cisgender sample survey. MTurk participant demographics can vary based on the time of day when a survey is posted, the keywords used to advertise a survey, and the qualifications researchers request of survey participants (e.g., nationality); therefore, the cisgender sample may differ from the two transgender samples in unknown ways. The essential validity of the cisgender data was clear, and the analyses functioned acceptably, supporting the decision to refrain from collecting additional comparison data. Future data collected using other methods can provide further information about the validity and generalizability of the UGDS-GS in other populations.

Future studies and analyses will need to deconstruct the capacity of pivot items, or those that function uniquely across groups, to predict distinct gender-related concepts. For example, the item, "I wish I had been born as my affirmed gender" is a clear pivot item that behaves differently across groups in the EFA, moving to the affirmed scale for binary transgender only. In the combined group, it loads onto the dysphoria scale, and fits acceptably in dysphoria within each group, albeit less so among binary transgender persons. Item response analyses, with differential item functioning on this item and others with the words "affirmed gender" in them can help to deconstruct how binary transgender, nonbinary/genderqueer, and cisgender LGBQ persons are responding to the concept of an affirmed gender differently.

Future research is needed with bigger sample sizes that proportionately represent the heterogeneity of gender and sexual diverse populations. While we understand dysphoria is related to negative mental health consequences for binary transgender persons, we do not know as much about dysphoria for nonbinary, genderqueer, and genderfluid persons. Further, gender may be experienced and described differently based on language and other cross-cultural factors, which limits generalizability of findings. The original UGDS was developed and normed in Dutch, and one of the changes to the UGDS-GS was to shift some of the language nuance to have a more smooth flow in English. Most of the subsamples were exclusively from the U.S., save for the LGBQ final subsample, which had about 30

people living outside the U.S. Further studies in other languages and from other cultural lenses will be necessary in the future.

The original UGDS was designed by clinicians representing the objectives of institutionalized medicine for the purpose of diagnostic treatment for gender dysphoric individuals (Cohen-Kettenis & van Goozen, 1997; Steensma et al., 2013). At the time of the UGDS construction, gender was largely viewed in binary terms, and research was predominately focused on "transsexual women" (terminology that is now considered pejorative by some and should only be used if an individual specifically identifies with this language). In the modern era of transgender research, we have come to understand gender as multidimensional, complex, and developmental (McGuire, Kuvalanka, Catalpa, & Toomey, 2016). The UGDS-GS can still be used as part of an ongoing clinical diagnostic assessment process where the objective is to understand how an individual is experiencing aspects of dysphoria for both diagnostic and treatment planning purposes. The impact of this revision to create the UGDS-GS should invite a wider spectrum of gender studies, including longitudinal research to determine whether the UGDS-GS is capable of assessing changes in gender dysphoria and affirmation over time.

Public significance statement

This study found that a longstanding measure of discomfort with gender, historically divided by sex, can be combined so everyone takes the same version and still work. This will make it easier for anyone to use the measure without having to say their gender first. Also, it lays the groundwork for other measures to be changed in the same way so that everyone takes the same version.

Declaration of conflict of interest

The authors declare they have no conflicts of interest.

ORCID

G. Nic Rider http://orcid.org/0000-0002-6730-5613

References

Beek, T. F., Kreukels, B. P. C., Cohen-Kettenis, P. T., & Steensma, T. D. (2015). Partial treatment requests and underlying motives of applicants for gender affirming interventions. *The Journal of Sexual Medicine*, 12(11), 2201–2205. doi:10.1111/jsm.13033

Berg, D., Spencer, K., McGuire, J., Becker-Warner, R., Vencill, J. A., & Catalpa, J. (2017, January). The Gender Affirmative Lifespan Approach: Promoting positive identity by building resiliency, increasing gender literacy, moving beyond the binary, and developing sex-positive pleasure and satisfaction. In G. Knudson (Chair), *WPATH presents: The inaugural USPATH Scientific Conference. Symposium conducted at the meeting of the United States Professional Association of Transgender Health*, Los Angeles, CA.

Buhrmaster, M., Kwang, T., & Gosling, S.D. (2011). Amazon's mechanical Turk: A new source of inexpensive. *Perspectives on Psychological Science*, 6, 3–5. doi:10.1177/1745691610393980

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456–466. doi:10.1037/0033-2909.105.3.456

Cassese, E., Huddy, L., Hartman, T., Mason, L., & Weber, C. (2013). Socially mediated internet surveys: Recruiting participants for online experiments. *PS: Political Science & Politics*, 46, 775–784. Retrieved from http://www.jstor.org/stable/43284764. doi:10.1017/S1049096513001029

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. doi:10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255. doi:10.1207/S15328007SEM0902_5

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 & Adolescent Psychiatry*, 36(2), 263–271. doi:10.1097/00004583-199702000-00017

de Vries, A. L. C., McGuire, J. K., Steensma, T. D., Wagenaar, E., Doreleijers, T., & Cohen-Kettenis, P. T. (2014). Prospective young adult outcomes of puberty suppression in transgender adolescents. *PEDIATRICS*, 134(4), 696–704. doi:10.1542/peds.2013-2958

de Vries, A. L. C., Steensma, T. D., Doreleijers, T. A. H., & Cohen-Kettenis, P. T. (2011). Puberty suppression in adolescents with gender identity disorder: A prospective follow-up study. *The Journal of Sexual Medicine*, 8(8), 2276–2283. doi:10.1111/j.1743-6109.2010.01943.x

Diamond, L. M., Pardo, S. T., & Butterworth, M. R. (2011). Transgender experience and identity. In S. J. Shwartz, K. Luyckx, & V. L. Vignoles (Eds.), *Handbook of identity*

theory and research, New York, NY: Springer, 629–647. doi:10.1007/978-1-4419-7988-9.

Jürgensen, M., Kleinemeier, E., Lux, A., Steensma, T. D., Cohen-Kettenis, P. T., Hiort, O., … Köhler, B. (2013). Psychosexual development in adolescents and adults with disorders of sex development-results from the German clinical evaluation study. *The Journal of Sexual Medicine*, *10*(11), 2703–2714. doi:10.1111/j.1743-6109.2012.02751.x

Khatchadourian, K., Amed, S., & Metzger, D. L. (2014). Clinical management of youth with gender dysphoria in Vancouver. *The Journal of Pediatrics*, *164*(4), 906–911. doi:10.1016/j.jpeds.2013.10.068

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

McGuire, J. K., Kuvalanka, K., Catalpa, J. M., & Toomey, R. B. (2016). Transfamily theory: How the presence of Trans family members informs gender development in families. *Journal of Family Theory & Review*, *8*(1), 60–73. doi:10.1111/jftr.12125

McGuire, J. K., Rider, G. N., Catalpa, J.M., Steensma, T. D., Cohen-Kettenis, P.T., & Berg, D. R. (2019). Utrecht Gender Dysphoria Scale – Gender Spectrum (UDGS-GS). In R. Milhausen, J. Sakaluk, T. Fisher, C. Davis & W. Yarber (Eds.), *Handbook of sexuality-related measures*. New York, NY: Routledge. doi:10.4324/9781315183169.

McGuire, J. K., & Catalpa, J. M. (2017, February). Revision of the Utrecht Gender Dysphoria Scale. Presentation at the Inaugural USPATH Symposium, Los Angeles, CA.

Olson, J., Schrager, S. M., Belzer, M., Simons, L. K., & Clark, L. F. (2015). Baseline physiologic and psychosocial characteristics of transgender youth seeking care for gender dysphoria. *Journal of Adolescent Health*, *57*(4), 374–380. doi:10.1016/j.jadohealth.2015.04.027

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. doi:10.1016/j.dr.2016.06.004

Rider, G. N., Vencill, J. A., Berg, D. R., Becker-Warner, R., Candelario-Perez, L., & Spencer, K. G. (2019). The gender affirmative lifespan approach (GALA): A framework for competent clinical care with nonbinary clients. *International Journal of Transgenderism*, *20*(2+3), 275–288. doi:10.1080/15532739.2018.1485069

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*(1), 89–99. doi:10.1017/S0033291704002776

Steensma, 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. In T. D. Steensma (Ed.), *From gender variance to gender dysphoria: Psychosexual development of gender atypical children and adolescents*, Amsterdam, NL: Ridderprint, 41–56.

Tabachnick, B. G., Fidell, L. S., & Osterlind, S. J. (2001). *Using multivariate statistics* (4th ed.). New York, NY: Harper Collins.

van de Grift, T. C., Cohen-Kettenis, P. T., de Vries, A. L. C., & Kreukels, B. P. C. (2018). Body image and self-esteem in disorders of sex development: A European multicenter study. *Health Psychology*, *37*(4), 334–343. doi:10.1037/hea0000600

van de Grift, T. C., Elaut, E., Cerwenka, S. C., Cohen-Kettenis, P. T., De Cuypere, G., Richter-Appelt, H., & Kreukels, B. P. C. (2017). Effects of medical interventions on gender dysphoria and body image: A follow-up study. *Psychosomatic Medicine*, *79*(7), 815–823. Retrieved from https://journals.lww.com/psychosomaticmedicine/Fulltext/2017/09000/Effects_of_Medical_Interventions_on_Gender.14.aspx.

Wallien, M. S. C., & Cohen-Kettenis, P. T. (2008). Psychosexual outcome of gender-dysphoric children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(12), 1413–1423. doi:10.1097/CHI.0b013e31818956b9