

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Trans Collaborations Academic Papers

Psychology, Department of

2022

Test–Retest Reliability and Sensitivity of a Brief Clinical Monitoring Measure for Transgender and Gender Diverse Adults: The Trans Collaborations Clinical Check-In (TC³)

T. Zachary Huit

Natalie R. Holt

Alexander Farquhar-Leicester

Rebecca L. Brock

Richard Mocarski

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.unl.edu/trans>



Part of the [Counseling Commons](#), [Developmental Psychology Commons](#), [Development Studies Commons](#), [Gender and Sexuality Commons](#), [Mental and Social Health Commons](#), [Other Psychiatry and Psychology Commons](#), [Other Psychology Commons](#), and the [Other Sociology Commons](#)

This Article is brought to you for free and open access by the Psychology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Trans Collaborations Academic Papers by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

T. Zachary Huit, Natalie R. Holt, Alexander Farquhar-Leicester, Rebecca L. Brock, Richard MocarSKI, Nathan Woodruff, and Debra Hope

Test-Retest Reliability and Sensitivity of a Brief Clinical Monitoring Measure for Transgender and Gender Diverse Adults: The Trans Collaborations Clinical Check-In (TC³)

T. Zachary Huit,¹ Natalie R. Holt,¹
Alexander Farquhar-Leicester,¹ Rebecca Brock,¹
Richard A. MocarSKI,² Nathan Woodruff,³
and Debra A. Hope¹

¹ Department of Psychology, University of Nebraska-Lincoln

² Office of Research, San José State University

³ Trans Collaborations Local Community Board, Lincoln, Nebraska,
United States

Correspondence – T. Zachary Huit, Department of Psychology, University
of Nebraska-Lincoln, 238 Burnett Hall, Lincoln, NE 68588-0308,
United States. Email: zach.huit@gmail.com

ORCID

T. Zachary Huit <https://orcid.org/0000-0003-1607-5763>

Rebecca Brock <https://orcid.org/0000-0001-7826-6421>

Richard A. MocarSKI <https://orcid.org/0000-0003-4220-0652>

Debra A. Hope <https://orcid.org/0000-0003-3685-9821>

Published in *Psychology of Sexual Orientation and Gender Diversity*, 2022

doi:10.1037/sgd0000598

Copyright © 2022 American Psychological Association. Used by permission.

Submitted July 9, 2021; revised June 4, 2022; accepted July 6, 2022; published August 18,
2022.

“This paper is not the copy of record and may not exactly replicate the authoritative
document published in the APA journal.”

Abstract

The current study aimed to examine the test–retest reliability and sensitivity of the Trans Collaborations Clinical Check-In (TC³) in a 3-month period with four assessment points at baseline, 1, 2, and 3 months to examine its utility as a clinical progress monitoring measure. This study builds on the initial validation study conducted by Holt et al. (2019). The sample of 32 transgender and gender diverse (TGD) participants were chosen who met screening for at least modest depression and anxiety, and did not have other significant risk factors (e.g., mania, self-harm). Participants completed a battery of measures that assessed mood, well-being, and gender-related constructs at each of the time points in addition to demographic questionnaires. Overall, the TC³ exhibited excellent test–retest reliability. While there was no systematic change in scores, there was some random variation of scores around the mean; and large, within-person correlations between time points. The TC³ also demonstrated convergence with many of the gender-related constructs, and to a lesser degree demonstrated criterion validity with mental health constructs. Further longitudinal study with larger samples in addition to study within intervention frameworks are necessary next steps to understand the utility of the TC³ for assessing systematic change over time. Overall, the current study highlights the initial utility of the TC³ to measure aspects of gender-related well-being across time, such as during health or behavioral health services.

Public Significance Statement

The overall findings of the study suggest that the Trans Collaborations Clinical Check-In (TC³) is a valid and reliable tool for use with transgender and gender diverse (TGD) people in clinical settings, which addresses the dearth of validated, brief TGD-specific assessments that are routine essentials for providing evidence-based care.

Keywords: transgender, gender diverse, well-being, measurement, psychometrics

Transgender and gender diverse (TGD) individuals have diverse intersections of identities and experiences (de Vries, 2012; Kuper et al., 2012). Though often LGBTQ+ (lesbian, gay, bisexual, transgender, queer, questioning, and all those among the diverse gender and sexuality spectrum) communities are treated as one homogenous group, it is important to understand the nuances of identities and experiences among these communities, namely for TGD people who are underrepresented in health care and health research broadly (Logie et al., 2012; Maragh-Bass et al., 2017). It is also necessary to consider the sociopolitical context with which TGD people exist. Historically, societal misconceptions of binary gender categories and heterosexist ideologies of sexuality have contributed to the stigmatization of TGD communities (Gartner & Sterzing, 2018; Javaid, 2018).

The mental health field as a whole has not been exempt from such assumptions. The treatment of gender variance as a mental illness (Austin & Craig, 2015; Hope et al., 2016) has perpetuated mental health disparities faced by many TGD individuals. Collectively, the lack of provider training, outdated understandings of standards of care, and lack of knowledge regarding gender diversity and experiences of TGD communities contribute to the overall paucity of affirming providers (Austin & Craig, 2015; Benson, 2013; Hope et al., 2016; Shipherd et al., 2010). The materiality of such disparities highlights the need for affirming evidence-based methods of treatment and assessment for TGD individuals to reduce stigma and combat barriers to seeking mental health services (Hope et al., 2016; Shipherd et al., 2010).

Another obstacle to providing culturally competent, evidence-based practice with TGD communities is the lack of validated measures. Many assessment tools that are used with TGD individuals rely on adaptations of existing measures, such as the *Transgender Stigma Scale* (Mizock & Mueser, 2014) which was modified from a scale of mental illness stigma (King et al., 2007). While adaptations such as these may be practical solutions to the dearth of TGD-specific measures, they may not fully capture distinct experiences and needs for gender diverse communities (Hope et al., 2016; Shulman et al., 2017). As such, there is a particular need for further development of TGD-specific measures.

Clinical progress monitoring has roots spanning back as far as the 1970s and earlier, which was highlighted by Barlow in the 1980s (Barlow, 1981) and is consistent with scientist-practitioner guidelines regarding evidence-based practice. While there has been some debate among mental health providers as to the clinical utility of such measures (Boswell et al., 2015), its role in evidence-based practice was reaffirmed by American Psychological Association (APA) guidelines in 2006 (APA Presidential Task Force on Evidence-Based Practice, 2006). The use of regular progress monitoring serves to confirm that treatment is appropriate and helpful for clients. Additionally, progress monitoring increases the efficacy of therapeutic interventions (Reese et al., 2009) and facilitates open communication between client and mental health professionals, which is a key factor in positive treatment impacts (Carrier et al., 2012). Due to the potential of

stigmatizing mental health practices (Austin & Craig, 2015) and various mental health disparities encountered by TGD individuals, regular monitoring may open the lines of communication between individuals and providers. Such communication may ensure that mental health professionals are cognizant of TGD clients' needs and enact TGD-affirming practices, enhancing treatment outcomes.

A review of TGD-specific measures conducted by Shulman and colleagues (Shulman et al., 2017) identified several measures spanning various domains that may be salient to TGD individuals across a broad range of experiences and identities. Some salient domains were: positive and negative thoughts regarding gender (Bauerband & Galupo, 2014), stressors and protective factors (Testa et al., 2015), salience of TGD identity and transition (Barr et al., 2016), community connectedness (Barr et al., 2016), congruence of desired gender and expression (Kozee et al., 2012), and voice quality (Dacakis et al., 2013). More recently, Jones and colleagues (Jones et al., 2019) created a comprehensive 38-item measure of gender congruence and life satisfaction. Several of these measures were recently included in a systemic review of tools to assess gender dysphoria (Bowman et al., 2021). While these constructs serve as valid and affirming measures for use with TGD individuals to assess overall functioning, the need for additional brief clinical monitoring tools that cover several functional areas was identified.

To address this gap in assessment methods with TGD individuals, the Trans Collaborations Clinical Check-In (TC³) was created to serve as a routine clinical monitoring measure to broadly assess constructs relevant to TGD care, rather than gender dysphoria or any other single construct alone (Holt et al., 2019).¹ The following serves as a general overview for the development of the TC³. More detailed accounts of the processes involved in the scale development for this measure can be found in the original validation study (Holt et al., 2019). This brief 18-item measure covers a range of areas such as social support, body dysphoria, comfort with identity expression, and coping with stigma and discrimination. Such areas were deemed important in the creation

1. The TC³ was included in the review of gender dysphoria measures by Bowman et al. (2021). The reason for its inclusion is unclear as the TC³ is not an overall measure of gender dysphoria. Some items assessing gender dysphoria are included as this is an important aspect of the lived experience of many TGD clients that a clinician should monitor.

of the TC³ through interviews with TGD individuals and affirming providers that were recommended by TGD communities, using community-based participatory research (CBPR) frameworks. These domains mentioned above also demonstrate content overlap with other existing measures, some of which are summarized by Shulman and colleagues (Shulman et al., 2017).

The next phase of measure development was an online validation study conducted with 217 TGD individuals who completed the TC³ along with other established TGD-specific and well-being constructs. The TC³ revealed four stable factors; however, the authors recommend using the total score. Correlational analyses revealed higher TC³ total scores (indicating more overall positive functioning) were associated with higher life satisfaction, lower depression and anxiety, as well as greater positive affect and lower negative affect. There were also several significant relationships between the TC³ with the various TGD-specific measures, with higher scores on the TC³ being related to greater transgender congruence, appearance and acceptance, and gender identity reflection, as well as less rumination about gender identity.

Test-retest reliability is an essential aspect of developing sound assessments (Gnambs, 2014; Schatz & Ferris, 2013; Weir, 2005). Therefore, the current study aims to further validate the TC³ by examining the test-retest reliability and sensitivity at multiple time points over a period of time as would be applicable to the course of a therapeutic intervention. As the TC³ is meant to measure domains related to TGD identity that may vary somewhat over time depending on fluctuations in social and well-being, there may be lower overall test-retest reliability. However, to examine the utility of the TC³ as a clinically applicable monitoring assessment, the examination of its sensitivity to track changes in TGD-specific constructs is necessary.

Additionally, it is important to understand how the TC³ covaries with other constructs related to well-being. Previous studies have demonstrated that TGD individuals experience higher rates of depression and anxiety (Bockting et al., 2013; Budge et al., 2013). However, positive affect may increase resilience in the face of adverse life situations (Cohn et al., 2009), and relatedly, greater gender congruence is related to higher life satisfaction (Kozee et al., 2012). Further, social support has been shown to relate to lower depression and anxiety

(Budge et al., 2013), and support from transgender peers may moderate the relationship between stigmatizing experiences and subsequent psychological distress (Bockting et al., 2013). Therefore, further examinations of reliability and covariance with other gender-specific (identity, stigma, and protective factors) and well-being (depression, anxiety, life satisfaction, and positive/negative affect) measures informed the utility of the TC³ as a therapeutic monitoring tool for use with TGD clients.

Hypotheses

Hypothesis 1a: As identity-related constructs are likely to remain relatively stable across a shorter time frames, scores on the TC³ may stay relatively consistent across time points; thus, demonstrating test-retest reliability.

Hypothesis 1b: Relatedly, it is anticipated that, on average, there will be no significant linear or curvilinear change in scores across the four time points when implementing growth curve analysis, further demonstrating stability of scores.

Hypothesis 2: As domains in the TC³ overlap with existing TGD measures, it is expected that the TC³ will be significantly correlated with at least medium sized effects ($r > .30$) with other TGD-specific constructs, demonstrating convergent validity.

Hypothesis 3: As experiences related to gender identity may impact overall well-being, it is also expected that measures related to overall mental health will be significantly correlated with at least small effects ($r > .10$) with TC³ scores, demonstrating criterion validity.

Participants

Inclusion/Exclusion Criteria

The following inclusion criteria were set forth for this study: participants had to be 19 years of age or older, identify as TGD, and endorse base levels of depression and anxiety given the TC³ is intended for clinical settings. This was measured as having item totals of 2 or higher

for the anxiety- and/or depression-related questions on the Abbreviated Brief Symptom Inventory (ABSI; Lang et al., 2009). Exclusion criteria were: current, untreated bipolar disorder, severe untreated alcohol and/or substance use, active psychosis, and acute suicidality requiring immediate services. These criteria were assessed using selected portions of the M.I.N.I. International Neuropsychiatric Interview 7.2 (M.I.N.I.; Sheehan, 2015; Sheehan et al., 1998).

Forty-five potential participants expressed some level of interest in participation in the study. Thirty-five participants attended the initial meeting and completed informed consent, with the other 10 withdrawing due to scheduling conflicts or no longer wishing to participate. Three of the consented participants did not meet the inclusion and exclusion criteria. All 32 eligible participants completed the initial baseline assessment and 29 completed all time points. Two participants did not complete the third and fourth time point and one participant did not complete the final time point. All eligible participants (32) were included for analyses; however, some were excluded on a case-by-case basis due to missing data during analyses.

Demographics

Participants were between the ages of 19 and 71 ($M = 32.38$, $SD = 14.71$). Gender identity was reported in two ways. Participants self-selected into one of three categories: transwoman/trans woman/MTF/woman ($n = 10$; 31.3%), transman/trans man/FTM/man ($n = 12$; 37.5%), nonbinary/gender nonconforming/genderqueer/agender/bigender/another gender minority ($n = 10$; 31.3%). They were also given the space to freely describe their gender (e.g., genderqueer; nonbinary; gender fluid; transmasculine; transexual woman; nonbinary agender; gender creative, both male and female; polygender; trans masculine - nonbinary; man of trans experience; very beautiful woman; or normal). Participants described their sexual orientation by selecting one of the following categories: heterosexual/straight ($n = 6$; 18.8%); gay/lesbian ($n = 6$; 18.8%); bisexual ($n = 5$; 15.6%); pansexual ($n = 9$; 28.1%); a different sexual orientation ($n = 6$; 18.8%). Those who chose “a different sexual orientation” were able to write in their identities (e.g., queer, mostly straight/bicurious). Participants were able to select multiple racial/ethnic identities. The majority

identified as European American/ Caucasian/White ($n = 28$; 87.5%), while others identified as African American/Black ($n = 1$; 3.1%), Asian American, including Pacific Islander ($n = 1$; 3.1%), Native American/ American Indian/Alaskan Native ($n = 1$; 3.7%), and Biracial/Multiracial ($n = 2$; 6.3%).

A majority of participants ($n = 20$; 62.5%), described their hometown as urban while the remainder described their hometowns as suburban and rural. Most participants ($n = 23$; 71.9%) described their yearly income as less than \$50,000, and were mostly employed either full- or part-time ($n = 21$; 65.7%) while the rest were either disabled/ unable to work or retired. The sample leaned more heavily toward having greater access to higher education with 27 (84.4%) of the participants having some college experience or degree.

Procedures

All study procedures were approved by the University of Nebraska-Lincoln Institutional Review Board as well as our Local TGD Community Board. Participants were recruited via flyers distributed among e-mail listservs, in-person recruitment at community-related or other events, by our local community board, and word-of-mouth. Scheduled participants met in-person for an initial session to determine eligibility, describe the study, and discuss consent procedures and safety planning before choosing to participate in the study.

During the initial in-person session, the M.I.N.I. and the ABSI were administered by either the first author or trained research personnel to determine inclusion and exclusion criteria. Eligible participants were then asked to complete the surveys via Qualtrics to assess a baseline measurement. The eligible participants were then contacted to complete online follow-up surveys at 1, 2, and 3 month intervals after the initial assessment to assess their functioning and were provided a separate survey link and individual identification number to mark their responses. Periodic correspondence to reduce attrition and engage participants throughout the study included a post-baseline thank you e-mail, as well as reminders before each follow-up survey, instructions at each time point, and prompts to complete the survey follow-ups. Participants were compensated up to a total of \$100 for

their participation, split between four assessment points, with the follow-up sessions having increasing amounts of compensation to reduce attrition.

Participants who qualified and consented to participate were compensated at the conclusion of each session: \$25 for the initial baseline (with an additional \$5 to cover the cost of parking/transportation if they traveled to campus or needed to park), \$15 for the 1-month follow-up, \$20 for the 2-month follow-up, and \$40 for the 3-month follow-up. The local TGD advisory board confirmed that these compensation schedules were appropriate for the nature and duration of the study. Those who consented to participate but did not qualify for the study were compensated \$20 at the conclusion of the inclusion/exclusion screening during the initial session (with an additional \$5 to cover for parking/transportation if needed). After the primary study at the first meeting, participants were also invited to engage in a separate research study on communications that was not a part of the current analysis.

Measures

Approach to Internal Consistency

With the exception of the M.I.N.I, Cronbach's α and McDonald's ω values are reported for each individual time point in **Table 1**. While Cronbach's α has been used as a convention in evaluating internal consistency, it may be misleading when there are fewer scale items and relies on the assumption of the unidimensionality of constructs (Graham, 2006; Hayes & Coutts, 2020) that may or may not be the case in certain scales. McDonald's ω does not rely on assumptions of τ -equivalence; therefore, it is a more general estimate of reliability (Hayes & Coutts, 2020). However, it may not be able to be calculated when there is an issue with covariance between scale items (i.e., due to a covariance of .00 or negative covariance between scale items). Therefore, both are reported as a metric for evaluation of internal consistency.

While there are differences in acceptability of the range of α and ω values (i.e., .70 to .90 is viewed as acceptable by some; Tavakol &

Table 1 Internal Consistency of Measures Using Cronbach's α and McDonald's ω

Measure	Timepoint			
	T1 α (ω)	T2 α (ω)	T3 α (ω)	T4 α (ω)
TC ³ total	.84 (—) ^a	.88 (.87)	.83 (—) ^a	.87 (.85)
TCS total	.89 (—) ^a	.90 (.91)	.91 (.91)	.90 (.90)
Appearance	.94 (.94)	.93 (.93)	.93 (.93)	.91 (.91)
Acceptance	.66 (.78)	.67 (1.71)	.64 (.64)	.79 (.87)
GRRS reflection	.61 (—) ^a	.82 (.82)	.83 (.82)	.90 (.90)
Rumination	.65 (.48)	.69 (.63)	.78 (.78)	.83 (.82)
Preoccupation	.78 (.77)	.83 (.84)	.73 (.71)	.76 (.74)
GMSR Nonaffirmation	.88 (.88)	.90 (.90)	.89 (.89)	.88 (.88)
Internalized transphobia	.93 (.93)	.90 (.90)	.92 (.92)	.95 (.95)
Negative expectations	.91 (.91)	.90 (.90)	.91 (.91)	.93 (.93)
Nondisclosure	.83 (.82)	.75 (.77)	.91 (.91)	.88 (.88)
Pride	.80 (.74)	.82 (.80)	.75 (.66)	.82 (.77)
Community	.89 (.90)	.83 (.81)	.83 (.94)	.83 (.92)
PHQ-9 total	.87 (.87)	.80 (.79)	.86 (.86)	.85 (.83)
GAD-7 total	.93 (.94)	.91 (.91)	.82 (.81)	.89 (.89)
I-PANAS-SF positive	.68 (.69)	.82 (.82)	.76 (.78)	.86 (.87)
Negative	.83 (.84)	.84 (.83)	.72 (.70)	.92 (.92)
SWLS total	.85 (.86)	.88 (.88)	.88 (.88)	.92 (.92)

Note. α = Cronbach's α ; ω = McDonald's ω ; TCS = Transgender Congruence Scale; TC³ = Trans Collaborations Clinical Check-In; GRRS = Gender Identity Reflection and Rumination Scale; GMSR = Gender Minority Stress and Resilience Scale; PHQ-9 = Patient Health Questionnaire (9-Item); GAD-7 = Generalized Anxiety Disorder (7-Item); I-PANAS-SF = International Positive and Negative Affect Scale (Short Form); SWLS = Satisfaction with Life Scale.

a. Omega values were unable to be obtained with the existing item composition in the respective scale.

Dennick, 2011), there are several factors to consider when evaluating the utility of a construct (Peters, 2014). The measures selected for comparison to the TC³ were chosen for the current study as they are widely used in research and have demonstrated conceptual and psychometric utility in previous research (e.g., Jones et al., 2019; Testa et al., 2017; van den Brink et al., 2020). These measures were also significantly related to the TC³ on various domains in the original validation study (Holt et al., 2019) and were retained for the current examination. Therefore, despite lower internal consistency at certain time points, they were retained in the current analyses given their previously demonstrated utility. Related limitations are further discussed below.

M.I.N.I. International Neuropsychiatric Interview 7.2 (M.I.N.I.; Sheehan, 2015; Sheehan et al., 1998)

The M.I.N.I. is a brief structured diagnostic interview that is designed to be useful in clinical and research contexts. It has been demonstrated as a valid and reliable for diagnostic criteria of psychological symptoms (Sheehan, 1998). The M.I.N.I. was selected for its short administration time of around 15 min covering several functional areas, making it a more feasible inclusion/exclusion tool. Only the suicidality, bipolar disorder, posttraumatic stress disorder, substance abuse, and psychotic disorders sections were used in the current study.

Trans Collaborations Clinical Check-In (TC³; Holt et al., 2019)

The TC³ is an 18-item scale of everyday experiences of TGD individuals. Items are assessed on a 1–5 Likert-scale with possible scores ranging from 18–90, and higher scores indicating greater functioning. Further measurement properties will be discussed further below. Domains of the TC³ include: social support (i.e., family, friends, work/school), body dysphoria (i.e., “How comfortable were you about your genitals?”; “How comfortable were you about your secondary sex characteristics?”), comfort with identity expression (i.e., “How many people that you care about know about your gender identity?”; “How close do you feel to your ideal self-expression”), and coping with stigma and discrimination (i.e., “How capable do you feel to handle any stressors that might arise due to your gender identity?”).

Transgender Congruence Scale (TCS; Huit et al., 2021; Kozee et al., 2012)

The TCS is a 12-item scale including a total and two subscale scores: *Appearance Congruence* and *Gender Identity Acceptance* that measures external and internal aspects of gender congruence, respectively. Higher scores indicate greater congruence and acceptance of gender identity. The TCS has been demonstrated to be a reliable and valid measurement (Kozee et al., 2012), though test–retest reliability has yet to be examined (Shulman et al., 2017). The TCS demonstrated mixed internal consistency in the current study as demonstrated in

Table 1. In particular, the *Gender Identity Acceptance* subscale demonstrated lower internal consistency for time-point 3, and McDonald's ω was unable to be obtained for time-point 1 for the *Total Score*.

Gender Identity Reflection and Rumination Scale (GRRS; Bauerband & Galupo, 2014)

The GRRS is a 15-item scale containing three subscales: *Reflection about Gender Identity*, *Rumination about Gender Identity*, and *Preoccupation with Other's Thinking*. Higher scores indicate greater amounts of individuals' positive and negative cognitions about their gender identity. The GRRS has shown internal consistency (Bauerband & Galupo, 2014), though other forms of validity and reliability, including test-retest reliability, have yet to be examined (Shulman et al., 2017). Internal consistency was mixed in the current sample (see Table 1), with lower internal consistency for the *Rumination* subscale in the first and second time points. McDonald's ω was unable to be obtained for time-point 1 for the *Reflection* subscale.

Gender Minority Stress and Resilience Scale (GMSR; Testa et al., 2015)

The GMSR is a 58-item measure containing nine subscales including: *Gender-Related Discrimination*, *Gender-Related Rejection*, *Gender-Related Victimization*, *Nonaffirmation of Gender Identity*, *Internalized Transphobia*, *Negative Expectations for Future Events*, *Nondisclosure* (of individual's gender identity), *Pride* (in one's gender identity), and *Community Connectedness*. Each of these subscales are independently scored, with higher scores indicating greater endorsement of experiences or alignment in each of the respective domains. The GMSR has demonstrated reliability and validity (Testa et al., 2015); however, test-retest reliability for the GMSR has yet to be examined. For the purpose of the current analyses, the *Discrimination*, *Rejection*, and *Victimization* scales were omitted from analyses due to their timeframe being broader than the scope of the current study. All other scales of the GMSR in the current analysis demonstrated internal consistency, which can be found in Table 1.

Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001)

The PHQ-9 is a nine-item measure with higher scores indicating greater depressive symptom presentation. The PHQ-9 has been demonstrated to be a reliable and valid depression instrument (Kroenke et al., 2001), and has also demonstrated excellent test-retest reliability (Löwe et al., 2004). Internal consistency for the PHQ-9 can be found in Table 1.

Generalized Anxiety Disorder-7 (GAD-7; Spitzer et al., 2006)

The GAD-7 is a seven-item scale that assesses generalized anxiety with higher scores suggesting more anxiety symptoms. The GAD-7 has demonstrated reliability and validity (Löwe et al., 2008), and has exhibited suitable test-retest reliability (Naeinian et al., 2011). The internal consistency statistics for this sample reported in Table 1 are in line with previous measurement reliability for the GAD-7.

International Positive and Negative Affect Schedule— Short Form (I-PANAS-SF; Thompson, 2007; Watson et al., 1988)

The PANAS contains two subscales: *Positive Affect* and *Negative Affect* with higher scores on either subscale indicate greater positive or negative mood states on their respective scales. The PANAS has been demonstrated to be a reliable and valid measure (Crawford & Henry, 2004), and has also demonstrated good test-retest reliability (Ostir et al., 2005). This shortened, 10-item version performed adequately in the current analysis (internal consistency is shown in Table 1) with the exception of *Positive Affect* at the first time point.

Satisfaction With Life Scale (Diener et al., 1985)

The Satisfaction with Life Scale is a five-item measure of global judgments of individuals' life satisfaction, where higher scores suggest greater life satisfaction. The Satisfaction with Life Scale has exhibited reliability and validity (Pavot et al., 1991), and has also demonstrated moderate test-retest stability (Pavot & Diener, 1993). Internal consistency for the SWLS is reported in Table 1.

Data Analytic Approach

SPSS Statistical Software packages, Version 22 were used for the majority of descriptive and correlational analyses. For Hypothesis 1a, Intraclass Correlations (ICCs) using two-way mixed effects and absolute agreement were examined to describe the test–retest reliability (Koo & Li, 2016) of the TC³. For Hypothesis 1b, using HLM7 software, we used multilevel modeling (MLM) with full information maximum likelihood estimation with robust standard errors (account for non-normality of data) to examine change in TC³ scores across four time points nested within-person. *Time* was calculated as days since the first assessment point for each participant and entered as uncentered. We also accounted for the possibility of curvilinear change over time with a quadratic time parameter. Nested model fit comparisons were conducted to determine (a) if including a quadratic time parameter improved the fit of the model and (b) whether time parameters should be modeled as fixed or random. Pearson’s correlations were used to examine the correlations between TC³ and TGD-specific and mental well-being measures for the second and third hypotheses. Correlations were computed in SPSS to address Hypotheses 2 and 3.

Results

Summary statistics including means, standard deviations, ranges, internal consistency, and correlations are presented in **Tables 1–3**. First, we examined the internal consistency of the measure across time points. The TC³ demonstrated overall internal consistency, with Cronbach’s α scores of .84, .88, .83, and .87 at each of the respective time-points 1 through 4. McDonald’s ω estimates were unable to be obtained for time-points 1 and 3; however, performed adequately for time-points 2 and 4 in terms of internal consistency. Follow-up analyses revealed that item nine (“In the last two weeks, how comfortable were you with your genitals?”) shared zero covariance with item 15 regarding satisfaction with support at work or school at time-point 1. Item nine also demonstrated negative covariances with other presentation-related and social support items at time-point 3. Therefore, this violated model assumptions for the TC³ time-points 1 and 3, making ω estimates unattainable in the current analysis.

Table 2 Summary Statistics by Measure and Time Point

Measures	Time 1		Time 2		Time 3		Time 4	
	M (SD)	Range	M (SD)	Range	M (SD)	Range	M (SD)	Range
TC ³ total	56.5 (11.68)	40.0–78.0	57.73 (12.97)	40.0–90.0	57.30 (11.17)	39.0–79.0	56.85 (12.05)	33.0–80.0
TCS total	31.52 (9.61)	14.33–55.42	34.15 (10.28)	17.33–55.42	33.25 (10.35)	17.33–55.42	32.84 (9.82)	15.42–54.42
Appearance congruence	21.20 (8.25)	8.11–40.56	23.33 (8.73)	8.11–40.56	23.01 (8.72)	9.11–40.56	22.70 (7.96)	8.11–40.56
Gender identity acceptance	8.94 (2.19)	3.67–11.67	9.32 (2.06)	3.67–11.67	8.84 (2.01)	3.33–11.67	8.73 (2.07)	5.0–11.67
GRRS-Reflection	12.87 (2.65)	7.0–20.0	11.97 (3.24)	5.0–20.0	12.20 (3.61)	5.0–20.0	11.59 (3.89)	5.0–20.0
Rumination	11.23 (3.04)	5.0–16.0	11.58 (2.90)	5.0–17.0	11.33 (3.50)	5.0–17.0	9.93 (3.47)	5.0–16.0
Preoccupation	12.10 (3.27)	5.0–19.0	11.74 (3.35)	5.0–18.0	11.73 (3.07)	6.0–18.0	10.55 (3.24)	5.0–16.0
GMSR Non affirmation	23.84 (6.35)	7.0–30.0	22.52 (6.50)	6.0–30.0	21.60 (6.63)	6.0–29.0	21.71 (6.12)	8.0–29.0
Internalized transphobia	20.26 (9.65)	8.0–40.0	21.06 (8.50)	8.0–40.0	21.70 (8.67)	8.0–40.0	21.11 (9.06)	8.0–37.0
Pride	27.0 (6.16)	11.0–38.0	27.0 (6.82)	13.0–39.0	26.07 (5.55)	12.0–37.0	26.00 (5.83)	16.0–39.0
Negative expectations	28.31 (9.11)	12.0–44.0	27.77 (8.26)	12.0–44.0	29.85 (8.41)	11.0–41.0	27.04 (8.69)	9.0–44.0
Nondisclosure of gender	16.14 (5.96)	5.0–25.0	16.65 (4.92)	8.0–25.0	17.81 (5.70)	5.0–25.0	17.04 (5.25)	5.0–25.0
Community connectedness	18.10 (5.43)	5.0–25.0	17.69 (4.73)	9.0–25.0	17.41 (5.89)	5.0–25.0	16.17 (5.61)	6.0–25.0
PHQ-9 total	8.13 (6.22)	0.0–22.0	8.97 (5.27)	2.0–18.0	9.20 (5.77)	1.0–22.0	8.36 (5.29)	0.0–20.0
GAD-7 total	15.23 (6.29)	8.0–28.0	14.27 (5.61)	7.0–28.0	16.47 (4.67)	10.0–25.0	15.17 (5.46)	7.0–27.0
I-PANAS-SF								
Positive	16.48 (3.88)	9.0–23.0	15.58 (4.40)	7.0–25.0	14.43 (4.32)	6.0–24.0	15.62 (5.27)	5.0–24.0
Negative	11.61 (4.22)	7.0–20.0	11.10 (4.72)	5.0–25.0	12.00 (4.24)	5.0–22.0	11.69 (5.20)	5.0–22.0
SWLS total	18.90 (7.30)	5.0–30.0	19.10 (7.17)	5.0–29.0	19.53 (7.04)	5.0–35.0	22.59 (8.01)	5.0–35.0

Note. TCS = Transgender Congruence Scale; TC³ = Trans Collaborations Clinical Check-In; GRRS = Gender Identity Reflection and Rumination Scale; GMSR = Gender Minority Stress and Resilience Scale; PHQ-9 = Patient Health Questionnaire (9-Item); GAD-7 = Generalized Anxiety Disorder (7-Item); I-PANAS-SF = International Positive and Negative Affect Scale (Short Form); SWLS = Satisfaction with Life Scale. Ns range from 23–31 due to attrition and missing data.

Table 3 Pearson's Correlations Between TC3 and Other Constructs by Time Point

Measures	TC ³ total			
	Time 1	Time 2	Time 3	Time 4
TCS total	.67***	.71***	.71***	.82***
Appearance congruence	.64***	.66***	.76***	.79***
Gender identity acceptance	.39*	.39*	.229	.531**
GRRS-Reflection	-.02	-.08	-.23	.21
Rumination	-.42*	-.56**	-.47**	-.58**
Preoccupation	-.62***	-.45*	-.58**	-.68***
GMSR Nonaffirmation	-.64***	-.43*	-.79***	-.66***
Internalized transphobia	-.41*	-.51**	-.38*	-.48*
Pride	.40*	.33	.15	.46*
Negative expectations	-.64***	-.58**	-.40*	-.56**
Nondisclosure of gender	-.19	-.51*	-.11	-.18
Community connectedness	.33	.15	.44*	.28
PHQ-9 total	-.29	-.21	-.36	-.43*
GAD-7 total	-.42*	-.43*	-.34	-.47*
I-PANAS-SF				
Positive	-.09	-.00	.11	.26
Negative	-.53**	-.43*	-.33	-.53**
SWLS total	.20	.35	.49**	.63***

Note. TCS = Transgender Congruence Scale; TC³ = Trans Collaborations Clinical Check-In; GRRS = Gender Identity Reflection and Rumination Scale; GMSR = Gender Minority Stress and Resilience Scale; PHQ-9 = Patient Health Questionnaire (9-Item); GAD-7 = Generalized Anxiety Disorder (7-Item); I-PANAS-SF = International Positive and Negative Affect Scale (Short Form); SWLS = Satisfaction with Life Scale. Ns range from 23–31 due to attrition and missing data.

* $p < .05$; ** $p < .01$; *** $p < .001$

Hypothesis 1a: Test-Retest Reliability

The TC³ demonstrated significant stability across measures with an average intraclass correlation coefficient (ICC) of .92 at a 95% confidence interval (CI) from .85–.96, $F(23, 69) = 12.36$, $p < .001$, indicating excellent test-retest reliability.

Hypothesis 1b: Change Over Time

Growth curve analyses were conducted in a MLM framework to test for average change in TC³ scores over time. The intercept only model, without time as a predictor ($TC^3 = \beta_{00} + r_0 + e$), demonstrated that 77% of variability in scores was between participants, and that the

expected correlation between any two time points within-person was .77. There was significant between-subjects variability in mean TC³ over time ($\chi^2(31) = 415.14, p < .001$). Next, *time* was added to the model as a fixed effect ($TC^3 = \beta_{00} + \beta_{10} * Time + r_o + e$). However, there was no significant linear effect of time on TC³ ratings ($t(84) = 1.59, p = .12$), nor did the model demonstrate better fit than the intercept only model ($\chi^2(1) = 2.51, p = .11$). Then, *time* was added to the model as a random effect ($TC^3 = \beta_{00} + \beta_{10} * Time + r_o + r_1 * Time + e$). Again, there was no significant linear effect of time on TC³ ratings when modeled within-persons ($t(31) = 1.60, p = .12$), nor did the model demonstrate better fit than the intercept only model ($\chi^2(3) = 2.53, p > .50$). Relatedly, when *time* was added as a fixed parameter and *QuadTime* was added to the model as fixed ($TC^3 = \beta_{00} + \beta_{10} * Time + \beta_{20} * QuadTime + r_o + e$), there was no significant curvilinear (quadratic) effect of time on TC³ scores ($t(83) = .02, p = .99$), nor did the model demonstrate better fit than the intercept only model ($\chi^2(2) = 2.51, p = .29$). Overall, there was no significant systematic relationship between time and TC³ ratings, suggesting that scores randomly varied around a mean level of TC³ with large within-person correlations between repeated assessments ($r = .77$).

Hypothesis 2: Convergent Validity With TGD-Constructs

Next, correlational analyses at each time point were examined to test for the convergent validity of the TC³ with TGD-specific constructs (see Table 3). Results of these analyses demonstrated reliable positive associations for the TC³ with gender congruence (TCS) including the *Total Score* and *Appearance Congruence*, across all four time points. *Gender Identity Acceptance* was positively associated at all time points except for the third assessment point. Additionally, there were significant negative correlations of the TC³ with the *GRRS Rumination* and *Preoccupation* scales at each time point; however, there were no significant relationships with *GRRS Reflection* scores.

For the distal stress factors of the GMSR, there were reliably significant negative relationships between the TC³ with the *Nonaffirmation* scale for all four assessment points. For the proximal stress factors of the GMSR, there were significant negative relationships between the TC³ and both the *Internalized Transphobia* and *Negative Expectations*

for the Future, whereas *Nondisclosure (Concealment)* was only significantly negatively related at the second time point. The GMSR resilience factors were not as consistently significantly correlated with the TC³ as *Pride* was only significantly positively related at the baseline and final assessment points, whereas *Community Connectedness* was only significantly positively associated at the third time point. Overall, the TC³ demonstrated at least medium ($r \geq .3$) to large ($r \geq .5$) effect sizes on most TGD-related constructs across most time points, except for Gender Identity Reflection and Nondisclosure of Gender Identity.

Hypothesis 3: Criterion Validity With Mental Health Measures

Finally, to assess for the criterion validity, correlations were conducted to examine the relationships between the TC³ and measures of mental health and well-being (see Table 3). Overall, the mental health-related measures exhibited nonsignificant to moderate relationships with the TC³. Both Generalized Anxiety (GAD-7) and Negative Affect (I-PANAS-SF) had significant negative relationships with the TC³ at most time points minus the third time point. Depression as measured by the PHQ-9 was only significantly negatively related at the final time point. Life Satisfaction (SWLS) likewise demonstrated significant positive relationships at the third and fourth assessment points only. Among the nonsignificant relationships for the TC³ with the GAD-7, PHQ-9, and SWLS, all were in the expected direction and exhibited mostly small ($r \geq .1$) to medium ($r \geq .3$) effect sizes though they did not reach significance at all time points ($p < .05$). Negative affect demonstrated medium to large effect sizes across the time points. However, there were no significant relationships between the TC³ and Positive Affect (PANAS).

Discussion

Overall, the TC³ demonstrated excellent internal consistency across time points, exhibiting the overall reliability of the items capturing the intended constructs. Consistent with the first hypothesis (Hypothesis 1a), the TC³ also demonstrated excellent test-retest reliability and was relatively stable over time according to the ICC

estimates. Relatedly, results of the growth curve analyses revealed no systematic change in scores over time which is consistent with Hypothesis 1b. Results pointing to the retention of the intercept only model suggest that scores wax and wane around a mean score, demonstrating some random fluctuation in scores over time; however, the expected correlation between any two time points within-person was .77 suggesting relatively stable scores. This may have been due to the shortened time frame of the study, where there may not be significant systematic fluctuations of gender identity, support, or lived experiences. Longer-term longitudinal research is needed to assess whether there are significant change patterns over time. Another possibility is that the specific inclusion criteria may have prevented those with more pronounced changes in identity and social support factors to participate as we specifically recruited those who identified as TGD and may not have been fully representative of those in the earlier process of gender exploration. What these findings do suggest is that there is some variability in scores that is being captured with the TC³, demonstrating its overall ability to detect changes in identity and support constructs, despite the lack of systematic change over time. Further study of the construct within specific intervention research that aims to increase gender-related well-being is important to understand the potential systematic change in scores that was not captured in the current study. This is particularly important to consider for its use as a clinical progress monitoring measure as it should also be sensitive to changes, while demonstrating test-retest reliability over time.

Our second hypothesis regarding the convergent validity of the TC³ with other TGD-specific measures was mostly supported. These findings reflect many of the same relationships found in the original validation study (Holt et al., 2019). Some correlations that were significant in the previous study were nonsignificant in this sample but that is likely attributable to the small sample size of the current study. In fact all relationships between the TC³ and the other TGD constructs were in the same direction with either similar or greater magnitudes of relationships in the current study compared with the validation sample.

Overall, the TC³ performed particularly well with gender congruence constructs, mostly the subscale related to appearance congruence.

This is not particularly surprising given the number of items related to gender presentation in the TC³ (e.g., comfort with presentation, knew how to present as gender identity). Additionally, it was related to fewer occurrences of being misgendered, suggesting that the ability to live as one's preferred gender is a marker of well-being that may be protective against the use of incorrect pronouns and identity markers in interpersonal interactions out in the community. Likewise, the TC³ was associated with internal resilience factors, namely greater gender-related pride, as well as lower negative cognitions related to gender identity, internalized transphobia, and negative expectations for future events related to gender expression. This highlights the importance of understanding gender-related well-being as a multifaceted cognitive construct that may mitigate the impact of societal stigma on TGD people. Community support and positive thoughts about gender appear to be separate constructs that are not captured in the breadth of items in the TC³. This may be due to differing conceptualizations of social support that is more instrumental (i.e., friends, family, or work/school) versus abstract (i.e., community connectedness) in nature or again attributable to small sample size. Relatedly, positive reflections of one's gender might not be something that one engages in regularly, and may be distinct self-actualizing constructs that are separate from other cognitive constructs of gender identity that may be driven by interpersonal environments.

The final hypothesis regarding the criterion validity of the TC³ with mental health measures was not supported; however, sample size and lack of power should be considered as factors in these nonsignificant relationships. There were overall modest relationships between the TC³ with negative affect and anxiety, in addition to inconsistent relationships with depression and life satisfaction. As the relationships between the TC³ and other mental health measures were generally in the expected direction and had small to medium effect sizes, it is likely lower power from the small sample size that explains these inconsistencies. Thus, the relationships between gender-related well-being and general mental health constructs are mostly consistent with previous findings (Holt et al., 2019), despite their nonsignificance overall. Unlike the findings of the original validation study, there was very little association with positive affect in the current sample. Further examination of the perceived differences between positive and negative

affect for TGD communities would be an interesting area to investigate to identify possible missed relationships or variables that would be better suited to understand protective factors.

As these findings suggest, there are a number of domains of resilience that are important to the well-being of TGD persons that are well captured by the TC³, including the ability to cope with stigma and discrimination, including specific gender-related cognitions, as well as being able to live in their affirmed gender. Thus, practitioners should use this measure to assess overall well-being of TGD clients in conjunction with other constructs for emotional distress (as appropriate to a given case) to have a thorough understanding of relevant domains that may be impacting the overall mental health of their TGD clients in an effort to provide evidence-based, affirming care.

Limitations and Future Directions

As mentioned above, there were a few limitations to the study that are important to note. First, the limited sample size has some added challenges in interpretation of the current findings. While there were some clearly reliable associations among certain measures with the TC³, some of the null findings may be simply due to a lack of power that would be remedied with a larger sample. Relatedly, the correlational results should be interpreted with some caution. Although the sample size between time points (between 21 and 32) has real-world applicability in regard to sample sizes found in clinical research, correlational estimates are more stable in larger sample sizes (Schönbrodt & Perugini, 2013); thus, the correlational results might not be reflective of scale patterns of results in a larger sample. Also, due to internal lower internal consistency at certain time points on the TCS, GRRS, and I-PANAS-SF Positive Affect, there may be the potential for attenuation of correlations, underestimating true effects. Although attenuation can be corrected statistically, this would cloud comparisons across time points. Given that the measures had been used in numerous previous studies and the fact that all had adequate internal consistency at some time points, it was decided to include McDonald's ω to gain a fuller understanding of internal consistency and to increase transparency about constructs at each time point. The size of the sample also limited the statistical power to perform more complex

multivariate analyses to understand the unique contributions of constructs in predicting TC³ scores.

Future studies should further examine the internal consistency of the TC³ and other TGD-specific measures across a larger sample over time to identify underperforming or redundant items. Particular attention should be paid to item nine on the TC³ related to comfort with genitals to examine how it relates to other items in these larger samples as it may perform differently based on many factors (e.g., gender presentation and identity; differences in the amount of dysphoria participants have about their lower body), or to examine whether it is a distinct construct from other items in the measure.

Relatedly, while there was a fairly even distribution of participants who identified as transfeminine, transmasculine, and gender diverse, the sample was mostly white and from urban areas. Future studies should aim to recruit a larger diverse sample with strategic stratification across racial, cultural, and locational factors to not only rectify the present issues with power, but also to examine differences among and between intersecting identities and the experiences that come with such identities.

Additionally, as the intended use of the TC³ is with TGD people in clinical environments, the current sample may not be representative of those in treatment contexts. The specific inclusion and exclusion criteria, while it only excluded three participants from the current pool, may not generalize to treatment-seeking TGD populations. Further for those engaging in treatment, the variability in scores of the general and TGD-specific well-being constructs may change over time as a function of focusing on quality of life improvements and further support and affirmation of gender identity which is not captured in the current general sample who may or may not have been engaged in treatment. Also, the limited timeframe of the current study may not fully capture the variation in constructs captured by the TC³ as it may take longer to notice significant fluctuations in scores. Future studies should assess the sensitivity and reliability of the measure in larger therapeutic and health care contexts over a longer window of time to understand how TC³ scores behave in treatment seeking contexts.

Conclusion

The current study extended the findings of the initial validation study of the TC³ by Holt and colleagues (Holt et al., 2019). While there are important limitations to note, the findings suggest that the TC³ demonstrates excellent test-retest reliability and reliably covaries with expected TGD-specific constructs over time, highlighting its utility as a general measure of TGD lived experiences. Its ability to track gender identity domains serves as a great complement to often implemented general mental health progress monitoring measures to provide effective and affirming evidence-based care for TGD clients. There were less reliable relationships with positive identity and well-being measures which may be due to smaller effects and the modest sample size. Further study is necessary to understand unique aspects of resilience factors that were not fully captured in the current study. More diverse and larger samples in clinical contexts are a logical next step to examine the real world performance of the measure across various experiential aspects, providing for greater generalizability of the current findings.

Acknowledgments We thank the participants and the Trans Collaborations research team, including the TC Local Community Board for their contributions. We would also like to acknowledge the time and effort of our participants whom are crucial to our work, and to express our gratitude for their engagement.

References

- APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *The American Psychologist*, 61(4), 271–285. <https://doi.org/10.1037/0003-066X.61.4.271>
- Austin, A., & Craig, S. L. (2015). Transgender affirmative cognitive behavioral therapy: Clinical considerations and applications. *Professional Psychology, Research and Practice*, 46(1), 21–29. <https://doi.org/10.1037/a0038642>
- Barlow, D. H. (1981). On the relation of clinical research to clinical practice: Current issues, new directions. *Journal of Consulting and Clinical Psychology*, 49(2), 147–155. <https://doi.org/10.1037/0022-006X.49.2.147>
- Barr, S. M., Budge, S. L., & Adelson, J. L. (2016). Transgender community belongingness as a mediator between strength of transgender identity and well-being. *Journal of Counseling Psychology*, 63(1), 87–97. <https://doi.org/10.1037/cou0000127>

- Bauerband, L. A., & Galupo, M. P. (2014). The gender identity reflection and rumination scale: Development and psychometric evaluation. *Journal of Counseling and Development*, 92(2), 219–231. <https://doi.org/10.1002/j.1556-6676.2014.00151.x>
- Benson, K. E. (2013). Seeking support: Transgender client experiences with mental health services. *Journal of Feminist Family Therapy*, 25(1), 17–40. <https://doi.org/10.1080/08952833.2013.755081>
- Bockting, W. O., Miner, M. H., Swinburne Romine, R. E., Hamilton, A., & Coleman, E. (2013). Stigma, mental health, and resilience in an online sample of the U.S. transgender population. *American Journal of Public Health*, 103(5), 943–951. <https://doi.org/10.2105/AJPH.2013.301241>
- Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research*, 25(1), 6–19. <https://doi.org/10.1080/10503307.2013.817696>
- Bowman, S. J., Casey, L. J., McAloon, J., & Wootton, B. M. (2021). Assessing gender dysphoria: A systematic review of patient-reported outcome measures. *Psychology of Sexual Orientation and Gender Diversity*. Advance online publication. <https://doi.org/10.1037/sgd0000486>
- Budge, S. L., Adelson, J. L., & Howard, K. A. (2013). Anxiety and depression in transgender individuals: The roles of transition status, loss, social support, and coping. *Journal of Consulting and Clinical Psychology*, 81(3), 545–557. <https://doi.org/10.1037/a0031774>
- Carlier, I. V., Meuldijk, D., Van Vliet, I. M., Van Fenema, E., Van der Wee, N. J., & Zitman, F. G. (2012). Routine outcome monitoring and feedback on physical or mental health status: Evidence and theory. *Journal of Evaluation in Clinical Practice*, 18(1), 104–110. <https://doi.org/10.1111/j.1365-2753.2010.01543.x>
- Cohn, M. A., Fredrickson, B. L., Brown, S. L., Mikels, J. A., & Conway, A. M. (2009). Happiness unpacked: Positive emotions increase life satisfaction by building resilience. *Emotion*, 9(3), 361–368. <https://doi.org/10.1037/a0015952>
- Crawford, J. R., & Henry, J. D. (2004). The positive and negative affect schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, 43(3), 245–265. <https://doi.org/10.1348/0144665031752934>
- Dacakis, G., Davies, S., Oates, J. M., Douglas, J. M., & Johnston, J. R. (2013). Development and preliminary evaluation of the transsexual voice questionnaire for male-to-female transsexuals. *Journal of Voice*, 27(3), 312–320. <https://doi.org/10.1016/j.jvoice.2012.11.005>
- de Vries, K. M. (2012). Intersectional identities and conceptions of the self: The experience of transgender people. *Symbolic Interaction*, 35(1), 49–67. <https://doi.org/10.1002/symb.2>
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

- Gartner, R. E., & Sterzing, P. R. (2018). Social ecological correlates of family-level interpersonal and environmental microaggressions toward sexual and gender minority adolescents. *Journal of Family Violence*, 33(1), 1–16. <https://doi.org/10.1007/s10896-017-9937-0>
- Gnambs, T. (2014). A meta-analysis of dependability coefficients (test– retest reliabilities) for measures of the Big Five. *Journal of Research in Personality*, 52, 20–28. <https://doi.org/10.1016/j.jrp.2014.06.003>
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement*, 66(6), 930–944. <https://doi.org/10.1177/0013164406288165>
- Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach’s alpha for estimating reliability. *But. . . . Communication Methods and Measures*, 14(1), 1–24. <https://doi.org/10.1080/19312458.2020.1718629>
- Holt, N. R., Huit, T. Z., Shulman, G. P., Meza, J. L., Smyth, J. D., Woodruff, N., MocarSKI, R., Puckett, J. A., & Hope, D. A. (2019). Trans Collaborations Clinical Check-in (TC3): Initial validation of a clinical measure for transgender and gender diverse adults receiving psychological services. *Behavior Therapy*, 50(6), 1136–1149. <https://doi.org/10.1016/j.beth.2019.04.001>
- Hope, D. A., MocarSKI, R., Bautista, C. L., & Holt, N. R. (2016). Culturally competent evidence-based behavioral health services for the transgender community: Progress and challenges. *American Journal of Orthopsychiatry*, 86(4), 361–365. <https://doi.org/10.1037/ort0000197>
- Huit, T. Z., Ralston, A. L., Holt, N. R., Hope, D. A., Puckett, J., MocarSKI, R. A., Haws, J. K., & Woodruff, N. (2021). Psychometric evaluation of the Transgender Congruence Scale. *Sexuality Research & Social Policy*. Advance online publication. <https://doi.org/10.1007/s13178-021-00659-7>
- Javaid, A. (2018). Out of place: Sexualities, sexual violence, and heteronormativity. *Aggression and Violent Behavior*, 39, 83–89. <https://doi.org/10.1016/j.avb.2018.02.007>
- Jones, B. A., Bouman, W. P., Haycraft, E., & Arcelus, J. (2019). The Gender Congruence and Life Satisfaction Scale (GCLS): Development and validation of a scale to measure outcomes from transgender health services. *International Journal of Transgenderism*, 20(1), 63–80. <https://doi.org/10.1080/15532739.2018.1453425>
- King, M., Dinos, S., Shaw, J., Watson, R., Stevens, S., Passetti, F., Weich, S., & Serfaty, M. (2007). The Stigma Scale: Development of a standardised measure of the stigma of mental illness. *The British Journal of Psychiatry*, 190(3), 248–254. <https://doi.org/10.1192/bjp.bp.106.024638>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kozee, H. B., Tylka, T. L., & Bauerband, L. A. (2012). Measuring transgender individuals’ comfort with gender identity and appearance: Development

- and validation of the Transgender Congruence Scale. *Psychology of Women Quarterly*, 36(2), 179–196. <https://doi.org/10.1177/0361684312442161>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Kuper, L. E., Nussbaum, R., & Mustanski, B. (2012). Exploring the diversity of gender and sexual orientation identities in an online sample of transgender individuals. *Journal of Sex Research*, 49(2–3), 244–254. <https://doi.org/10.1080/00224499.2011.596954>
- Lang, A. J., Norman, S. B., Means-Christensen, A., & Stein, M. B. (2009). Abbreviated brief symptom inventory for use as an anxiety and depression screening instrument in primary care. *Depression and Anxiety*, 26(6), 537–543. <https://doi.org/10.1002/da.20471>
- Logie, C. H., James, L., Tharao, W., & Loutfy, M. R. (2012). We don't exist": A qualitative study of marginalization experienced by HIV-positive lesbian, bisexual, queer and transgender women in Toronto, Canada. *Journal of the International AIDS Society*, 15(2), 17392. <https://doi.org/10.7448/IAS.15.2.17392>
- Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care*, 46(3), 266–274. <https://doi.org/10.1097/MLR.0b013e318160d093>
- Löwe, B., Ünlützer, J., Callahan, C. M., Perkins, A. J., & Kroenke, K. (2004). Monitoring depression treatment outcomes with the patient health questionnaire-9. *Medical Care*, 42(12), 1194–1201. <https://doi.org/10.1097/00005650-200412000-00006>
- Maragh-Bass, A. C., Torain, M., Adler, R., Ranjit, A., Schneider, E., Shields, R. Y., Kodadek, L. M., Snyder, C. F., German, D., Peterson, S., Schuur, J., Lau, B. D., & Haider, A. H. (2017). Is it okay to ask: Transgender patient perspectives on sexual orientation and gender identity collection in healthcare. *Academic Emergency Medicine*, 24(6), 655–667. <https://doi.org/10.1111/acem.13182>
- Mizock, L., & Mueser, K. T. (2014). Employment, mental health, internalized stigma, and coping with transphobia among transgender individuals. *Psychology of Sexual Orientation and Gender Diversity*, 1(2), 146–158. <https://doi.org/10.1037/sgd0000029>
- Naeinian, M. R., Shairi, M., Sharifi, M., & Hadian, M. (2011). To study reliability and validity for a brief measure for assessing Generalized Anxiety Disorder (GAD-7).
- Ostir, G. V., Smith, P. M., Smith, D., & Ottenbacher, K. J. (2005). Reliability of the positive and negative affect schedule (PANAS) in medical rehabilitation. *Clinical Rehabilitation*, 19(7), 767–769. <https://doi.org/10.1191/0269215505cr8940a>
- Pavot, W., Diener, E., Colvin, C. R., & Sandvik, E. (1991). Further validation of the Satisfaction with Life Scale: Evidence for the cross-method convergence of well-being measures. *Journal of Personality Assessment*, 57(1), 149–161. https://doi.org/10.1207/s15327752jpa5701_17

- Pavot, W., & Diener, E. (1993). Review of the satisfaction with life scale. *Psychological Assessment*, 5(2), 164-172. <https://doi.org/10.1037/1040-3590.5.2.164>
- Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *The European Health Psychologist*, 16(2), 56-69. <https://doi.org/10.31234/osf.io/h47fv>
- Reese, R. J., Norsworthy, L. A., & Rowlands, S. R. (2009). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy: Theory, Research, Practice, Training*, 46(4), 418-431. <https://doi.org/10.1037/a0017901>
- Schatz, P., & Ferris, C. S. (2013). One-month test-retest reliability of the ImPACT test battery. *Archives of Clinical Neuropsychology*, 28(5), 499-504. <https://doi.org/10.1093/arclin/act034>
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5), 609-612. <https://doi.org/10.1016/j.jrp.2013.05.009>
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini- International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, 59(20), 22-33.
- Sheehan, D., Janavs, J., Baker, R., Sheehan, K. H., Knapp, E., & Sheehan, M. (2015). *Mini international neuropsychiatric interview-version 7.0. o DSM-5*. <https://harmresearch.org/product/mini-international-neuropsychiatric-interview-mini-7-0-2-13/>
- Shpherd, J. C., Green, K. E., & Abramovitz, S. (2010). Transgender clients: Identifying and minimizing barriers to mental health treatment. *Journal of Gay & Lesbian Mental Health*, 14(2), 94-108. <https://doi.org/10.1080/19359701003622875>
- Shulman, G. P., Holt, N. R., Hope, D. A., Mocarski, R., Eyer, J., & Woodruff, N. (2017). A review of contemporary assessment tools for use with transgender and gender nonconforming adults. *Psychology of Sexual Orientation and Gender Diversity*, 4(3), 304-313. <https://doi.org/10.1037/sgd0000233>
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166(10), 1092-1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Testa, R. J., Habarth, J., Peta, J., Balsam, K., & Bockting, W. (2015). Development of the gender minority stress and resilience measure. *Psychology of Sexual Orientation and Gender Diversity*, 2(1), 65-77. <https://doi.org/10.1037/sgd0000081>
- Testa, R. J., Michaels, M. S., Bliss, W., Rogers, M. L., Balsam, K. F., & Joiner, T. (2017). Suicidal ideation in transgender people: Gender minority stress and

- interpersonal theory factors. *Journal of Abnormal Psychology*, 126(1), 125–136. <https://doi.org/10.1037/abn0000234>
- Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology*, 38(2), 227–242. <https://doi.org/10.1177/0022022106297301>
- van den Brink, F., Vollmann, M., & van Weelie, S. (2020). Relationships between transgender congruence, gender identity rumination, and self-esteem in transgender and gender-nonconforming individuals. *Psychology of Sexual Orientation and Gender Diversity*, 7(2), 230–235. <https://doi.org/10.1037/sgd0000357>
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*, 19(1), 231–240.