Authors’ version of submitted paper

Measuring compulsivity as a self-reported multi-dimensional transdiagnostic construct:

 Large-scale (N=182,000) validation of the Cambridge–Chicago Compulsivity Trait Scale

Jeggan Tiego1\*, William Trender2, Peter J. Hellyer3 and, Jon E. Grant4, Adam Hampshire5,

and Samuel R. Chamberlain5

1Turner Institute for Brain and Mental Health and School of Psychological Sciences, Monash University.

2 UKDRI Care Research and Technology Centre, Department of Brain Sciences, Imperial College London, UK

3Neuroimaging Sciences, Institute of Psychiatry Psychology and Neuroscience, King's College London

4Department of Psychiatry, University of Chicago, Chicago, USA.

5DRI Care Research and Technology Centre, and Department of Brain Sciences Imperial College London.

5Department of Psychiatry, Faculty of Medicine, University of Southampton, Southampton, UK; and Southern Health NHS Foundation Trust, NHS, UK.

\*Correspondence regarding this article should be addressed to Jeggan Tiego, Monash Biomedical Imaging, 770 Blackburn Road, Clayton VIC 3168, Australia

October 13th, 2022

Abstract word count: 149

Word Count: 8,181

**Funding and acknowledgements:**

This research was funded in whole, or in part, by Wellcome [110049/Z/15/Z & 110049/Z/15/A]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. Jeggan Tiego was supported by a Turner Impact Fellowship from the Turner Institute for Brain and Mental Health, Monash University. William Trender was supported by funding from the EPSRC through the Imperial College Centre for doctoral training for Neurotechnology. Peter Hellyer was funded by the King's college London Biomedical Research Centre. Prof. Chamberlain’s role in this study was funded by a Wellcome Trust Clinical Fellowship (110049/Z/15/Z & 110049/Z/15/A). The Cambridge–Chicago Compulsivity Trait Scale (CHI-T) is copyright Professor Jon Grant and Professor Sam Chamberlain and may not be used, reproduced, or modified without permission. Please contact the copyright holders for information about how to access and use the scale.

**Conflict of interest statement:**

Dr Grant has received research grants from Otsuka and Biohaven Pharmaceuticals. He receives yearly compensation from Springer Publishing for acting as Editor-in-Chief of the Journal of Gambling Studies and has received royalties from Oxford University Press, American Psychiatric Publishing, Inc., Norton Press, and McGraw Hill.  Prof. Chamberlain receives honoraria from Elsevier for editorial work. Prof. Chamberlain and Prof. Grant are copyright holders for the Cambridge-Chicago Trait Compulsivity Scale (CHI-T).

**Author contribution statements using CRediT:**

Jeggan Tiego was responsible for conceptualization, data curation, formal analysis, methodology, software, visualization, writing - original draft, and writing - review and editing. William Trender and Peter Hellyer were responsible for data curation, investigation, software development, writing - review and editing. Peter Hellyer was responsible for data curation, investigation, software development, writing - review and editing.  Jon Grant was responsible for conceptualization, funding acquisition, resources, supervision, writing - review and editing. Adam Hampshire was responsible for conceptualization, data curation, funding acquisition, investigation, methodology, project administration, resource, software supervision, and writing – review and editing. Sam Chamberlain was responsible for conceptualization, funding acquisition, methodology, project administration, resources, supervision, writing - review and editing.

**Abstract**

Compulsivity has potential transdiagnostic relevance to a range of psychiatric disorders, but it has not been well-characterized and there are few existing measures available for measuring the construct across clinical and non-clinical samples that have been validated at large population scale. We aimed to characterize the multidimensional latent structure of self-reported compulsivity in a population-based sample of British children and adults (*N* = 182,145) using the Cambridge–Chicago Compulsivity Trait Scale (CHI-T). Exploratory structural equation modeling provided evidence for a correlated two-factor model consisting of 1) Perfectionism and 2) Reward Drive dimensions. Evidence was obtained for discriminant validity in relation to the big five personality dimensions and acceptable test-retest reliability. The CHI-T, here validated at extremely large scale, is suitable for use in studies seeking to understand the correlates and basis of compulsivity in clinical and non-clinical participants. We provide extensive normative data to facilitate interpretation in future studies.

**Key words:** compulsivity; transdiagnostic dimensional phenotype; perfectionism; reward drive; psychopathology; exploratory structural equation modeling.

**Introduction**

Compulsivity describes a tendency towards excessive and repetitive behavior that is experienced as urge-driven and is maintained despite being inconsistent with a person’s goals and/or is accompanied by undesirable consequences (Luigjes et al., 2019; Robbins, Gillan, Smith, de Wit, & Ersche, 2012a). While conventionally measured for one disorder per time (using scales such as those capturing criteria for obsessive-compulsive personality disorder [OCPD] or obsessive-compulsive disorder [OCD]), compulsivity can instead be viewed as a transdiagnostic dimensional phenotype (i.e., trait) that is characteristic of several psychiatric disorders and therefore has potential utility for improving understanding of the etiology and treatment of mental disorders (Fineberg et al., 2010; Hollander & Rosen, 2002; Morris & Voon, 2016). The importance of compulsivity as a transdiagnostic dimensional phenotype is highlighted by the incorporation of an obsessive-compulsive and related disorders (OCRDs) grouping in the Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition (DSM-5), encompassing obsessive-compulsive disorder (OCD), body dysmorphic disorder, hoarding disorder, trichotillomania, and excoriation disorder (American Psychiatric Association., 2013). The comorbidity and phenotypic overlap of substance use disorders, eating disorders, and behavioral addictions with OCRDs and each other implicates compulsivity as a common contributing mechanism (Berlin & Hollander, 2014; Gillan, Fineberg, & Robbins, 2017; Tiego et al., 2018). Transdiagnostic and dimensional assessment of compulsivity beyond clinical through subclinical and non-clinical levels of severity has utility for earlier identification and treatment of OCRDs. In this way, such dimensional approaches complement and extend beyond more traditional boundaried (i.e., one disorder per time) categorical approaches towards understanding mental disorders. Furthermore, exploring compulsivity as a transdiagnostic phenotype is consistent with the dimensional framework of the Research Domain Criteria (RDoC), with the potential for biologically-informed discovery of etiological mechanisms (Cuthbert, 2014; Cuthbert & Insel, 2013). In particular, large population-based, data-driven studies are well-positioned to characterize the dimensional structure of psychiatric phenotypes, such as compulsivity (Gillan & Daw, 2016; Gillan et al., 2017).

Compulsivity is a somewhat ambiguous and protean construct that varies in character depending upon the context of assessment (Yücel & Fontenelle, 2012). Compulsivity has – for example – been conceptualized with reference to compulsions in obsessive-compulsive disorder (OCD), which is considered the ‘*archetypal*’ compulsive disorder (Berlin & Hollander, 2014; Chamberlain, Leppink, Redden, & Grant, 2016; Hollander & Benzaquen, 1997). As a result, some existing instruments were developed to measure compulsivity in the context of OCD assessment (Hook et al., 2021). Studies that used scales designed to quantify OCD symptoms specifically as ‘compulsivity’ thus conflate obsessions with compulsions and the content of both as indexed by such scales is clearly specific for OCD and not transdiagnostic. For example, the archetypal compulsions in OCD would not be endorsed for compulsivity in other disorders such as gambling disorder. Furthermore, compulsivity is likely a multidimensional construct; however, there is no consensus on the exact number and nature of the subdimensions comprising the construct in recent work (Fineberg et al., 2014; Fineberg et al., 2010; Gillan et al., 2017).

Several scales have been developed that quantify aspect(s) of compulsivity. The Dimensional Pathological Personality Predicting Personality Disorders Basic Questionnaire (DAPP-BQ) in full form comprises 290 items (Livesley & Jackson, 2009). It was developed primarily for use to measure dimensions of personality disorder in clinical populations (Livesley & Jackson, 2009). Shorter forms exist, but it remains a relatively long assessment (e.g., 136 items) (van Kampen, de Beurs, & Andrea, 2008)). Within this scale, compulsivity relates to rigidity, self-control, and orderliness (Gutiérrez-Zotes et al., 2008). The Personality Inventory for DSM-5 (PID-5) (Krueger, Derringer, Markon, Watson, & Skodol, 2012); comprises 220 items in its original form, and was designed to measure DSM-5 personality disorder tendencies. Within this instrument, compulsivity was generally held to be the inverse of impulsivity (i.e., compulsivity is considered as reflecting ‘excessive’ control)(Krueger et al., 2012). The Personality Inventory for ICD-11 (PICD; 60 items) was developed within the five-domain maladaptive personality trait model (Oltmanns & Widiger, 2018). and includes a dimension for ‘anankastic’ reflecting obsessive-compulsive personality disorder tendencies. This domain was formulated and found to have a negative correlation with external measures of disinhibition; i.e., was associated with higher levels of self-control thusly operationalized (Oltmanns & Widiger, 2018).The Computerized Adaptive Assessment of Personality Disorder (CAT-PD) was developed to model higher and lower order personality traits (Simms et al., 2011), with a focus on Negative Emotionality, Positive Emotionality, Antagonism, Constraint, and Oddity (Simms et al., 2011). In its more recent version, the Comprehensive Assessment of Traits Relevant to Personality Disorder-Static Form (CAT-PD-SF) has 33 domains, with a total of 246 items (Wright & Simms, 2014). It does not have an explicit ‘compulsivity’ domain but does include domains of relevance (e.g., workaholism, rigidity, perfectionism). Another relevant scale is the Five Factor Obsessive Compulsive Inventory (FFOCI), which was developed to measure 12 traits (e.g., inflexibility, risk aversion) relevant to OCPD (Samuel, Riddell, Lynam, Miller, & Widiger, 2012). The FFOCI has 120 items though a shorter form has been developed with 48 items (Griffin et al., 2018).

While valuable, the existing scales have several potential limitations. Extant scales typically have large numbers of items (rendering their use in mainstream research and clinical practice challenging) and have typically not received validation in large samples (i.e., order of tens-of-thousands or higher participants per given study). Scales were largely developed from models of maladaptive personality rather than necessarily also considering how compulsivity is operationalised across what were formerly considered Axis-I disorders (e.g., OCD, gambling disorder, etc.) Another issue is that they were for the most part developed prior to more recent discoveries from the neurosciences, indicating that compulsivity is linked to a loss of top-down control, contrary to the assumption that compulsivity reflects ‘over control’ or the antithesis of impulsivity (Figee et al., 2016; Ioannidis, Hook, Wickham, Grant, & Chamberlain, 2019; Luigjes et al., 2019; van den Heuvel et al., 2016). As such, a conceptualisation of compulsivity should ideally include items of relevance to both reward and more traditional domains.

The Cambridge Chicago Trait Compulsivity Scale (CHI-T) was developed based on extensive expertise from translational research and clinical contexts; this included involvement in extensive studies in normative populations and in clinical disorders, and assessing many patients with (but not limited to) substance use disorders, gambling disorder, OCD, OCPD (and other personality disorders) – and contrasting disorders such as attention-deficit hyperactivity disorder (ADHD). The CHI-T was developed to include common features of compulsivity in keeping with earlier scales, but also to capture a largely overlooked domain implicated in recent compulsivity work (e.g., reward drive). The psychometric properties of the CHI-T were initially previously examined in a small pilot sample of 112 participants recruited via community advertisements, with results suggesting high internal consistency reliability (α = .80) and convergent validity with self-reported obsessive-compulsive symptoms on the Padua inventory, as well as less risk adjustment on the Cambridge Gambling Task (Chamberlain & Grant, 2018). Further validation was conducted by Albertella et al. (2019) in a sample of 260 adults recruited via Mechanical Turk, where total scores on the CHI-T were found to be associated with obsessive-compulsive symptoms (*r* = .52, *p* <.001), problematic usage of the Internet (*r* = .36, *p* <.001), problematic gambling (*r* = .17, *p* =.041), compulsive alcohol use (*r* = .16, *p* =.009), and compulsive eating behavior (*r* = .18, *p* =.004), as well as with family history of obsessive-compulsive disorder and addiction. Interestingly, it was also found that CHI-T total scores were positively associated with attentional capture by reward-related stimuli (*r* = .26, *p* <.001). This finding suggests that the CHI-T is capturing elements of reward sensitivity / reward-seeking previously thought to exclusively define impulsivity at the opposite end of a risk-avoidant/risk-seeking continuum with compulsivity (Hollander, 1993; Hollander & Benzaquen, 1997). However, more recently compulsivity and impulsivity have been recognized as overlapping constructs at phenotypic and neurobiological levels (Chamberlain et al., 2019; Chye et al., 2021; Parkes et al., 2019; Romero-Garcia et al., 2021; Tiego et al., 2020; Tiego et al., 2018). No existing transdiagnostic measures of compulsivity measure reward-seeking as an element of compulsivity. If the CHI-T is capturing a unique subdimension of compulsivity related to reward, this would be a novel finding with potential implications for the study of compulsivity and related disorders. However, the CHI-T is a recently developed instrument and needs validation at population scale, incorporating people with and without mental disorders (Albertella et al., 2020; Hook et al., 2021) .

The purpose of the present study was to characterize self-reported compulsivity as a transdiagnostic dimensional phenotype, as well as establish the psychometric properties of the CHI-T in a very large mixed clinical and non-clinical population-based sample of children and adults. First, we wished to determine the factorial structure of the CHI-T. Compulsivity is theorized to be a multidimensional construct; however, the number and nature of these subconstructs are not well-established bearing in mind more recent neurocognitive and biological findings. This is important to determine, because different subconstructs may have different correlates and functional consequences (Clark & Watson, 2019). For example, there is preliminary evidence to suggest that a ‘Perfectionism’ factor derived from the CHI-T may operate as a protective factor against adverse circumstances, such as stress related to the COVID-19 pandemic (Hampshire et al., 2021). This is based on some preliminary analyses that yielded a three-factor solution; namely 1) Perfectionism – characterized by the need for tasks to be completed and to a high standard; 2) Reward drive – acting on urges and drives to undertake activities that are intrinsically rewarding; and 3) Cognitive rigidity – characterized by repetitive thinking and habitual behavior (Hampshire et al., 2021). However, these results were based on a suboptimal approach using principal components analysis instead of factor analysis (Costello & Osborne, 2005; Howard, 2016). Given the preliminary nature of these analyses we decided to adopt a purely exploratory approach to investigating the factorial structure of the CHI-T.

Second, we aimed to test for measurement invariance of the CHI-T across subgroups within our sample to ensure the scoring metric could be generalized across groups. If factor loadings were equal across groups, then compulsivity had the same substantive interpretation regardless of subgroup. However, if factor loadings differed, then compulsivity as measured by the CHI-T has a different interpretation for each group. Furthermore, differences in factor loadings, intercepts, and item uniquenesses across groups necessitates the use of group-specific norms. We conducted invariance testing across several groups. There are well-documented sex differences in impulsivity and compulsivity, which are theorized to have an evolutionary and biological bases (Cross, Copping, & Campbell, 2011; Fattore & Melis, 2016). Thus, we wished to determine if the CHI-T was measuring compulsivity equivalently across the sexes. Biases are observed in different language versions of a test or measure (Edelen, Thissen, Teresi, Kleinman, & Ocepek-Welikson, 2006). Therefore, we tested whether first-spoken language other than English lead to group differences in measurement properties of the CHI-T. We also tested for differences across participants with and without a psychiatric history, to determine if there were differences in response styles, such as acquiescence (‘yea-saying’), disaquiescence (‘nay-saying’), and extreme (selecting extreme response categories in Likert-type ordinal scales) response styles (Podsakoff, MacKenzie, & Podsakoff, 2012).

We also wished to explore whether there were age-related changes in compulsivity across the life-span, as no studies to date have explored how age may relate to increases and/or decreases in compulsivity in different age groups. For example, impulsivity is known to increase during adolescence before declining in adulthood (Galvan, Hare, Voss, Glover, & Casey, 2007; Spear, 2000). Impulsivity and compulsivity are related constructs (Dalley et al., 2011; Fineberg et al., 2014; Tiego et al., 2018). Thus, it would be interesting to investigate whether compulsivity exhibited the same or different developmental trend as impulsivity, which may be informative with respect to the emergence of psychopathology that is underpinned by elevated levels of compulsivity, such as OCD and habitual drug use.

A notable frequent limitation of previous validation studies of self-report instruments for compulsivity was that they used either exploratory factor analysis (EFA) and/or confirmatory factor analysis (CFA), both of which have substantial limitations. EFA does not provide an index to evaluate model-data consistency. In contrast, CFA imposes unrealistic restrictions on model parameters by constraining non-target loadings to zero as an independent clusters model (ICM) (Marsh, Morin, Parker, & Kaur, 2014). It is not uncommon for CFA-ICM models to provide a very poor fit to the data, even for established constructs, such as the Big Five personality traits (Marsh et al., 2010). Moreover, imposing an independent clusters model on complex multidimensional data can inflate factor intercorrelations, thereby distorting the underlying latent structure (Asparouhov, Muthen, & Morin, 2015; Marsh et al., 2009). Multidimensionality can be fundamental to the appropriate use of an instrument (Marsh et al., 2009; Reise, Bonifay, & Haviland, 2013). It is therefore important to use a technique that preserves the empirical relationships between the multiple constructs measured by a scale. A more viable modern approach to modeling compulsivity questionnaire data is to use a technique that allows for item cross-loadings on non-target factors, such as exploratory structural equation modelling (ESEM) and/or Bayesian structural equation modelling (BSEM) (Asparouhov & Muthén, 2009; Asparouhov et al., 2015). ESEM represents a synthesis of EFA and CFA, enabling evaluation of model fit whilst also accommodating item cross-loadings (Asparouhov & Muthén, 2009; Marsh et al., 2014). Here, we used ESEM because simulation studies have demonstrated that it performs better than BSEM with ordinal data and in large samples (Guo et al., 2019; Liang, Yang, & Cao, 2020).

Using an ESEM approach enabled us to address several other noteworthy limitations of previous validation studies of self-report instruments for compulsivity. First, we report model-based internal consistency reliability estimates (ω) instead of Cronbach’s alpha (α), which does not assume that the items are equally important indices of the underlying construct, taking into account differences in the strength of the factor loading estimates (Hayes & Coutts, 2020). Second, we conducted tests of measurement invariance using the comprehensive taxonomy proposed by Marsh et al. (2009), which evaluates different combinations of invariance across 13 models. Measurement invariance describes the property of a scale, such that the association of the items and the unobserved trait being measured is not dependent on group or measurement occasion (Vandenberg & Lance, 2000). It is a fundamental aspect of construct validity and determines whether findings from a measurement instrument can be generalized across different groups (e.g., sex, ethnicity, language). Third, we obtained follow-up data from 13,702 participants, enabling us to evaluate the test-retest reliability of the CHI-T at the level of latent variables, which is much more accurate than manifest variables (i.e., raw scores) (Marsh et al., 2010). We also obtained concurrent data on the big five personality dimensions in 59,789 participants, enabling us to evaluate the discriminant validity of the CHI-T.

Our primary hypothesis was that ESEM would reveal multiple (i.e., two or more) dimensions of compulsivity as measured by the CHI-T. In particular, based on previous work (Hampshire et al., 2021), we hypothesised three dimensions: 1) Perfectionism – characterized by the need for tasks to be completed and to a high standard; 2) Reward drive – acting on urges and drives to undertake activities that are intrinsically rewarding; and 3) Cognitive rigidity – characterized by repetitive thinking and habitual behavior. Our secondary hypothesis was that self-reported compulsivity would be higher in individuals reporting a psychiatric history of OCD compared to those that did not. We also expected self-reported compulsivity to have discriminant validity from the big five dimensions of personality, particularly Conscientiousness and Neuroticism, as well as Extraversion, Openness, and Agreeableness as indicated by weak to less than moderately strong latent correlations (φ = .100 - .600). Our exploratory hypotheses were that weak, strong, and strict measurement non-invariance would be demonstrated across one or more sample subgroups, necessitating the establishment of group-specific norms for comparison and interpretation in future studies.

**Methods**

**Participants**

A sample of 182,145 (94,385 females; 86,401 males; 1,334 non-binary) individuals aged 9- to 86-years (*M* = 43.82, *SD* = 16.25) participated in this study from the online Great British Intelligence Test (GBIT) study, an ongoing collaborative citizen science project with BBC2 Horizon. A breakdown of the participant sample with respect to demographic and clinical variables is provided in Tables S1 and S2 supplemental materials. Participants were asked to complete a series of cognitive tasks and questionnaires online in the GBIT study. Participants were randomly split into a calibration subsample (*n* = 91,073) and a validation subsample (*n* = 91, 072) for the purposes of cross-validation of the results using invariance testing. As expected, there was no meaningful difference in the mean age (*t* (182,143) = 1.459, *p* = .145; δ = .007 [95%*CI* = -.002, .016]); nor were there meaningful differences in proportion of the sexes (χ2(2) = 3.671, *p* = .160), first spoken language (χ2(1) = 0.146, *p* = .703), ethnicity (χ2(6) = 3.807, *p* = .703), nor education (χ2(5) = 2.069, *p* = .840), between the subsamples. Power for determining a close-fitting model based on the Room Mean Square Error of Approximation (ε) (null hypothesis ε = 00, alternative hypothesis ε < .05) was determined to be greater than .99 even for the largest models with minimum degrees of freedom (Preacher & Coffman, May 2006). The study was approved by the Imperial College Research Ethics Committee (17IC4009) and complied with the American Psychiatric Association ethics standards involving participants. We have complied with the Transparency and Openness Promotion guidelines (see supplemental materials for details).

**Cambridge–Chicago Compulsivity Trait Scale (CHI-T)**

The CHI-T is a 15-item self-report questionnaire that covers broad aspects of compulsivity, including perfectionism, habit formation, cognitive rigidity, intolerance of uncertainty, and reward seeking (for details see (Chamberlain & Grant, 2018)). An intermediate response option was added to the response scale (“*neither agree or disagree*”) from the original validation study, such that participants responded 0 – 4 (0 = “*strongly disagree*”, 1 = “*disagree*”, 2 = “*neither agree nor disagree*”, 3= “*agree*”, 4 = “*strongly agree*”). Broadening the response scale had the advantage of increasing psychometric precision and improving the distributional properties of the data to facilitate analysis using the more efficient robust maximum likelihood estimator compared to the weighted least squares estimator for ordered categorical (i.e., ordinal) data (Finney & DiStefano, 2013; Muthén & Muthén, 1998 - 2017; Simms, Zelazny, Williams, & Bernstein, 2019).

**Big Five Inventory**

Since by ‘trait’ we refer to measures that are transdiagnostic in nature, the most relevant instruments for establishing discriminant validity are personality questionnaires rather than disorder specific instruments. Of these, we chose to include the well-validated Big Five Inventory (BFI), which is probably the most widely used personality inventory around the globe today. The BFI comprises 44 questions assessing Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (John, Robins, & Pervin, 2010). Each question consists of a short phrase and participants endorse their level of agreement with the statement on a 5-point Likert scale (1 ‘*strongly disagree*’ – 5 ‘*strongly agree’*). The BFI is based on the work of Goldberg (1992).

**Statistical Analysis**

**Exploratory Structural equation modelling.**

Preliminary analyses and descriptive statistics were conducted in IBM SPSS Version 27 (IBM Corp, Released 2020). Missing data are summarized in Table S3 of supplemental materials. There were no missing data for CHI-T item 1 and only a very small proportion of responses were missing for CHI-T items 2 – 11 and 13 – 15 (<.005%). However, a large proportion of data (32.83%) was missing for item 12 (domain of scope for improvement / not being good enough) due to this question being excluded from initial data collection (on request from BBC Horizon collaborators). All but two of the covariates (education & employment) were identified as correlates of missingness (see Table S4 supplemental materials). We therefore used multiple imputation (MI) with demographic and clinical information, as well as responses to all CHI-T items which were intercorrelated (see Table S5 supplemental material), incorporated as auxiliary variables in the imputation model to maximize the precision of the imputations (Enders, 2010; Graham, 2009). Parameter estimates, standard errors, and fit statistics were averaged across the 100 imputed datasets (Muthén & Muthén, 1998 - 2017). MI analyses also enabled us to estimate the fraction of missing information (γ), which quantifies sampling variability in the parameter estimates due to missing data (Pan & Wei, 2016; Savalei & Rhemtulla, 2012). All models were estimated in Mplus 8.3 using the covariance matrix (for maximum likelihood estimation) or polychoric correlation matrix (for weighted least square estimation). We used maximum likelihood with robust standard errors (MLR) to estimate the models, which is appropriate for ordinal data with four or more categories (Finney & DiStefano, 2013), and is robust to non-normality (Muthén & Muthén, 1998 - 2017) (see *Figures S1*-*S2*, supplemental materials for CHI-T item distributions). Using the MLR estimator also enabled all 13 levels of invariance in the Marsh et al. (2009) taxonomy to be evaluated. Default settings were used for model estimation. Number of iterations for the H1 model was 2,000 and the convergence criterion was 0.100 D-03. For the rotation algorithm, we used 30 random starts with a maximum number of iterations of 10,000 and a derivative convergence criterion of 0.100D-04.

The chi square (χ2) test statistic is the gold-standard metric for evaluating model-data consistency in structural equation modeling; however, it is overly sensitive to minor model misspecifications in very large samples (Kline, 2015; Marsh, Hau, & Wen, 2004). For this reason, we follow Marsh et al. (2009) in relying upon approximate fit indices to adjudge model fit, including the Root Mean Square Error of Approximation (RMSEA) (ε < .05 close approximate fit; ε = .05 - .08 reasonable approximate fit), Comparative Fit Index (CFI) (>.90 acceptable fit; >.95 good fit), and Standardized Root Mean Squared Residual (SRMR) (<.08 good fit) (Bagozzi & Yi, 2012; Bentler & Bonett, 1980; Browne & Cudeck, 1993; Marsh et al., 2004).

We conducted invariance testing within the ESEM framework (Marsh et al., 2009). Invariance testing is a statistical approach to evaluating generalizability and reproducibility of model parameters across discrete groups by imposing a series of increasingly restrictive equality constraints on the model parameter estimates (Meredith, 1993; Vandenberg, 2002). In the interest of thoroughness, we follow the 13-model taxonomy of Marsh et al. (2009) in testing the invariance of five sets of parameters in various combinations (i.e., factor loadings, intercepts, factor means, item uniquenesses, and factor variance-covariance matrix) and report the results in supplemental material. However, as these 13 levels reflect different combinations of invariance of a small number of key parameters we focus on the 4 levels of invariance traditionally tested within a CFA framework: weak (equality of factor loadings); strong (equality of factor loadings and intercepts, enabling latent mean comparisons between groups); strict (equality of factor loadings, intercepts item uniquenesses, enabling comparison of observed/raw scores between groups); and structural invariance (equality of latent variable variances and covariances) (Grevenstein, 2020).

We tested invariance between two randomly split subsamples to evaluate model reproducibility, as well as invariance across the sexes, first-spoken language, ethnicity, and psychiatric and neurological diagnostic status to determine if results on the CHI-T could be meaningfully compared across these demographic and clinical groups. To evaluate invariance, we compared statistically nested models by change in the RMSEA (ΔRMSEA), CFI (ΔCFI), and SRMR (ΔSRMR) approximate fit indices (Marsh et al., 2009), due to the limitations of chi square difference testing (Δχ2) in very large samples (Sass, Schmitt, & Marsh, 2014). We adopt slightly more conservative cut-offs for the difference in alternative fit indices (ΔAFI) as recommended for ordinal data (ΔRMSEA < 0.010, ΔCFI >.002, & ΔSRMR <.010) (Sass et al., 2014), compared with the conventional guidelines for establishing measurement invariance used in CFA studies with continuous data (ΔRMSEA < 0.015 & ΔCFI >.010) (Chen, 2007; Cheung & Rensvold, 2002). However, because the RMSEA includes a penalty for parsimony, it is also possible for it to improve (i.e., be lower) with the more restricted models (Marsh, 2007). We also note that whilst updated cut-offs for approximate fit indices have been provided for evaluating measurement invariance in the ESEM context with ordinal data, these cut-offs presume the use of weighted least squares mean- and variance-adjusted estimator (WLMSV) rather than the MLR estimator (Jin, 2020). We follow Marsh et al. (2009) in placing less emphasis on information criterion (i.e., Akaike information criterion [AIC] and Bayesian Information criterion [BIC]) that penalize more complex models, which is less relevant for very large sample sizes where the chance of capitalization on sampling variability is low (Marsh et al., 2009). To refer to model parameters in-text we used LISREL notation, which uses lowercase letters of the Greek alphabet to refer to different elements of the model (Jöreskog, 2006). Factor loadings are referred to by the lowercase lambda (λ). Covariances between exogenous (independent) latent variables are referred to by lowercase phi (φ). Regression parameters of endogenous (dependent) on exogenous latent variables are denoted by gamma (γ).

The data analyzed as part of the current study, as well as the Mplus syntax used to model the data, can be obtained by contacting the corresponding author.

**Results**

**Exploratory Structural Equation Modeling**

 Intercorrelations amongst the CHI-T items and overall fit statistics for the competing ESEMs are displayed in Tables S5 and S6. A two-factor model provided the best fit to the data (see Table 1 for the pattern of factor loadings & *Figure S3* supplemental materials). It had a clear and interpretable loading structure, with the two factors only weakly correlated (φ = .272, *SE* = .014, [95%*CI* = .244, .299], *p* <.001). With exploratory factor modeling it is possible to obtain better-fitting models in terms of global fit statistics simply by increasing the number of factors. However, this can result in overparameterized models with nonsensical factor loading structures and junk factors that do not capture meaningful variance.

To ensure the results were robust to estimation method, we re-estimated the competing models with the WLSMV estimator, which is optimized for ordinal data. Tables S13 – S18 display the correlation residuals for the competing models based on estimation with the WLSMV estimator, where a pattern of residuals > .100 indicates poor local fit (Kline, 2015). These tables show that the three- and four-factor models had more correlation residuals > .100 (18 and 8, respectively compared to 2 for the two-factor model), indicating that they did not reproduce the bivariate relationships between the items as well as the two-factor despite being less parsimonious (Goodboy & Kline, 2017; Kline, 2015). Furthermore, the four-factor model was characterized by a factor with primary loadings from only two items, suggesting it did have a sensible loading structure. Similarly, the bifactor one model had more correlation residuals > .100 (3). Whereas the other two bifactor models had no correlation residuals > .100, this was clearly the result of overparameterization. The factor loadings structures revealed insufficient support for a general factor in all cases, because it did not have moderate to strong primary loadings from all items (Bornovalova, Choate, Fatimah, Petersen, & Wiernik, 2020; Markon, 2019). Furthermore, the bifactor model with three specific factors was associated with a junk factor with only two items with primary loadings. Additionally, unidimensionality statistics computed for the bifactor models, specifically the explained common variance (ECV) and percentage of uncontaminated correlations (PUC) did not support collapsing the CHI-T into a unidimensional scale (ECV < .70 + PUC < .70) (Rodriguez, Reise, & Haviland, 2016a, 2016b). In combination, these results supported the two-factor model as the most parsimonious model, which provided the best fit to the CHI-T item data.

The two-factor model accounted for approximately 33% of the item variance. The first factor captured approximately 22.3% of the item variance and was characterized by moderately strong loadings from items 2 (doing things just right), 3 (repetition to achieve high standards), 13 (completion leading to soothing), 10 (difficulty moving between tasks if not perfect) and 1 (dislike of non-completion of tasks) and was labelled ‘Perfectionism’. Marker variables for the second factor included items 9 (doing things that are immediately rewarding), 8 (difficulty not acting on urges), and 6 (addictive propensity), with weaker loadings from items 7 (rigidity) and 4 (getting stuck with thoughts). This factor was therefore labelled ‘Reward Drive’ and captured approximately 10.2% of the item variance.

Based on the pattern of factor loadings observed across the sample (Table 1) and within subsamples (see Tables S47 – S52 supplemental material) we propose the following scoring approach for the CHI-T: items 1, 2, 3, 5, 10, 11, 12, 13, and 15 are summed to yield a raw score for Perfectionism and items 6, 8, 9 are summed to yield a raw score for Reward Drive. We calculated the reliability of the two subscales using omega (ω) which is a model-based reliability estimate taking into account the variable importance of the indicators according to their standardized factor loadings (Hayes & Coutts, 2020; Rodriguez et al., 2016b). Reliability estimates, based on this raw scoring approach, were adequate both for Perfectionism (ω = .80) and Reward Drive (ω = .63). We provide norms for the raw scores of both scales (see supplementary tables S53 – S58). However, we also provide norms for the factor-based scores (see Tables S59 – S64 supplemental material), which take into account the full loading structure of all 15 CHI-T items on the Perfectionism and Reward Drive factors. Further validation of the two-factor model, including construct replicability are provided in supplemental materials.

**Age-Related Effects**

We estimated latent polynomial growth models incorporating orthogonal linear, squared, cubic, and quadratic effects of age to determine how self-reported compulsivity changed across the age range 9 – 86-years of age (Marsh et al., 2009). There was a complex interaction of Perfectionism with age, such that there were positive linear (γ = .204, *SE* = .007, [95%*CI* = .191, .218], *p* <.001) and cubic (γ = .666, *SE* = .011, [95%*CI* = .644, .689], *p* <.001) effects and negative squared (γ = -.657, *SE* = .008, [95%*CI* = -.673, -.640], *p* <.001) and quadratic (γ = -.230, *SE* = .011, [95%*CI* = -.252, -.207], *p* <.001) effects. Thus, Perfectionism appeared to increase from ages 9 through to around 30, before decreasing gradually over older ages through to 86-years old. In contrast, there was only a weak positive linear effect of age on Reward Drive (γ = .232, *SE* = .016, [95%*CI* = .201, .262], *p* =.068). We examined the relationship of the education and income background variables using a multiple indicators multiple causes (MIMIC) model, because each consisted of too many categories for a multigroup analysis (Marsh et al., 2009). Education had no meaningful effect on either Perfectionism (γ = -.008, *SE* = .004 [95%*CI* = -.015, -.001], *p* = .020) or Reward Drive (γ = -.021, *SE* = .004 [95%*CI* = -.029, -.013], *p* < .001), nor did level of income on either factor (γ = -.009, *SE* = .004 [95%*CI* = -.016, -.002], *p* = .001 & γ = -.018, *SE* = .004 [95%*CI* = -.026, -.009], *p* < .001).

**Invariance Testing**

The results of invariance testing are summarized in Table 2 and Tables S39 to S46 of supplemental materials. In brief, we found that the two-factor ESEM was fully invariant (i.e., invariance of the factor loadings, item uniquenesses, factor variance-covariances, item intercepts, and factor means) across the randomly split subsamples, indicating high stability and reproducibility. Full invariance was also established across participants with English and non-English as their first-spoken language, as well as those with and without a history of a neurological condition. In contrast, we found that the CHI-T was sex non-invariant, such that the unstandardized factor loadings and intercepts varied across males, females, and non-binary participants. This indicates that compulsivity differs across genders, necessitating gender-specific norms. The variable importance of specific indicators to defining the Perfectionism and Reward drive constructs can be seem in the group-specific factor loadings provided in Tables S47 – S52 supplemental material. Weak invariance (equality of factor loadings) across participants with and without a psychiatric history was established, indicating that the CHI-T can be used to measure compulsivity in a dimensional and transdiagnostic fashion. However, there was evidence of non-invariance of the intercepts, suggesting that participants with history of one or more psychiatric conditions exhibited systematic differences in individual item level responding to the CHI-T independently of their standing on the underlying latent variables (e.g., response bias). Caution should be exercised when comparing raw scores across subgroups. We provide the factor loadings (Tables S47 – S52) and SPSS syntax as well as extensive normative data (Tables S53 – S64) in supplemental materials to facilitate scoring, comparison, and interpretation in future studies.

We established strong invariance of the two-factor ESEM model between participants with and without a history of OCD (χ2*M* (193) = 52249.702, *p* < .001; RMSEA*M* = .054; CFI*M* = .901; SRMR*M* = .040; Model 5 – Model 2 ΔRMSEA = -.002, ΔCFI = -.001; ΔSRMR = .050), enabling comparison of latent means. As would be expected, we found that Perfectionism (*Z* = 12.217, *p* < .001) and Reward Drive (*Z* = 16.424, *p* < .001) were substantially higher in participants with a history of OCD compared to those without a history of OCD.

**Test-Retest Reliability**

We evaluated test-retest reliability in a subsample of 13,702 participants that completed the CHI-T over two successive timepoints over 6-12 months by correlating their raw scores for the two measurement occasions. Overall test-retest reliability for Perfectionism was *r* = .77, [95%*CI* =.76,.78], *p* < .001, *n* = 9,990 and for Reward Drive *r* = .69, [95%*CI* = .68, .70], *p* < .001, *n* = 13,702. For participants without a psychiatric history, stability coefficients were also acceptable for males (Perfectionism *r* = .75, [95%*CI* = .74,.76], *p* < .001, *n* = 3,357; Reward Drive *r* = .69, [95%*CI* = .68, .70], *p* < .001, *n* = 5,207); females (Perfectionism *r* = .78, [95%*CI* = .77, .79], *p* < .001, *n* = 4,750; Reward Drive *r* = .68, [95%*CI* = .67, .69], *p* < .001, *n* = 6,578); and non-binary (Perfectionism *r* = .67, [95%*CI* = .42, .92], *p* < .001, *n* =22; Reward Drive *r* = .77, [95%*CI* = .66, .88], *p* < .001, *n* = 56) participants. For participants with a psychiatric history, stability coefficients were also acceptable for males (Perfectionism *r* = .74, [95%*CI* = .70, 0.78], *p* < .001, *n* = 614; Reward Drive *r* = .68, [95%*CI* = .64, .72], *p* < .001, *n* = 614); females (Perfectionism *r* = .77, [95%*CI* = .75, .79], *p* < .001, *n* =1,231; Reward Drive *r* = .67, [95%*CI* = .64, .70], *p* < .001, *n* =1,231); and non-binary participants for Perfectionism (*r* = .77, [95%*CI* = .53, 1.00], *p* = .001, *n* =14), but not Reward Drive (*r* = .45, [95%*CI* = -.02, .92], *p* =.103, *n* = 14); however, this may be due to small sample size. We also evaluated test-retest reliability using an ESEM approach (Marsh et al., 2010; Muthén & Muthén, 1998 - 2017). Not surprisingly, the results revealed stability coefficients that were higher than raw scores (φ = .79 - .85), because latent variables are not subject to attenuation bias due to unreliability in contrast to raw scores. Detailed results are reported in supplemental material.

**Discriminant Validity**

To examine the discriminant validity of the CHI-T in relation to normal personality variables, we conducted bivariate correlation analyses between the CHI-T subscale raw scores and raw scores obtained using the Big Five inventory (Goldberg, 1993; John & Srivastava, 1999) in a subset of participants (*n* = 59,790) from whom we collected data on both measures . Perfectionism raw scores were only weakly to modestly correlated with the Big Five dimensions of personality, most strongly to Conscientiousness (*n* = 59, 789, Neuroticism *r* = .20, [95%*CI* = .20,.21], *p* <.001; Extraversion *r* = -.08 [95%*CI* = -.09, -.07], *p* <.001; Openness *r* = -.03, [95%*CI* = -.04, -.024], *p* <.001; Agreeableness *r* = -.11, [95%*CI* = -.12, -.10], *p* <.001; Conscientiousness *r* = .28, [95%*CI* = .27, .28], *p* <.001), as were Reward Drive raw scores (*n* = 59, 790, Neuroticism *r* = .26, [95%*CI* = .25, .26], *p* <.001; Extraversion *r* = .051 [95%*CI* =.04, .06], *p* <.001; Openness *r* = .037, [95%*CI* = .03, .05], *p* <.001; Agreeableness *r* = -.21, [95%*CI* =- .22, -.21 ], *p* <.001; Conscientiousness *r* = -.39, [95%*CI* = -.40, -.38], *p* <.001). Thus, the CHI-T subscale shared less than 16% variance with any of the Big 5 dimensions. We also conducted these analyses using an ESEM approach because of the strong item cross-loading for the scales of both the CHI-T and Big 5 measurement instruments (Marsh et al., 2010). These results are presented in supplemental material and were generally comparable in demonstrating discriminant validity, with the highest associations with Conscientiousness and the two CH-T subscales. The difference was that the strengths of the associations were generally stronger using the ESEM approach, which enables the empirical overlap between these scales to be estimated without attenuation bias due to error variance in the items.

**Discussion**

 Categorical models have limited utility for understanding the genetic and neurobiological mechanisms of OCRDs and related disorders, including substance use disorders and behavioral addictions; and for understanding variation in treatment response in clinical disorders (Gillan et al., 2017). Instead, compulsivity can be characterized and measured as a transdiagnostic dimensional phenotype, which can facilitate insights into the etiology and treatment of compulsive disorders (Robbins, Gillan, Smith, de Wit, & Ersche, 2012b). However, progress in this endeavor has been limited by the lack of psychometrically-sound self-report instruments that incorporate more recent findings from the neurosciences such as a role for reward and the notion that disinhibition (rather than ‘increased inhibition’ is characteristic of compulsivity. Previous studies indicated the need for further evaluation of the CHI-T in clinical and non-clinical cohorts (Albertella et al., 2020). Here, we undertook an extensive evaluation of the psychometric properties of the CHI-T to examine its suitability for dimensional and transdiagnostic assessment of compulsivity in future studies. We validated the scale in a population-based sample that was diverse and included people with and without psychiatric and neurological conditions, as well as across a broad age range, including adults and young people.

A strength of the present study compared to previous validation studies of self-report compulsivity measures was that we leveraged data from a very large, population-based sample and utilized advanced statistical techniques, including ESEM and invariance testing (Marsh et al., 2009; Vandenberg & Lance, 2000). The use of ESEM enabled us to recover the complex two-factor structure of the CHI-T without imposing unrealistically restrictive constraints on item cross-loadings, which would have distorted measurement of the underlying factors, including their intercorrelation (Marsh et al., 2014; Marsh et al., 2009). We also used a hybrid approach to examining the generalizability of the psychometric properties of the CHI-T across discrete groups, combining multigroup invariance testing and MIMIC models (Marsh et al., 2009). ESEM also facilitated examination of the discriminant validity of the CHI-T compulsivity factors with respect to the Big Five personality dimensions. These procedures allowed us to provide extensive validation of the psychometric properties of the CHI-T.

The two-factor structure of transdiagnostic compulsivity as measured by the CHI-T, consisting of nearly orthogonal (1) Perfectionism and (2) Reward drive dimensions, parallels previous research of self-reported compulsivity suggesting this multidimensional construct includes both a ‘risk-avoidant’ and a ‘reward-seeking’ dimension, the latter which overlaps with impulsivity (Chamberlain et al., 2019; Tiego et al., 2018). Compulsivity has been theorized to reflect the end of a behavioral continuum, in which previously rewarded ‘impulsive’ behaviors become increasingly habitualized through mechanisms of reinforcement learning (Dalley et al., 2011). Thus, these two constructs, Perfectionism and Reward drive, may contribute to negative and positive reinforcement mechanisms, respectively, both of which play crucial roles in the instantiation of compulsivity across disorders. Measuring compulsivity subconstructs as dimensional phenotypes in future studies may yield important insights into the neurobiology of compulsivity and could also be included in clinical trials to identify candidate treatments capable of operating trans-diagnostically (Dalley et al., 2011).

There are also parallels between the factor structure of the CHI-T and facets of existing compulsivity measures. For example, our Perfectionism factor parallels the Perfectionism subscales of the CAT-PD and PIC-D the Rigid Perfectionism scale of the PID-5 (Krueger et al., 2012; Oltmanns & Widiger, 2018; Simms et al., 2011). There are also parallels between the CHI-T items assessing cognitive rigidity and the rigidity and inflexibility dimensions assessed by the CAT-PD, DAPP-BQ, and the FFOCI (Livesley & Jackson, 2009; Samuel et al., 2012; Simms et al., 2011; van Kampen et al., 2008). These conceptual overlaps suggest that perfectionism and cognitive and behavioral rigidity are core components of a multi-dimensional compulsivity construct. Unique to the CHI-T is the reward drive dimension, which assesses addictive propensity, sensitivity to reward and proclivity to act on urges. No other existing transdiagnostic measures of compulsivity assess this dimension, likely because reward-seeking has historically been conceptualized at the opposite end of the risk-seeking/harm avoidance continuum to compulsivity (Hollander, 1993; Hollander & Benzaquen, 1997). However, more recent evidence suggested that impulsivity and compulsivity are overlapping constructs sharing a disinhibitory component, indicating that they may share phenomenological features including sensitivity to reward (Chamberlain et al., 2019; Tiego et al., 2020; Tiego et al., 2018). For example, compulsivity may incorporate a component of ‘urgency’ in which individuals experience urges to engage with rewarding stimuli when experiencing, and to cope with, strong emotions; and experience ‘attentional grab’ for relevant environmental stimuli (Chamberlain et al., 2019; Cyders & Smith, 2008; Tiego et al., 2020; Tiego et al., 2018). This would be in keeping with previous findings that showed a correlation between scores on the CHI-T and attentional capture by reward-related stimuli (Albertella et al., 2020). Future research is needed to investigate the criterion validity of the Reward Drive dimension of the CHI-T, as well assess the degree of empirical overlap with existing transdiagnostic dimensional compulsivity scales, particularly the Perfectionism factor.We found that the measurement properties of the CHI-T were fully invariant across subsamples, participants who spoke English as their first language and those that did not, as well as those with and without a history of a neurological history. However, the strength of factor loadings and the intercepts varied as a function of sex and ethnicity. From a theoretical perspective, these results indicate that Perfectionism and Reward drive subscales can have different statistical interpretations across the sexes and ethnic groups, because the linear weights of the items contributing to the factor variances differ between male, female, and non-binary persons and participants of different ethnicity. This finding is interesting as it may speak to population level sex-related differences in compulsivity that should be explored in future work. From a practical perspective, this means that group-specific norms are desirable for interpreting findings. This is likely to be a pervasive issue across many mental health scales, rather than being specific to the CHI-T: it is just that studies have seldom addressed and highlighted this psychometric issue for other scales. However, rigorous direct comparisons are possible using differential item functioning analysis within an item response theory framework in which subsamples are scaled on the same latent trait continuum (Tay, Meade, & Cao, 2015). Although the Perfectionism and Reward drive factors were measured equivalently across individuals with and without a psychiatric history, we found group differences in the intercepts. This suggests systematic differences in item responding independent of the underlying latent variable, for example acquiescence bias, negative, or extreme responding. Again, this is likely to be the case for many mental health scales and is unsurprising. However, researchers can use latent variable modeling techniques to measure transdiagnostic compulsivity across the latent trait continuum with high precision (Bollen & Noble, 2011). Critically, the latent trait distributions for the two factors were approximately normal suggesting that compulsivity as measured by the CHI-T is optimally represented as a transdiagnostic and fully dimensional phenotype. For pragmatic purposes, we suggest that researchers and clinicians can use the scale across populations/groups without needing “group-specific norms” as long as they are mindful of these issues, as would be the case for other mental health scales; though the steps suggested above would help to make such findings more rigorous statistically.

 Crucially, we also determined that the two compulsivity factors measured by the CHI-T have discriminant validity in terms of the Big Five personality dimensions. The maximum amount of shared variance was less than 15% based on the upper bound of the 95% confidence interval of the latent correlations, suggesting that the CHI-T is clearly measuring phenotypes distinct from conventional personality domains. Interestingly, and as may be expected, there were moderate positive associations between Perfectionism and Conscientiousness and between Reward drive and Neuroticism. Future work will be able to further clarify the nature and correlates of the Perfectionism and Reward drive compulsivity dimensions articulated in the present study.

**Limitations**

 A limitation of the present study was that we modified the response scale to 5 response categories from the previous 4-point response scale. This had the advantage of increasing psychometric precision and improving the distributional properties of the item data to enable the use of maximum likelihood, which is a more efficient estimator than those used for ordinal data (e.g., weighted least squares). However, this means that the results of the current study cannot necessarily be compared directly to previous results. Future studies should adopt the 5-point Likert scale for all future applications of the CHI-T.

 Another limitation is inherent to the field: there is a lack of contemporary consensus on definitions (and consequently lack of agreement about likely subdomains) (see e.g., Luigjes et al., 2019), albeit lessons can be gleaned from the personality disorder literature and earlier compulsivity instruments as described in the introduction. As such, it is possible future work may identify other subdomains relevant to compulsivity that are not in the CHI-T. At the same time, we feel the scale contains comprehensive facets of compulsivity and is completely in keeping with the most recent ‘high level’ definitions of the term (e.g., Luigjes et al., 2019) whilst also offering richness and depth of measurement beyond that which would be captured by a single question reflecting a high-level definition. A further limitation of the study was that the we were missing a large portion of data for item 12 (*n* = 59, 796, 33%). Multiple imputation is considered to be a gold-standard approach to handling missing data (Enders, 2010). Furthermore, we followed recommended guidelines for imputing the missing data, including the use of many auxiliary variables that were associated with missingness, as well as generating a large number of imputations (Graham, 2009; Lang & Little, 2016). However, the amount of missing data can lead to variability of parameter estimates across imputations, which is quantified as the fraction of missing information (Pan & Wei, 2016; Savalei & Rhemtulla, 2012). It can be seen in Tables 1 and S7 – S11 that the factor loading estimates for item 12 had a high fraction of missing information which equates to a higher degree of uncertainty in the actual population parameter. Thus, the missing data introduces some indeterminacy and the results of our factor analysis should be interpreted with some caution.

 As traits are transdiagnostic in nature, we felt that personality instruments were the most relevant to establish discriminant validity. Since we were limited in the number of questionnaire items we could include in this large-scale citizen science study, we included the Big Five Inventory – this being one of the most widely used scales. Future work should also examine discriminant validity against other personality instruments. A further potential limitation (from the perspective of ease of interpretation) is that measurement of the Perfectionism and Reward drive constructs using the CHI-T was non-invariant across the sexes and participants with and without a psychiatric history. We found that the unstandardized factor loadings and intercepts varied across males, females, and non-binary participants. This indicates that compulsivity differs across the sexes, with different interpretations of the Perfectionism and Reward drive constructs. This is not surprising given that gender differences are already established for other traits (Feingold, 1994). The differences were most notable for non-binary participants (see Tables S47 – 52), where their factor loadings for items 1, 2, 12, and 13 on Perfectionism and items 8 and 9 on Reward drive were larger, indicating that these constructs were more strongly defined by these items. Further analysis, such as with differential item functioning, would be needed to understand the sources and causes of these differences (Stark, Chernyshenko, & Drasgow, 2006). Furthermore, participants with and without a psychiatric history differed on the intercepts, suggesting systematic differences in responding to items independent of the underlying latent variable. The reasons for differences in responding are not clear from this study, but may be attributable to differences in response styles, such as acquiescence (‘yea-saying’), disaquiescence (‘nay-saying’), and extreme (selecting extreme response categories in Likert-type ordinal scales) response styles (Podsakoff et al., 2012).

**Conclusions**

Compulsivity can be viewed as an important transdiagnostic construct implicated in several psychiatric disorders, including OCD, addictions, and personality disorders, as well as being relevant to understanding normative behaviour. However, there is a paucity of self-report measures available for measuring compulsivity transdiagnostically in non-clinical as well as clinical populations that have been validated at large scale, and incorporate domains implicated from the neurosciences. The CHI-T, here validated at extremely large scale, is suitable for use in studies seeking to understand the correlates and basis of compulsivity in clinical and non-clinical participants. We provide extensive normative data to facilitate interpretation in future studies.

**References**

Albertella, L., Chamberlain, S. R., Le Pelley, M. E., Greenwood, L. M., Lee, R. S., Den Ouden, L., . . . Yucel, M. (2020). Compulsivity is measurable across distinct psychiatric symptom domains and is associated with familial risk and reward-related attentional capture. *CNS Spectr, 25*(4), 519-526. doi:10.1017/S1092852919001330

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. Washington, DC: American Psychiatric Association.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural equation modeling: a multidisciplinary journal, 16*(3), 397-438. doi:10.1080/10705510903008204

Asparouhov, T., Muthen, B., & Morin, A. J. S. (2015). Bayesian Structural Equation Modeling With Cross-Loadings and Residual Covariances: Comments on Stromeyer et al. *Journal of Management, 41*(6), 1561-1577. doi:10.1177/0149206315591075

Bagozzi, R. P., & Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the academy of marketing science, 40*(1), 8 - 34.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(1), 588 - 606.

Berlin, G. S., & Hollander, E. (2014). Compulsivity, impulsivity, and the DSM-5 process. *CNS Spectrums, 19*(1), 62 - 68. doi:10.1017/S1092852913000722

Bollen, K. A., & Noble, M. D. (2011). Structural equation models and the quantification of behavior. *Proceedings of the National Academy of Sciences, 108*(Supplement 3), 15639-15646. doi:10.1073/pnas.1010661108

Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate use of bifactor analysis in psychopathology research: Appreciating benefits and limitations. *Biological Psychiatry, 88*(1), 18-27. doi:10.1016/j.biopsych.2020.01.013

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136 -162 ). Newbury Park, CA: Sage.

Chamberlain, S. R., & Grant, J. E. (2018). Initial validation of a transdiagnostic compulsivity questionnaire: The Cambridge–Chicago Compulsivity Trait Scale. *CNS Spectrums, 23*(5), 340-346. doi:10.1017/S1092852918000810

Chamberlain, S. R., Leppink, E. W., Redden, S. A., & Grant, J. E. (2016). Are obsessive–compulsive symptoms impulsive, compulsive or both? *Compr Psychiatry, 68*(Supplement C), 111-118. doi:<https://doi.org/10.1016/j.comppsych.2016.04.010>

Chamberlain, S. R., Tiego, J., Fontenelle, L. F., Hook, R., Parkes, L., Segrave, R., . . . Yucel, M. (2019). Fractionation of impulsive and compulsive trans-diagnostic phenotypes and their longitudinal associations. *Aust N Z J Psychiatry, 53*(9), 896-907. doi:10.1177/0004867419844325

Chen, F. 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:Doi 10.1207/S15328007sem0902\_5

Chye, Y., Suo, C., Romero-Garcia, R., Bethlehem, R. A. I., Hook, R., Tiego, J., . . . Chamberlain, S. R. (2021). Examining the relationship between altered brain functional connectome and disinhibition across 33 impulsive and compulsive behaviours. *The British Journal of Psychiatry*, 1-3. doi:10.1192/bjp.2021.49

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychol Assess, 31*(12), 1412-1427. doi:10.1037/pas0000626

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.

Cross, C. P., Copping, L. T., & Campbell, A. (2011). Sex differences in impulsivity: a meta-analysis. *Psychological Bulletin, 137*(1), 97-130. doi:<https://doi.org/10.1037/a0021591>

Cuthbert, B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry, 13*(1), 28-35. doi:10.1002/wps.20087

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Med, 11*, 126. doi:10.1186/1741-7015-11-126

Cyders, M. A., & Coskunpinar, A. (2011). Measurement of constructs using self-report and behavioral lab tasks: Is there overlap in nomothetic span and construct representation for impulsivity? *Clinical Psychology Review, 31*, 965-982. doi:10.1016/j.cpr.2011.06.001

Cyders, M. A., & Smith, G. T. (2008). Emotion-based dispositions to rash action: Positive and negative urgency. *Psychol Bull, 134*(6), 807-828. doi:10.1037/a0013341

Dalley, J. W., Everitt, B. J., & Robbins, T. W. (2011). Impulsivity, compulsivity, and top-down cognitive control. *Neuron, 69*(4), 680 - 694. doi:10.1016/j.neuron.2011.01.020

Edelen, M. O., Thissen, D., Teresi, J. A., Kleinman, M., & Ocepek-Welikson, K. (2006). Identification of differential item functioning using item response theory and the likelihood-based model comparison approach: Application to the Mini-Mental State Examination. *Medical Care, 44*(11), S134-S142.

Eisenberg, I. W., Bissett, P. G., Zeynep, E. A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications, 10*(1), 2319. doi:10.1038/s41467-019-10301-1

Enders, C. K. (2010). *Applied missing data analysis.* New York: The Guilford Press.

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, 201818430. doi:10.1073/pnas.1818430116

Fattore, L., & Melis, M. (2016). Sex differences in impulsive and compulsive behaviors: a focus on drug addiction. *Addiction Biology, 21*(5), 1043-1051. doi:<https://doi.org/10.1111/adb.12381>

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116*(3), 429-456. doi:<https://doi.org/10.1037/0033-2909.116.3.429>

Figee, M., Pattij, T., Willuhn, I., Luigjes, J., van den Brink, W., Goudriaan, A., . . . Denys, D. (2016). Compulsivity in obsessive-compulsive disorder and addictions. *Eur Neuropsychopharmacol, 26*(5), 856-868. doi:10.1016/j.euroneuro.2015.12.003

Fineberg, N. A., Chamberlain, S. R., Goudriaan, A. E., Stein, D. J., Vanderschuren, L. J. M. J., Gillan, C. M., . . . Potenza, M. N. (2014). New developments in human neurocognition: Clinical, genetic, and brain imaging correlates of impulsivity and compulsivity. *CNS Spectrums, 19*(1), 69 - 89. doi:10.1017/S1092852913000801

Fineberg, N. A., Potenza, M. N., Chamberlain, S. R., Berlin, H. A., Menzies, L., Bechara, A., . . . Hollander, E. (2010). Probing compulsive and impulsive behaviors, from animal models to endophenotypes: A narrative review. *Neuropsychopharmacology, 35*, 591 - 604. doi:10.1038/npp.2009.185

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In *Structural equation modeling: A second course, 2nd ed.* (pp. 439-492). Charlotte, NC, US: IAP Information Age Publishing.

Galvan, A., Hare, T., Voss, H., Glover, G., & Casey, B. (2007). Risk‐taking and the adolescent brain: Who is at risk? *Developmental Science, 10*(2), F8-F14. doi:<https://doi.org/10.1111/j.1467-7687.2006.00579.x>

Gillan, C. M., & Daw, N. D. (2016). Taking psychiatry research online. *Neuron, 91*(1), 19 - 23. doi:<http://dx.doi.org/10.1016/j.neuron.2016.06.002>

Gillan, C. M., Fineberg, N. A., & Robbins, T. W. (2017). A trans-diagnostic perspective on obsessive-compulsive disorder. *Psychol Med, 47*(9), 1528-1548. doi:10.1017/S0033291716002786

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26.

Goodboy, A. K., & Kline, R. B. (2017). Statitical and practical concerns with published communication research featuring structural equation modeling. *Communication Research Reports, 34*(1), 68-77.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*(1), 549-576. doi:10.1146/annurev.psych.58.110405.085530

Grevenstein, D. (2020). Factorial validity and measurement invariance across gender groups of the German version of the Interpersonal Reactivity Index. *Measurement Instruments for the Social Sciences, 2*(1), 8. doi:10.1186/s42409-020-00015-2

Griffin, S. A., Suzuki, T., Lynam, D. R., Crego, C., Widiger, T. A., Miller, J. D., & Samuel, D. B. (2018). Development and Examination of the Five-Factor Obsessive-Compulsive Inventory–Short Form. *Assessment, 25*(1), 56-68. doi:10.1177/1073191116643818

Guo, J., Marsh, H. W., Parker, P. D., Dicke, T., Lüdtke, O., & Diallo, T. M. O. (2019). A Systematic Evaluation and Comparison Between Exploratory Structural Equation Modeling and Bayesian Structural Equation Modeling. *Structural equation modeling: a multidisciplinary journal, 26*(4), 529-556. doi:10.1080/10705511.2018.1554999

Gutiérrez-Zotes, J. A., Gutiérrez, F., Valero, J., Gallego, E., Baillés, E., Torres, X., . . . Livesley, W. J. (2008). Structure of personality pathology in normal and clinical samples: Spanish validation of the DAPP-BQ. *J Pers Disord, 22*(4), 389-404. doi:10.1521/pedi.2008.22.4.389

Hampshire, A., Hellyer, P. J., Soreq, E., Mehta, M. A., Ioannidis, K., Trender, W., . . . Chamberlain, S. R. (2021). Associations between dimensions of behaviour, personality traits, and mental-health during the COVID-19 pandemic in the United Kingdom. *Nature Communications, 12*(1), 4111. doi:10.1038/s41467-021-24365-5

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. doi:10.1080/19312458.2020.1718629

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 1–21. doi:<https://doi.org/10.3758/s13428-017-0935-1>

Hollander, E. (1993). Obsessive-compulsive spectrum disorders: An overview. *Psychiatric Annals, 23*(7), 355 -358.

Hollander, E., & Benzaquen, D. S. (1997). The obsessive compulsive spectrum disorders. *International Review of Psychiatry, 9*(1), 99 - 110. doi:10.1080/09540269775628

Hollander, E., & Rosen, J. (2002). Obsessive-compulsive spectrum disorders: A review. In M. Maj, N. Sartorius, A. Okasha, & J. Zohar (Eds.), *Obsessive-compulsive disorder* (2nd ed., pp. 203 - 224). West Sussex: John Wiley & Sons Ltd.

Hook, R. W., Grant, J. E., Ioannidis, K., Tiego, J., Yücel, M., Wilkinson, P., & Chamberlain, S. R. (2021). Trans-diagnostic measurement of impulsivity and compulsivity: A review of self-report tools. *Neuroscience and Biobehavioral Reviews, 120*, 455-469. doi:10.1016/j.neubiorev.2020.10.007

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human–Computer Interaction, 32*(1), 51-62. doi:10.1080/10447318.2015.1087664

IBM Corp. (Released 2020). IBM SPSS Statistics for Windows. (Version 27.0). Armonk, NY: IBM Corp.

Ioannidis, K., Hook, R., Wickham, K., Grant, J. E., & Chamberlain, S. R. (2019). Impulsivity in Gambling Disorder and problem gambling: a meta-analysis. *Neuropsychopharmacology, 44*(8), 1354-1361. doi:10.1038/s41386-019-0393-9

Jin, Y. (2020). A note on the cutoff values of alternative fit indices to evaluate measurement invariance for ESEM models. *International Journal of Behavioral Development, 44*(2), 166-174. doi:10.1177/0165025419866911

John, O. P., Robins, R. W., & Pervin, L. A. (2010). *Handbook of personality: Theory and research*: Guilford Press.

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In O. P. John & L. A. Pervin (Eds.), *Handbook of personality: Theory and research.* (Vol. 2, pp. 102 - 138). New York: Guilford Press.

Jöreskog, K. G. (2006). *Lisrel*: Wiley Online Library.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York: The Guilford Press.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychol Med, 42*(9), 1879-1890. doi:10.1017/s0033291711002674

Lang, K. M., & Little, T. D. (2016). Principled missing data treatments. *Prevention Science, 19*(3), 284-294. doi:10.1007/s11121-016-0644-5

Liang, X., Yang, Y., & Cao, C. (2020). The Performance of ESEM and BSEM in Structural Equation Models with Ordinal Indicators. *Structural equation modeling: a multidisciplinary journal, 27*(6), 874-887. doi:10.1080/10705511.2020.1716770

Livesley, W. J., & Jackson, D. N. (2009). *Manual for the Dimensional Assessmentof Personality Pathology–Basic Questionnaire (DAPP-BQ)*. Port Huron, MI: Sigma Assessment Systems.

Luigjes, J., Lorenzetti, V., de Haan, S., Youssef, G. J., Murawski, C., Sjoerds, Z., . . . Yücel, M. (2019). Defining Compulsive Behavior. *Neuropsychology Review, 29*(1), 4-13. doi:10.1007/s11065-019-09404-9

MacKillop, J., Weafer, J., J, C. G., Oshri, A., Palmer, A., & de Wit, H. (2016). The latent structure of impulsivity: impulsive choice, impulsive action, and impulsive personality traits. *Psychopharmacology (Berl), 233*(18), 3361-3370. doi:10.1007/s00213-016-4372-0

Markon, K. E. (2019). Bifactor and hierarchical models: Specification, inference, and interpretation. *Annu Rev Clin Psychol, 15*, 51-69. doi:10.1146/annurev-clinpsy-050718-095522

Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport and exercise psychology. In *Handbook of sport psychology, 3rd ed.* (pp. 774-798). Hoboken, NJ, US: John Wiley & Sons, Inc.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural equation modeling, 11*(3), 320-341. doi:<https://doi.org/10.1207/s15328007sem1103_2>

Marsh, H. W., Ludtke, O., Muthen, B., Asparouhov, T., Morin, A. J., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychol Assess, 22*(3), 471-491. doi:10.1037/a0019227

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*(1), 85-110. doi:10.1146/annurev-clinpsy-032813-153700

Marsh, H. W., Muthen, B., Asparouhov, T., Ludtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. *Structural Equation Modeling-a Multidisciplinary Journal, 16*(3), 439-476. doi:10.1080/10705510903008220

Meredith, W. (1993). Measurement Invariance, Factor-Analysis and Factorial Invariance. *Psychometrika, 58*(4), 525-543. doi:Doi 10.1007/Bf02294825

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *Am J Psychiatry, 158*(11), 1783 - 1793. doi:<https://doi.org/10.1176/appi.ajp.158.11.1783>

Morris, L. S., & Voon, V. (2016). Dimensionality of cognitions in behavioral addiction. *Curr Behav Neurosci Rep, 3*(1), 49-57. doi:<https://doi.org/10.1007/s40473-016-0068-3>

Muthén, L. K., & Muthén, B. O. (1998 - 2017). *Mplus User’s Guide.* (Eighth ed.). Los Angeles, CA: Muthén & Muthén.

Oltmanns, J. R., & Widiger, T. A. (2018). A self-report measure for the ICD-11 dimensional trait model proposal: The personality inventory for ICD-11. *Psychol Assess, 30*(2), 154-169. doi:10.1037/pas0000459

Pan, Q., & Wei, R. (2016). Fraction of Missing Information (γ) at Different Missing Data Fractions in the 2012 NAMCS Physician Workflow Mail Survey. *Applied mathematics, 7*(10), 1057-1067. doi:10.4236/am.2016.710093

Parkes, L., Tiego, J., Aquino, K., Braganza, L., Chamberlain, S. R., Fontenelle, L. F., . . . Yucel, M. (2019). Transdiagnostic variations in impulsivity and compulsivity in obsessive-compulsive disorder and gambling disorder correlate with effective connectivity in cortical-striatal-thalamic-cortical circuits. *Neuroimage, 202*, 116070. doi:10.1016/j.neuroimage.2019.116070

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. In S. T. Fiske, D. L. Schacter, & S. E. Taylor (Eds.), *Annual Review of Psychology, Vol 63* (Vol. 63, pp. 539-569).

Preacher, K. J., & Coffman, D. L. (May 2006). Computing power and minimum sample size for RMSEA [Computer software]. . Available from <http://quantpsy.org/>.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129-140. doi:10.1080/00223891.2012.725437

Robbins, T. W., Gillan, C. M., Smith, D. G., de Wit, S., & Ersche, K. D. (2012a). Neurocognitive endophenotypes of impulsivity and compulsivity: Towards dimensional psychiatry. *Trends in Cognitive Sciences, 16*, 81 - 91. doi:10.1016/j.tics.2011.11.009

Robbins, T. W., Gillan, C. M., Smith, D. G., de Wit, S., & Ersche, K. D. (2012b). Neurocognitive endophenotypes of impulsivity and compulsivity: Towards dimensional psychiatry. *Trends in Cognitive Sciences, 16*(1), 81 - 91. doi:<https://doi.org/10.1016/j.tics.2011.11.009>

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *J Pers Assess, 98*, 223-237. doi:<https://doi.org/10.1080/00223891.2015.1089249>

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological methods, 21*(2), 137-150. doi:<http://dx.doi.org/10.1037/met0000045>

Romero-Garcia, R., Hook, R. W., Tiego, J., Bethlehem, R. A. I., Goodyer, I. M., Jones, P. B., . . . Chamberlain, S. R. (2021). Brain micro-architecture and disinhibition: a latent phenotyping study across 33 impulsive and compulsive behaviours. *Neuropsychopharmacology, 46*(2), 423-431. doi:10.1038/s41386-020-00848-9

Samuel, D. B., Riddell, A. D., Lynam, D. R., Miller, J. D., & Widiger, T. A. (2012). A five-factor measure of obsessive-compulsive personality traits. *J Pers Assess, 94*(5), 456-465. doi:10.1080/00223891.2012.677885

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural equation modeling: a multidisciplinary journal, 21*(2), 167-180. doi:10.1080/10705511.2014.882658

Savalei, V., & Rhemtulla, M. (2012). On Obtaining Estimates of the Fraction of Missing Information From Full Information Maximum Likelihood. *Structural equation modeling: a multidisciplinary journal, 19*(3), 477-494. doi:10.1080/10705511.2012.687669

Simms, L. J., Goldberg, L. R., Roberts, J. E., Watson, D., Welte, J., & Rotterman, J. H. (2011). Computerized adaptive assessment of personality disorder: Introducing the CAT–PD project. *Journal of Personality Assessment, 93*(4), 380-389. doi:<https://doi.org/10.1080/00223891.2011.577475>

Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the Number of Response Options Matter? Psychometric Perspectives Using Personality Questionnaire Data. *Psychological Assessment, 31*(4), 557-566. doi:10.1037/pas0000648

Spear, L. P. (2000). The adolescent brain and age-related behavioral manifestations. *Neuroscience & biobehavioral reviews, 24*(4), 417-463. doi:[https://doi.org/10.1016/S0149-7634(00)00014-2](https://doi.org/10.1016/S0149-7634%2800%2900014-2)

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology, 91*(6), 1292 - 1306. doi:10.1037/0021-9010.91.6.1292

Tay, L., Meade, A. W., & Cao, M. (2015). An Overview and Practical Guide to IRT Measurement Equivalence Analysis. *Organizational Research Methods, 18*(1), 3-46. doi:10.1177/1094428114553062

Tiego, J., Chamberlain, S. R., Harrison, B. J., Dawson, A., Albertella, L., Youssef, G. J., . . . Yucel, M. (2020). Heritability of overlapping impulsivity and compulsivity dimensional phenotypes. *Sci Rep, 10*(1), 14378. doi:10.1038/s41598-020-71013-x

Tiego, J., Oostermeijer, S., Prochazkova, L., Parkes, L., Dawson, A., Youssef, G., . . . Yucel, M. (2018). Overlapping dimensional phenotypes of impulsivity and compulsivity explain co-occurrence of addictive and related behaviors. *CNS Spectr, 24*(4), 426-440. doi:10.1017/S1092852918001244

van den Heuvel, O. A., van Wingen, G., Soriano-Mas, C., Alonso, P., Chamberlain, S. R., Nakamae, T., . . . Veltman, D. J. (2016). Brain circuitry of compulsivity. *Eur Neuropsychopharmacol, 26*(5), 810-827. doi:10.1016/j.euroneuro.2015.12.005

van Kampen, D., de Beurs, E., & Andrea, H. (2008). A short form of the Dimensional Assessment of Personality Pathology-Basic Questionnaire (DAPP-BQ): the DAPP-SF. *Psychiatry Res, 160*(1), 115-128. doi:10.1016/j.psychres.2007.05.002

Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods, 5*(2), 139-158. doi:10.1177/1094428102005002001

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70. doi:<https://doi.org/10.1177/109442810031002>

Wright, A. G. C., & Simms, L. J. (2014). On the structure of personality disorder traits: conjoint analyses of the CAT-PD, PID-5, and NEO-PI-3 trait models. *Personal Disord, 5*(1), 43-54. doi:10.1037/per0000037

Yücel, M., & Fontenelle, L. F. (2012). Compulsivity as an endophenotype: The search for a hazy moving target. *Addiction, 107*(10), 1735-1736.

Table 1

*Standardized Factor Loading Estimates with Standard Errors and Fraction of Missing Information for Each Parameter as Estimated for the Two-Factor Model of Compulsivity Measured Using the Cambridge-Chicago Compulsivity Trait Scale in the Calibration Subsample*

|  |  |  |
| --- | --- | --- |
| CHI-T Items | Standardized Factor Loadings λ (*SE*) | Fraction of missing informationγ |
| Perfectionism  | Reward Drive |
| Perfectionism | Reward Drive  |
| 1 | Need for completion  | **.585 (.005)** | -.208 (.005) | .374 | .025 |
| 2 | Doing things just right | **.721 (.004)** | -.095 (.005) | .386 | .012 |
| 3 | Repetition to meet high standard | **.734 (.003)** | -.067 (.005) | .442 | .008 |
| 4 | Getting stuck in thoughts | .361 (.004) | .350 (.005) | .072 | .148 |
| 5 | Habit propensity | **.348 (.005)** | .185 (.005) | .082 | .032 |
| 6 | Addictive propensity | .129 (.005) | **.535 (.004)** | .021 | .215 |
| 7 | Rigidity | .234 (.004) | .324 (.004) | .057 | .126 |
| 8 | Tendency to act on urges | -.002 (.001)\*\* | **.581 (.004)** | .001 | .124 |
| 9 | Doing things that are immediately rewarding | -.080 (.004) | **.676 (.004)** | .016 | .055 |
| 10 | Difficulty moving from task to task | **.605 (.003)** | .194 (.005) | .128 | .085 |
| 11 | High standards | **.632 (.003)** | .000(.001)\* | .332 | .003 |
| 12 | Scope for improvement / nothing is good enough | **.231 (.005)** | .068 (.005) | .265 | .284 |
| 13 | Completion leads to soothing  | **.674 (.003)** | .051 (.004) | .247 | .005 |
| 14 | Need for control | .376 (.004) | .201 (.005) | .095 | .043 |
| 15 | Needing to be the best | **.444 (.004)** | .227 (.004) | .119 | .096 |

*Note.* Bold typeface = primary loading; Standard typeface = secondary loading. Standard errors appear in brackets. All loadings were *p* <.001 unless otherwise specified. \*\* *p* <.01. \* *p* > .05. Factor intercorrelation φ = .272, *SE* = .014, [95%*CI* = .244, .299], *p* <.001, (γ = .742). Perfectionism ω = .80. Reward Drive ω = .63.

Table 2

*Summary of Fit Statistics for Invariance Testing of the Two-Factor Exploratory Structural Equation Model in the Calibration and Validation Subsamples*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Invariance Model | Nested Models Compared | Δ*df* | ΔRMSEA*M* | ΔCFI*M* | ΔSRMR*M* | ΔBIC*M* | ΔAIC*M* |
| Subsamples |  |  |  |  |  |  |  |
| Weak | Model 2 – Model 1 | 26 | -.004 | .000 | -.002 | -286.143 | -23.216 |
| Strong | Model 5 – Model 2 | 15 | -.002 | -.001 | .002 | -167.845 | -16.157 |
| Strict | Model 7 – Model 5 | 15 | -.002 | -.001 | .000 | -169.486 | -17.797 |
| Latent Means | Model 10 – Model 5 | 2 | .000 | .000 | .000 | 23.763 | 3.538 |
|  Variance-Covariance Matrix | Model 4 – Model 2 | 3 | -.001 | -.001 | .005 | 546.194 | 576.532 |
|  |  |  |  |  |  |  |  |
| Sex |  |  |  |  |  |  |  |
| Weak | Model 2 – Model 1 | 52 | -.005 | -.003\* | .001 | 476.288 | 1,002.134 |
| Intercepts only | Model 51 – Model 2 | 30 | -.001 | -.009\* | .003 | 4,100.954 | 4,404.326 |
|  |  |  |  |  |  |  |  |
| Language |  |  |  |  |  |  |  |
| Weak | Model 2 – Model 1 | 26 | -.004 | -.001 | .001 | 141.214 | 404.14 |
| Strong | Model 5 – Model 2 | 15 | -.001 | -.005 | .001 | 2,217.238 | 2,368.927 |
| Strict | Model 7 – Model 5 | 15 | -.002 | -.003 | .001 | 982.214 | 1,133.902 |
| Latent Means | Model 10 – Model 5 | 2 | .000 | .000 | .004 | -371.52 | -391.745 |
|  Variance-Covariance Matrix | Model 4 – Model 2 | 3 | -.001 | -.001 | .000 | 96.61 | 126.948 |
|  |  |  |  |  |  |  |  |
| Ethnicity |  |  |  |  |  |  |  |
| Weak | Model 2 – Model 1 | 156 | -.007 | -.003 | .001 | -1,138.516 | 439.044 |
| Intercepts only | Model 5a – Model 2 | 90 | -.003 | -.007 | .003 | 1,752.413 | 2,662.544 |
|  |  |  |  |  |  |  |  |
| Psychiatric Status |  |  |  |  |  |  |  |
| Weak | Model 2 – Model 1 | 26 | -.004 | .000 | .002 | -98.142 | 164.785 |
| Strong | Model 5 – Model 2 | 15 | .000 | -.009\* | .011\* | 4,849.571 | 5,001.259 |
|  |  |  |  |  |  |  |  |
| Neurological Status |  |  |  |  |  |  |  |
| Weak | Model 2 – Model 1 | 26 | -.004 | -.001 | .000 | -262.042 | 0.884 |
| Strong | Model 5 – Model 2 | 15 | -.002 | -.001 | .000 | 41.467 | 193.156 |
| Strict | Model 7 – Model 5 | 15 | -.002 | .000 | .000 | -122.035 | 29.654 |
| Latent Means | Model 10 – Model 5 | 2 | .000 | .000 | .000 | -61.591 | -81.817 |
|  Variance-Covariance Matrix | Model 4 – Model 2 | 3 | -.001 | .000 | .000 | -13.57 | 16.769 |

 *Note.* Δ*df* = difference in the degrees of freedom between the more- and less-restricted nested invariance models; ΔRMSEA*M* = difference in the mean imputed Root Mean Square Error of Approximation between the more- and less-restrictive statistically nested invariance models; ΔCFI*M* = difference in the mean imputed Comparative Fit Index between the more- and less-restrictive invariance models; ΔSRMR*M* = difference in the mean imputed Standardized Root Mean Squared Residual between the more- and less-restrictive invariance models; ΔBIC*M*= difference in the mean imputed Bayesian information criterion values between nested model; ΔAIC*M*= difference in the mean imputed Akaike information criterion values between nested model. Calibration subsample *n* = 91,073. Validation subsample *n* = 91,072. Male *n* = 86,401, Female *n* = 94,385, non-Binary *n* = 1,334. English as first spoken language subsample *n* = 169,167. Non-English as first spoken language *n* = 12,978. No psychiatric condition *n* = 165,021; Psychiatric condition *n* = 17,124. No neurological condition *n* = 178,010; Neurological condition *n* = 4,135. a Factor loadings were freely estimated due to weak non-invariance.

\*ΔRMSEA > .010; \*ΔCFI > -.002. \*ΔSRMR > .020.