Published in final edited form as:

Assessment. 2019 January ; 26(1): 45–55. doi:10.1177/1073191117725171.

Is the Assessment of Personality Comparable in Persons Who Have and Have Not Experienced Depressive, Anxiety, and Substance Use Disorders? An Examination of Measurement Invariance

Thomas M. Olino¹, Laura Benini¹, Grace Icenogle¹, Sylia Wilson², Daniel N. Klein³, John R. Seeley^{4,5}, and Peter M. Lewinsohn⁵

¹Temple University, Philadelphia, PA, USA

²University of Minnesota, Minneapolis, MN, USA

³Stony Brook University, Stony Brook, NY, USA

⁴University of Oregon, Eugene, OR, USA

⁵Oregon Research Institute, Eugene, OR, USA

Abstract

Numerous studies have focused on characterizing personality differences between individuals with and without psychopathology. For drawing valid conclusions for these comparisons, the personality instruments used must demonstrate psychometric equivalence. However, we are unaware of any studies that examine measurement invariance in personality across individuals with and without psychopathology. This study conducted tests of measurement invariance for positive emotionality, negative emotionality, and disinhibition across individuals with and without histories of depressive, anxiety, and substance use disorders. We found consistent evidence that positive emotionality, negative emotionality, and disinhibition were assessed equivalently across all comparisons with each demonstrating strict invariance. Overall, results suggest that comparisons of personality measures between diagnostic groups satisfy the assumption of measurement invariance and these scales represent the same psychological constructs. Thus, mean-level comparisons across these groups are valid tests.

Keywords

personality; measurement invariance; psychopathology

There is long-standing interest in the conceptual (Klein, Durbin, Shankman, & Santiago, 2002; Klein, Kotov, & Bufferd, 2011) and empirical (Kotov, Gamez, Schmidt, & Watson,

Reprints and permissions: sagepub.com/journalsPermissions.nav

Corresponding Author: Thomas M. Olino, Department of Psychology, Temple University, 1701 North 13th Street, Weiss Hall, Philadelphia, PA 19122, USA. thomas.olino@temple.edu.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

2010) relationships between personality and psychopathology. However, despite the large number of studies conducted on personality differences between individuals with and without specific forms of psychopathology, there are important psychometric tests that have not yet been conducted—results of which may have significant impacts on the interpretation of previous studies and tests of these conceptual models. Dimensions of personality, especially neuroticism, extraversion, and disinhibition, provide a framework for understanding the structure of psychopathology, as they are linked in theoretically meaningful ways to hierarchical models of psychiatric disorders (Markon, Krueger, & Watson, 2005; Watson, Clark, & Harkness, 1994). Moreover, they are closely related to the biobehavioral domains and constructs that serve as the foundation for the National Institute of Mental Health's Research Domain Criteria (Cuthbert, 2005; Sanislow et al., 2010).

There are multiple competing and complementary models of personality. Big Five models include the dimensions of neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness (Goldberg, 1993; McCrae et al., 2000). Other models have emphasized three broad dimensions, including positive emotionality, negative emotionality, and constraint/disinhibition (Clark & Watson, 1999; Tellegen, 1985). Across these models, there is consistency in the meaning of neuroticism/negative emotionality (N/NE), which refers to individual differences in reactivity to negative emotions, and extraversion/positive emotionality (E/PE), which reflects individual differences in the experience of positive emotions, reward seeking, and interest and pleasure in social interactions. Disinhibition includes elements of both Big Five low agreeableness and conscientiousness.

Many studies have examined whether individuals with and without multiple forms of psychiatric disorders differ on these personality dimensions. Kotov et al. (2010) summarized the results of this literature in an extensive meta-analysis and found that individuals with depressive, anxiety, or substance use disorders (SUDs) reported higher levels of N/NE; individuals with a variety of disorders, and especially chronic depression, reported lower levels of E/PE; and individuals with SUDs reported higher levels of disinhibition. Negligible differences between individuals with and without disorders were found for agreeableness and openness. The cross-sectional studies included in the meta-analysis support concurrent associations between neuroticism, extraversion, and disinhibition and psychiatric disorders. The growing consensus regarding the continuity between personality and psychopathology may, in part, be a measurement artifact (e.g., Bienvenu, Hettema, Neale, Prescott, & Kendler, 2007). The magnitude of association may be distorted due to measurement differences between clinical and control samples. Thus, it is incumbent to demonstrate measurement invariance before concluding that there is a high degree of overlap, or continuity, between traits and psychopathology. There are important, available psychometric tests that can evaluate these possibilities.

When comparing groups using scores on a particular measure, it is critical that the psychometric quantities reflected in the measure's items are equivalent across groups. Measurement invariance (Vandenberg & Lance, 2000; Widaman, Ferrer, & Conger, 2010) is one approach to determine whether the measurement functioning of a target construct is comparable across different groupings (e.g., groups, informants, or occasions). In the absence of measurement invariance, it is not possible to determine whether observed

differences reflect true differences in the construct and/or differences in measurement properties (Millsap, 2011; Widaman et al., 2010). More concretely, in the absence of measurement invariance across individuals with and without psychopathology, it is not possible to interpret mean-level differences on measured personality traits. Thus, establishing measurement invariance for key personality measures across individuals with and without different psychiatric disorders is of fundamental importance for the literature on the relationship between personality and psychopathology.

There are multiple levels of measurement invariance, characterized by increasing strictness, that address different psychometric questions (Millsap, 2011; Widaman et al., 2010). The most basic requirement is that the same items are associated with the same construct across both groups (i.e., Do the same items load on the same factors when assessed within the different groups?). This is referred to as configural invariance. If the items assessing what are purportedly the same constructs differ across groups, the items hold different meanings within each group. Next, it is important that the magnitude of the associations between the items and the underlying construct are the same across both groups (i.e., Are the factor loadings for each factor comparable when assessed within the different groups?). This is referred to as metric invariance. Finally, the probability of endorsing a specific item in the same manner should be the same across both groups (Reise, Widaman, & Pugh, 1993; Vandenberg & Lance, 2000). This is referred to as scalar invariance. When configural, metric, and scalar invariance have been established for a particular measure across groups, scale scores can be considered to reflect the same psychometric quantities among the groups, and observed differences across groups can be attributed to the characteristic of interest, as opposed to psychometric factors. Thus, the interpretation of the results of empirical studies comparing personality traits across individuals with and without different psychiatric disorders hinges on first establishing measurement invariance for the personality instruments used. If metric and scalar invariance are not evinced across groups, there are differences in the underlying constructs between groups. Thus, differences between groups are not interpretable. However, as yet it remains untested whether the psychometric properties of commonly used personality instruments differ across individuals with and without a range of forms of psychopathology.

There are several reasons to think that the psychometric properties of items may be differentially associated with their underlying latent construct (i.e., fail to show measurement invariance) among groups with and without psychopathology. For example, N/NE items may have stronger factor loadings among individuals with depression relative to those without depression due to a negativity bias in reports of these items, such that individuals with depression are more likely to endorse negatively valenced items than individuals without depression (Sato & Kawahara, 2011). Alternatively, due to restricted range in levels of N/NE among individuals with depression, factor loadings may be attenuated. Similarly, it may be easier for individuals with some forms of psychopathology to endorse certain items relative to individuals without psychopathology. Statistically, this would be manifested when negatively valenced items have lower thresholds among individuals with psychopathology than without psychopathology. For example, individuals with a history of depression may endorse items assessing N/NE more easily than individuals without a history of depression. Conversely, positively valenced items may have higher

thresholds for individuals with a history of depression. Thus, individuals with a history of depression may endorse items assessing E/PE with greater difficulty than individuals without a history of depression. Thus, psychometric functioning of items may differ across important individual difference factors, such as history of psychopathology—evidenced by a lack of measurement invariance across individuals with and without psychiatric disorders. Failing to demonstrate measurement invariance for key personality measures across individuals with and without psychiatric disorders would raise significant questions about the validity and interpretation of the well-known associations between personality and psychopathology in the literature.

Measurement invariance of personality measures has been studied across genders (Laverdiere, Morin, & St-Hilaire, 2013; Marsh et al., 2010; Marsh, Nagengast, & Morin, 2013; Merz et al., 2013; Mor et al., 2008; Reise, Smith, & Furr, 2001; Rowinski, Cieciuch, & Oakland, 2014; Smith & Reise, 1998), developmental periods (Laverdiere et al., 2013; Marsh et al., 2010; Merz et al., 2013; Rowinski et al., 2014; Spence, Owens, & Goodyer, 2013), cultures (Church et al., 2011; Johnson, Spinath, Krueger, Angleitner, & Riemann, 2008; Woo et al., 2014), and self- and informant-reports (Olino & Klein, 2015). In each of these domains of study, authors have reported that the underlying psychometric functioning of these measures does not differ across these characteristics, supporting examination of mean-level personality differences across genders, time and development, cultures, and informants. Examination of mean-level personality differences across individuals with and without psychiatric disorders requires the same level of scrutiny to determine whether personality measures demonstrate measurement invariance with respect to the experience of psychopathology.

Previous studies have provided some evidence bearing on these issues. Some of this work has focused on the meta-structure of psychopathology, which has generally found that the structure of personality traits is similar among individuals with and without psychopathology (O'Connor, 2002). Other studies have focused on whether specific dimensions of personality are consistent across individuals with and without psychopathology. For example, Watson et al. (1995) conducted exploratory factor analyses on dimensions of positive affect/anhedonia, general distress, and somatic anxiety across samples of students, adults, and patients from a Veterans Affairs Medical Center. They found that the patterns of factor loadings across the samples were highly consistent, with very strong congruence coefficients (mean coefficient = .95). Similarly, Bagby et al. (1999) examined the structure of the revised NEO Personality Inventory (Costa & McCrae, 1992) between individuals with psychopathology, inclusive of schizophrenia, bipolar disorder, and major depression, and normative data on the revised NEO Personality Inventory. The authors found a high degree of consistency as reflected by strong congruence coefficients (with a range of .95 to .97). Thus, these studies provide preliminary evidence for configural invariance. Yet these studies have focused on higher order dimensions by focusing on analyses of facets of personality.

Only few studies have examined item-level functioning of personality measures across samples of individuals with and without psychopathology. Some investigations have found invariance in item functioning between individuals with and without psychopathology on

constructs relevant to personality, such as alexithymia and maladaptive cognitive styles (Meganck, Vanheule, & Desmet, 2008; Rijkeboer & van den Bergh, 2006). More recently, Eigenhuis, Kamphuis, and Noordhof (2016) examined multiple levels of measurement invariance using the Multidimensional Personality Questionnaire (Tellegen & Waller, 2008) between individuals without psychopathology and patients from a specialized clinic for individuals with personality disorders. Using sophisticated tests of measurement invariance, the authors found no evidence for measurement differences between individuals with and without psychopathology. However, as this study compared individuals in treatment for a multiple forms of psychopathology with individuals in the community, with presumably low levels of psychopathology, there is still little information on whether specific disorders differentially influence reporting of personality. Directions of biases in item properties may work in opposite directions for specific disorders (e.g., internalizing vs. externalizing conditions) and may mask how particular disorders influence item properties.

The present study provides the first examination of measurement invariance of personality assessment in individuals with and without specific common forms of psychopathology. We examined measurement invariance of a personality measure, the General Temperament Survey (GTS), which assesses N/NE, E/PE, and disinhibition/constraint, in a community-based sample with representative rates of psychopathology. We selected depressive, anxiety, and SUDs because these are relatively common psychiatric disorders in the general population (Grant et al., 2004; Kessler, Chiu, Demler, & Walters, 2005), and because each has been found to show mean-level differences on primary dimensions of personality (Kotov et al., 2010). We followed established guidelines for evaluating measurement invariance (Millsap, 2011), applying increasingly restrictive models to examine whether personality traits demonstrated configural, metric, and scalar invariance. Our results have direct implications for personality–psychopathology research to identify whether observed differences in personality between individuals with and without common forms of psychopathology are biased due to differences in measurement.

Method

Participants

The present study uses data from the Oregon Adolescent Depression Project (Lewinsohn, Hops, Roberts, Seeley, & Andrews, 1993), a longitudinal study of a large cohort of high school students who completed diagnostic assessments twice during adolescence, and a third time at an average age of 24 years. Participants completed a personality assessment several years later, at an average age of 28 years. Participants were randomly selected from nine high schools in Western Oregon. A total of 1,709 adolescents (ages 14–18; mean age = 16.6, $SD = 1.2$) completed the initial (T_1) assessments between 1987 and 1989. The participation rate at T_1 was 61%. All youth provided informed consent before completing research procedures. Approximately 1 year later, 1,507 of the adolescents (88%) returned for a second evaluation (T_2). Differences between the sample and the larger population from which it was selected, and between participants and those who declined to participate or dropped out of the study before T_2 , were small (Lewinsohn et al., 1993). However,

individuals with a history of disruptive behavior disorder at T₁ were more likely to drop out of the study, 16.8% versus 6.0%, $\chi^2(1, N = 1,709) = 31.22, p < .001$.

All participants with a history of a depressive disorder by T₂ ($n = 360$) or history of any other disorder(s) ($n = 284$), and a random sample of participants with no history of psychopathology by T₂ ($n = 457$), were invited to participate in a third evaluation (T₃) at approximately age 24. All non-White T₂ participants were also retained in the T₃ sample to maximize ethnic diversity. Of the 1,101 T₂ participants selected for a T₃ interview, 941 (85%) completed the age 24 evaluation. The T₂ diagnostic groups did not differ on the rate of participation at T₃.

When participants were 28.6 years ($SD = 9.7$), they were invited to complete a questionnaire battery that included a measure of personality. As diagnostic and personality assessments were not administered concurrently, we focus here on the participants who completed both the age 24 diagnostic assessment and the age 28 personality assessment. This yielded a sample of 734 participants. Of this specific subsample, the mean age was 27.7 years at the time of the questionnaire assessment ($SD = 1.4$, range = 24.4 to 31.7) and 60.9% ($n = 447$) were female.

Measures

Diagnostic Assessment—At T₁ and T₂, participants were interviewed with a version of the Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS; Orvaschel, Puig-Antich, Chambers, Tabrizi, & Johnson, 1982), which combined features of the Epidemiologic and Present Episode versions, and included additional items to derive *Diagnostic and Statistical Manual of Mental Disorders—Third edition—Revised (DSM-III-R*; American Psychiatric Association, 1987) diagnoses. Follow-up assessments at T₂ and T₃ were jointly administered with the Longitudinal Interval Follow-Up Evaluation (LIFE; Keller et al., 1987). The K-SADS/LIFE procedure provided information regarding the onset and course of disorders since the previous interview. Diagnoses were based on *DSM-III-R* criteria (American Psychiatric Association, 1987) for T₁ and T₂ and *Diagnostic and Statistical Manual of Mental Disorders—Fourth edition (DSM-IV*; American Psychiatric Association, 1994) criteria for T₃. Lifetime diagnoses were used as the indicators of psychopathology in this article, such that an individual was considered to have a diagnosis if it was identified as a past or current disorder at any of the assessments.

A subset of interviews from each wave was rated from audio or videotapes by a second interviewer for reliability purposes: T₁, $n = 263$; T₂, $n = 162$; and T₃, $n = 190$ interviews. Diagnostic agreement among raters was indexed by kappa. To avoid potential inflation, deflation, and/or unreliability of the kappa statistic, reliability was calculated only for categories diagnosed 10 or more times by both raters combined. Fleiss (1981) provides guidelines for the interpretation of kappa, whereby values $\geq .75$ denote excellent agreement beyond chance, those between .75 and .40 are indicative of good to fair agreement, and coefficients $< .40$ reflect poor agreement. Across the four assessment waves, interrater diagnostic reliability was good to excellent for disorders that occurred with sufficient frequency. The reliability of one disorder, major depressive disorder, could be determined for each of the four waves ($M\kappa = .84$; range: .81 to .86). Alcohol abuse/dependence and

cannabis abuse/dependence were diagnosed with sufficient frequency among raters during three of the four waves, and the mean kappas were .77 (range: .74 to .82) and .79 (range: .72 to .83), respectively. Hard drug abuse/dependence was diagnosed with sufficient frequency among raters during two of the four waves, with the mean kappa for this disorder being .76 (range: .69 to .83). Kappa coefficients for dysthymia (.56), posttraumatic stress disorder (.73), specific phobias (.66), panic disorder (.81), separation anxiety disorder (.83), attention-deficit/hyperactivity disorder (.89), and oppositional defiant disorder (.77) could only be determined for one of the four assessment waves. Generalized anxiety disorder, social phobia, obsessive-compulsive disorder, and conduct disorder were not diagnosed with sufficient frequency during any assessment wave to allow an evaluation of diagnostic reliability.

Personality Assessment—As part of a substudy, the Oregon Adolescent Depression Project participants were invited to complete a battery of self-report measures that included the GTS (Clark & Watson, 1990) at age 28. The GTS is a 90-item measure that assesses PE, NE, and disinhibition/constraint using True–False endorsement of items. Internal consistency estimates for these scales were excellent (α s = .88, .92, and .81 for PE, NE, and disinhibition, respectively). Each of the scales demonstrates good concurrent and discriminant validity in their associations with other measures of personality (Clark & Watson, 1999; Watson & Clark, 1992). For example, Watson and Clark (1992) showed strong correspondence between the GTS NE scale with the NEO-Five-Factor Inventory neuroticism (Costa & McCrae, 1992) and Goldberg Neuroticism scales (McCrae & Costa, 1985), and between the GTS PE scale with the NEO-Five-Factor Inventory Extraversion and Goldberg Extraversion scales in exploratory factor analytic models. This factor analytic model also showed substantial overlap between the GTS Disinhibition scale with measures of conscientiousness and moderate associations with agreeableness. The three GTS scales are included in the Schedule for Nonadaptive and Adaptive Personality (Clark, 1993), which has been used extensively in the literature.

Data Analytic Strategy

All models were estimated using *Mplus* 7.4 (Muthén & Muthén, 1998–2012) and applied sample weighting as participants were nonrandomly selected for the T₃ assessment. As all observed variables were dichotomous, we relied on estimation methods appropriate for this format. Initial models to demonstrate adequacy of single factor models were estimated using a weighted least squares estimator with a diagonal weight matrix and robust standard errors and a mean- and variance-adjusted chi-square test statistic (Flora & Curran, 2004) in theta parameterization. This was done to identify model fit in the overall sample and in each sample of individuals with and without specific forms of psychopathology. Model fit was assessed using the comparative fit index (CFI; Bentler, 1990) and the root mean square error of approximation (RMSEA; Steiger, 1990). Although existing guidelines are somewhat arbitrary (Marsh, Hau, & Wen, 2004), current conventions suggest that excellent fit is indicated by a CFI greater than .95 (Hu & Bentler, 1999) and a RMSEA below or equal to .05 (MacCallum, Browne, & Sugawara, 1996) and good fit is indicated by a CFI greater than .90 and a RMSEA between .05 and .10. We first fit unidimensional models to PE, NE,

and disinhibition factors for the full sample. We then estimated these factor models for portions of the sample with and without specific forms of psychopathology.

We conducted tests of measurement invariance using the sequence of models described in Ezpeleta and Penelo (2015). There are some modifications to the model constraints to permit evaluation of invariance of factor loadings and thresholds independently when binary outcomes are compared across groups. Specifically, residual errors of all indicators are constrained to be equal to one for all items across both groups in the configural, metric, and scalar invariance models. This is done for (a) model identification purposes and (b) so that factor loadings and thresholds are estimated independently. For the configural (common form) invariance model, presence of factor loadings are identical across groups, but are freely estimated for each group. For the metric (weak) invariance model, factor loadings are constrained to be equal across groups. For the scalar (strong) invariance model, item thresholds, in addition to factor loadings, are constrained to be equal across groups. Testing strict invariance typically is estimated by adding constraints on residuals across groups. However, as previously noted, with binary observed variables, previous tests imposed equality constraints on the residuals. Thus, we evaluated strict invariance by removing those constraints and examining whether relaxing those constraints resulted in improved model fit. As factor loadings and thresholds are constrained to be equal across groups, this model is identified and estimable. Model fit comparisons are evaluated by investigating change in both CFI and RMSEA. We relied on recommendations from Chen (2007) in evaluating the presence of measurement invariance. Chen (2007) recommended interpreting reductions in CFI and RMSEA of .010 as indicating noninvariance (i.e., failure to demonstrate measurement invariance). We estimated configural (same presence of factor loadings), metric (equality of factor loadings across groups), scalar (equality of thresholds across groups), and strict invariance (equality of residual variances across groups) for comparisons of those with and without specific forms of psychopathology (i.e., depression, anxiety, and SUD). When the RMSEA and CFI changes led to different conclusions, we relied on the RMSEA as the primary index of model change. We present chi-square difference tests (as implemented by the DIFFTEST option in *Mplus*) for completeness, but do not interpret the results as the chi-square difference test is overly sensitive to sample size (Asparouhov & Muthén, 2006).

Results

Of the 734 participants, 220 did not have a history of depression (neither major depressive disorder nor dysthymia), anxiety disorders, or SUDs. Fewer numbers of cases had only one specific form of psychopathology: 28 had a history of anxiety disorders, only; 86 had a history of SUD, only; 155 had a history of depression, only. Comorbidity was present in many cases: 57 had a history of depression and anxiety; 121 had a history of depression and SUD; 13 had a history of anxiety and SUD; and 54 had a history of depression, anxiety, and SUD. As there were too few cases of disorders without comorbidity to conduct analyses of measurement invariance, our analyses focused on comparisons between individuals with and without a history of depression, anxiety, and SUD, regardless of comorbid conditions.

Overall and Group-Specific Model Fit

We fit one-factor models for each of the personality dimensions in the overall sample and in each of the specific groups of interest. Model fit information is presented in Table 1. For the overall sample, model fit for PE was good and NE was excellent, as indicated by both CFI and RMSEA. However, model fit for disinhibition was equivocal, with RMSEA indicating excellent fit and CFI indicating poor fit, $\chi^2(560) = 1065.57, p < .001$; CFI = .858, RMSEA = .035 [.032, .038].

Before proceeding with further analyses, we examined the factor loadings contributing to poor model fit. Four items had standardized factor loadings less than |.32|, indicating that the factor accounted for less than approximately 10% of variance in the indicator. After removing these items, fit was not adequate, $\chi^2(434) = 860.58, p < .001$; CFI = .877, RMSEA = .037 [.033, .040]. Next, we examined items for similar content and/or item structure that could indicate that the residuals of the items could be correlated. We identified two item pairs that shared content *and* structure (i.e., were both reverse scored). Thus, we included two residual correlations and reestimated the model. This resulting model was an adequate fit to the data, $\chi^2(432) = 770.45, p < .001$; CFI = .903, RMSEA = .033 [.029, .036]. We computed intraclass correlations to examine the consistency of factor loadings in the models with and without the residual correlations. The intraclass correlation was .943, suggesting that the addition of the residual correlations was not distorting the factor loadings, the key structural elements, in the model. We proceeded with this as the baseline model for the disinhibition factor further analysis.

Fit for PE and NE was good or excellent for individuals with depression, anxiety, SUDs, and their complement groups without those disorders. For the disinhibition scale, RMSEA indicated excellent fit and CFI indicated adequate fit when focusing on individuals without psychopathology. However, disinhibition factor models for individuals with psychopathology had excellent fit as indexed by the RMSEA, but less than adequate fit as indexed by the CFI (range from .87 to .89). As these are only marginally lower than .90, we proceed with further analyses cautiously. As these individual group analyses serve as baseline models for conducting measurement invariance analyses, we next examined invariance for each of the three scales in each of the three disorder comparisons, including the models for disinhibition that demonstrated equivocal fit.

Configural, Weak, Strong, and Strict Models Across Psychiatric Groups

Model fit comparisons between individuals with and without depressive disorder diagnoses are presented in Table 2. For NE, the configural invariance model fit the data well. Imposing constraints on the factor loadings (metric invariance) and, in a second step, thresholds (scalar invariance model) did not substantively worsen model fit, as indicated by small changes in RMSEA/CFI. In addition, removing equality constraints on the residual variances did not substantively improve model fit, as indicated by small changes in RMSEA/CFI. Thus, for NE, strict invariance was supported. For PE, the configural invariance model was a good fit to the data. Imposing constraints on the factor loadings (i.e., metric invariance) did not lead to substantive changes in model fit. When additional constraints were imposed on the thresholds (i.e., scalar invariance), model fit was not substantively worsened. Removing

equality constraints on the residual variances did not substantively improve model fit. Thus, for PE, strict invariance was supported. For disinhibition, the configural invariance model was an adequate fit to the data. Imposing constraints on the factor loadings (i.e., metric invariance) led to a small change in the RMSEA (change = .002), but a larger change in the CFI (change = .021). In light of the inconsistent result, we rely on the RMSEA. When additional constraints were imposed on the thresholds (i.e., scalar invariance), model fit was not substantively worsened. Removing equality constraints on the residual variances did not substantively improve model fit. Thus, for disinhibition, strict invariance was supported.

Model fit comparisons between individuals with and without anxiety disorder diagnoses are presented in Table 3. For NE, the configural invariance model fit the data well. Imposing constraints on the factor loadings (i.e., metric invariance) and thresholds (i.e., scalar invariance) did not substantively worsen model fit. In addition, removing equality constraints on the residual variances did not substantively improve model fit. Thus, for NE, strict invariance was supported. For PE, imposing constraints on the factor loadings (i.e., metric invariance) and thresholds (i.e., scalar invariance) did not substantively worsen model fit. In addition, removing equality constraints on the residual variances did not substantively improve model fit. Thus, for PE, strict invariance was supported. For Disinhibition, the configural invariance model was an adequate fit to the data. Imposing constraints on the factor loadings (i.e., metric invariance) led to a small change in the RMSEA (change = .002), but a larger change in the CFI (change = .016). Given the inconsistent result, we relied on the RMSEA. When additional constraints were imposed on the thresholds, model fit was not substantively worsened. Removing equality constraints on the residual variances did not substantively improve model fit. Thus, for Disinhibition, strict invariance was supported.

Model fit comparisons between individuals with and without SUD diagnoses are presented in Table 4. For NE, the configural invariance model fit the data well. Imposing constraints on the factor loadings (i.e., metric invariance) and thresholds (i.e., scalar invariance) did not substantively worsen model fit. In addition, removing equality constraints on the residual variances did not substantively improve model fit. Thus, for NE, strict invariance was supported. For PE, the configural invariance model was a good fit to the data. Imposing constraints on the factor loadings (i.e., metric invariance) and thresholds (i.e., scalar invariance) did not substantively worsen model fit. In addition, removing equality constraints on the residual variances did not substantively improve model fit. Thus, for PE, strict invariance was supported. For Disinhibition, the configural invariance model was an adequate fit to the data. Imposing constraints on the factor loadings (i.e., metric invariance) and thresholds (i.e., scalar invariance) did not substantively worsen model fit. In addition, removing equality constraints on the residual variances did not substantively improve model fit. Thus, for Disinhibition, strict invariance was supported.

Sensitivity Analysis

To address whether comorbidity in our groups may have affected our results, we conducted a set of additional analyses comparing the largest single diagnosis group, individuals with depression only (i.e., no other disorders), with individuals with no lifetime history of

psychopathology. Consistent with the initial analyses, NE, PE, and disinhibition all indicated support for configural, metric, scalar, and strict invariance.

Discussion

Numerous previous studies have examined differences between individuals with and without psychopathology on multiple dimensions of personality. This literature has recently been summarized in a quantitative review by Kotov et al. (2010). However, there are critical unexamined psychometric questions that need to be addressed to interpret differences between those with and without psychopathology. Previous work has tested configural invariance in samples of individuals with and without broad forms of psychopathology (Bagby et al., 1999; Watson et al., 1995) and in comparisons between individuals with and without personality pathology (Eigenhuis et al., 2016). However, no studies have examined whether measurement invariance holds between those with and without the experience of specific forms of psychopathology. In the present study, we conducted tests of measurement invariance on PE, NE, and disinhibition, measured by the GTS, between individuals with and without a history of depressive, anxiety, and SUDs. We found support for measurement invariance for these personality dimensions across groups. These findings suggest that tests of mean-level differences between groups with and without psychopathology are valid and interpretable. For example, we can be assured that the higher levels of NE among individuals with depression found in previous studies is due to true differences in NE, not measurement effects. Thus, we extend previous work supporting the validity of personality assessment in individuals with and without specific forms of psychopathology.

The results of this work are most straightforward for PE and NE. For the overall sample and each subsample, the unidimensional models fit the data well. In our model, comparisons from free (configural) to strict invariance model for depression, anxiety, and SUDs, we found only trivial differences in model fit. These findings indicate that the PE and NE scales are functioning similarly across groups and that we can meaningfully compare scores between groups.

The results for disinhibition were somewhat weaker. Our initial model fit the data well according to only one of two primary fit indices; therefore, several modifications were made to the scale. Nonetheless, we used the unidimensional structure as the baseline model to test for measurement invariance. Our modeling found that, overall, there was support for strict measurement invariance (including configural, weak, strong, or strict) in the disinhibition scale across all three forms of psychopathology examined. Because model fit was less than adequate (as indexed by the CFI) for individuals with psychopathology, caution is needed in interpreting these results.

The present study provides reassurance that personality comparisons across individuals with and without histories of various forms of psychopathology are likely to be valid as trait measures and have a high degree of measurement invariance. These conclusions are supported through the use of sophisticated analyses in a large, unselected community-based sample.

However, there are several limitations to the work. First, our analyses compared individuals with a target disorder with individuals without that disorder. Given the substantial comorbidity in the sample, there were individuals with other (nonindex) forms of psychopathology in the comparison group. Thus, we cannot rule out the possibility that some of our results are due to the presence of psychopathology in the comparison groups. However, results of a sensitivity analysis comparing individuals with depression only (our largest single disorder group) to individuals without any form of psychopathology mirrored those found for the primary analyses. Thus, we have some assurance that our results are not fully accounted for by comorbidity. Second, our analyses focused on individuals with a history of, but not necessarily current, psychopathology. Thus, our data and analyses cannot precisely address whether measurement invariance holds when individuals are currently in episode. Furthermore, the personality assessment followed the diagnostic assessments, so it is possible that some individuals developed episodes (or had recurrent episodes) after the diagnostic assessment. Third, this work focuses on only one measurement occasion. More powerful tests can come from studies that assess longitudinal measurement invariance in samples assessed before, during, and after disorder onset. Fourth, the disinhibition scale required modifications to yield adequate fit in the entire sample. However, fit of the models among individuals with common forms of psychopathology was marginally less than adequate. Thus, our conclusions for that scale are cautious. Finally, we conducted these analyses with a well-validated, albeit single, measure of personality and focused on broadband dimensions. Future studies would benefit from testing similar questions with a broader array of personality measures and examining narrowband dimensions of personality and broader structural questions.

Despite these limitations, the present study provides important reassurance for the field that tests of mean-level differences between individuals with and without specific forms of psychopathology generally satisfy the assumption of measurement invariance. Tests of measurement invariance across diagnostic groupings are needed for other frequently used measures of personality, other diagnoses, and with individuals in current episodes of psychopathology. Furthermore, as there is a substantial body of research linking dimensions of personality to psychopathology (Kotov et al., 2010), there may be questions about how varying levels of symptomatology influence personality assessment and vice versa. Recent analytic developments permit testing of measurement invariance using continuous predictors, such that individual levels of psychopathology, rather than diagnosis, could be used to examine differences in personality item functioning (Bauer, 2016).

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was partially supported by National Institute of Mental Health Grants RO1 MH40501, RO1 MH50522, and RO1 MH52858 (Dr. Lewinsohn), RO1 MH107495 (Dr. Olino), and K01 DA037280 (Dr. Wilson).

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3. Washington, DC: Author; 1987. rev

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: Author; 1994. rev
- Asparouhov, T; Muthén, BO. Robust chi square difference testing with mean and variance adjusted test statistics. 2006. Retrieved from <https://www.statmodel.com/download/webnotes/webnote10.pdf>
- Bagby RM, Costa PT, McCrae RR, Livesley WJ, Kennedy SH, Levitan RD, ... Young LT. 1999; Replicating the five factor model of personality in a psychiatric sample. *Personality and Individual Differences*. 27:1135–1139.
- Bauer DJ. 2016; A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*. doi: 10.1037/met0000077
- Bentler PM. 1990; Comparative fit indexes in structural models. *Psychological Bulletin*. 107:238–246. [PubMed: 2320703]
- Bienvenu OJ, Hettema JM, Neale MC, Prescott CA, Kendler KS. 2007; Low extraversion and high neuroticism as indices of genetic and environmental risk for social phobia, agoraphobia, and animal phobia. *American Journal of Psychiatry*. 164:1714–1721. [PubMed: 17974937]
- Chen FF. 2007; Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*. 14:464–504.
- Church A, Alvarez JM, Mai NT, French BF, Katigbak MS, Ortiz FA. 2011; Are cross-cultural comparisons of personality profiles meaningful? Differential item and facet functioning in the Revised NEO Personality Inventory. *Journal of Personality and Social Psychology*. 101:1068–1089. [PubMed: 21910552]
- Clark, LA. Manual for the schedule for nonadaptive and adaptive personality. Minneapolis: University of Minnesota Press; 1993.
- Clark, LA, Watson, D. The General Temperament Survey (GTS). Iowa City: University of Iowa; 1990.
- Clark, LA, Watson, D. Temperament: A new paradigm for trait psychology. In: Pervin, LA, John, OP, editors *Handbook of personality: Theory and research*. 2. New York, NY: Guilford Press; 1999. 399–423.
- Costa, PT, McCrae, RR. Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources; 1992.
- Cuthbert BN. 2005; Dimensional models of psychopathology: Research agenda and clinical utility. *Journal of Abnormal Psychology*. 114:565–569. [PubMed: 16351380]
- Eigenhuis A, Kamphuis J, Noordhof A. 2016; Personality in general and clinical samples: Measurement invariance of the Multidimensional Personality Questionnaire. *Psychological Assessment*. doi: 10.1037/pas0000408
- Ezpeleta L, Penelo E. 2015; Measurement invariance of oppositional defiant disorder dimensions in 3-year-old preschoolers. *European Journal of Psychological Assessment*. 31:45–53.
- Fleiss, JL. Statistical methods for rates and proportions. Vol. 2. New York, NY: John Wiley; 1981. The measurement of interrater agreement; 212–236.
- Flora DB, Curran PJ. 2004; An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*. 9:466–491. [PubMed: 15598100]
- Goldberg LR. 1993; The structure of phenotypic personality traits. *American Psychologist*. 48:26–34. [PubMed: 8427480]
- Grant BF, Stinson FS, Dawson DA, Chou SP, Dufour MC, Compton W, ... Kaplan K. 2004; Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Archives of General Psychiatry*. 61:807–816. [PubMed: 15289279]
- Hu L, Bentler PM. 1999; Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 6:1–55.
- Johnson W, Spinath F, Krueger RF, Angleitner A, Riemann R. 2008; Personality in Germany and Minnesota: An IRT-based comparison of MPQ self-reports. *Journal of Personality*. 76:665–706. [PubMed: 18399949]
- Keller MB, Lavori PW, Friedman B, Nielsen E, Endicott J, McDonald-Scott P, Andreasen NC. 1987; The Longitudinal Interval Follow-up Evaluation: A comprehensive method for assessing outcome

in prospective longitudinal-studies. *Archives of General Psychiatry*. 44:540–548. [PubMed: 3579500]

- Kessler RC, Chiu WT, Demler O, Walters EE. 2005; Prevalence, severity, and comorbidity of 12-month *DSM-IV* disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*. 62:617–627. [PubMed: 15939839]
- Klein, DN, Durbin, CE, Shankman, SA, Santiago, NJ. Depression and personality. In: Gotlib, IH, Hammen, CL, editors *Handbook of depression*. New York, NY: Guilford Press; 2002. 115–140.
- Klein DN, Kotov R, Bufferd SJ. 2011; Personality and depression: Explanatory models and review of the evidence. *Annual Review of Clinical Psychology*. 7:269–295.
- Kotov R, Gamez W, Schmidt F, Watson D. 2010; Linking “big” personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*. 136:768–821. [PubMed: 20804236]
- Laverdiere O, Morin AJ, St-Hilaire F. 2013; Factor structure and measurement invariance of a short measure of the Big Five personality traits. *Personality and Individual Differences*. 55:739–743.
- Lewinsohn PM, Hops H, Roberts RE, Seeley JR, Andrews JA. 1993; Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-III-R disorders in high school students. *Journal of Abnormal Psychology*. 102:133–144. [PubMed: 8436689]
- MacCallum RC, Browne MW, Sugawara HM. 1996; Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*. 1:130–149.
- Markon KE, Krueger RF, Watson D. 2005; Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*. 88:139–157. [PubMed: 15631580]
- Marsh HW, Hau KT, 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*. 11:320–341.
- Marsh HW, Ludtke O, Muthén B, Asparouhov T, Morin AJ, Trautwein U, Nagengast B. 2010; A new look at the Big Five factor structure through exploratory structural equation modeling. *Psychological Assessment*. 22:471–491. [PubMed: 20822261]
- Marsh HW, Nagengast B, Morin AJ. 2013; Measurement invariance of Big-Five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology*. 49:1194–1218. [PubMed: 22250996]
- McCrae RR, Costa PT. 1985; Updating Norman’s “adequacy taxonomy”: Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology*. 49:710–721. [PubMed: 4045699]
- McCrae RR, Costa PT Jr, Ostendorf F, Angleitner A, Hrebáčková M, Avia MD, ... Smith PB. 2000; Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology*. 78:173–186. [PubMed: 10653513]
- Meganck R, Vanheule S, Desmet M. 2008; Factorial validity and measurement invariance of the 20-item Toronto Alexithymia Scale in clinical and nonclinical samples. *Assessment*. 15:36–47. [PubMed: 18258730]
- Merz EL, Malcarne VL, Roesch SC, Ko CM, Emerson M, Roma VG, ... Sadler GR. 2013; Psychometric properties of Positive and Negative Affect Schedule (PANAS) original and short forms in an African American community sample. *Journal of Affective Disorders*. 151:942–949. [PubMed: 24051099]
- Millsap, RE. *Statistical approaches to measurement invariance*. New York, NY: Taylor & Francis; 2011.
- Mor N, Zinbarg RE, Craske MG, Mineka S, Uliaszek A, Rose R, ... Waters AM. 2008; Evaluating the invariance of the factor structure of the EPQ-R-N among adolescents. *Journal of Personality Assessment*. 90:66–75. [PubMed: 18444097]
- Muthén, LK, Muthén, BO. *Mplus user’s guide*. 7. Los Angeles, CA: Muthén & Muthén; 1998–2012.
- O’Connor BP. 2002; The search for dimensional structure differences between normality and abnormality: A statistical review of published data on personality and psychopathology. *Journal of Personality and Social Psychology*. 83:962–982. [PubMed: 12374447]

- Olino TM, Klein DN. 2015; Psychometric comparison of self-and informant-reports of personality. *Assessment*. 22:655–664. [PubMed: 25612626]
- Orvaschel H, Puig-Antich J, Chambers WJ, Tabrizi MA, Johnson R. 1982; Retrospective assessment of prepubertal major depression with the Kiddie-SADS-E. *Journal of the American Academy of Child & Adolescent Psychiatry*. 21:392–397.
- Reise SP, Smith L, Furr RM. 2001; Invariance on the NEO PI-R neuroticism scale. *Multivariate Behavioral Research*. 36:83–110.
- Reise SP, Widaman KF, Pugh RH. 1993; Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*. 114:552–566. [PubMed: 8272470]
- Rijkeboer MM, van den Bergh H. 2006; Multiple group confirmatory factor analysis of the Young Schema-Questionnaire in a Dutch clinical versus non-clinical population. *Cognitive Therapy and Research*. 30:263–278.
- Rowinski T, Cieciuch J, Oakland T. 2014; The factorial structure of four temperament styles and measurement invariance across gender and age groups. *Journal of Psychoeducational Assessment*. 32:77–82.
- Sanislow CA, Pine DS, Quinn KJ, Kozak MJ, Garvey MA, Heinssen RK, ... Cuthbert BN. 2010; Developing constructs for psychopathology research: Research domain criteria. *Journal of Abnormal Psychology*. 119:631–639. [PubMed: 20939653]
- Sato H, Kawahara J-i. 2011; Selective bias in retrospective self-reports of negative mood states. *Anxiety, Stress, & Coping*. 24:359–367.
- Smith LL, Reise SP. 1998; Gender differences on negative affectivity: An IRT study of differential item functioning on the Multidimensional Personality Questionnaire Stress Reaction scale. *Journal of Personality and Social Psychology*. 75:1350–1362. [PubMed: 9866192]
- Spence R, Owens M, Goodyer I. 2013; The longitudinal psychometric properties of the EAS Temperament Survey in adolescence. *Journal of Personality Assessment*. 95:633–639. [PubMed: 23905630]
- Steiger JH. 1990; Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*. 25:173–180. [PubMed: 26794479]
- Tellegen, A. Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In: Hussain, TA, Maser, JD, editors *Anxiety and the anxiety disorders*. Hillsdale, NJ: Lawrence Erlbaum; 1985. 681–706.
- Tellegen, A, Waller, NG. Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. In: Boyle, GJ, Matthews, G, Saklofske, DH, editors *The SAGE handbook of personality theory and assessment*, Vol. 2: Personality measurement and testing. Thousand Oaks, CA: Sage; 2008. 261–292.
- Vandenberg RJ, Lance CE. 2000; A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*. 3:4–69.
- Watson D, Clark LA. 1992; On traits and temperament: General and specific factors of emotional experience and their relation to the five-factor model. *Journal of Personality*. 60:441–476. [PubMed: 1635050]
- Watson D, Clark LA, Harkness AR. 1994; Structures of personality and their relevance to psychopathology. *Journal of Abnormal Psychology*. 103:18–31. [PubMed: 8040477]
- Watson D, Clark LA, Weber K, Assenheimer JS, Strauss ME, McCormick RA. 1995; Testing a tripartite model: II. Exploring the symptom structure of anxiety and depression in student, adult, and patient samples. *Journal of Abnormal Psychology*. 104:15–25. [PubMed: 7897037]
- Widaman KF, Ferrer E, Conger RD. 2010; Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*. 4:10–18. [PubMed: 20369028]
- Woo SE, Chernyshenko OS, Longley A, Zhang ZX, Chiu CY, Stark SE. 2014; Openness to experience: Its lower level structure, measurement, and cross-cultural equivalence. *Journal of Personality Assessment*. 96:29–45. [PubMed: 23795950]

Table 1

Tests of Model Fit for Personality Dimensions Within the Complete and Psychiatric Disorder Samples.

	χ^2	<i>df</i>	RMSEA [90% CI]	CFI
Complete sample (<i>n</i> = 733)				
Positive emotionality	1082.06	324	.06 [.05, .06]	.91
Negative emotionality	838.33	350	.04 [.04, .05]	.96
Disinhibition	770.45	432	.03 [.03, .04]	.90
No anxiety (<i>n</i> = 581)				
Positive emotionality	873.48	324	.05 [.05, .06]	.91
Negative emotionality	683.54	350	.04 [.04, .04]	.96
Disinhibition	689.26	432	.03 [.03, .04]	.90
Anxiety (<i>n</i> = 152)				
Positive emotionality	441.61	324	.05 [.04, .06]	.95
Negative emotionality	460.37	350	.05 [.03, .06]	.97
Disinhibition	498.99	432	.03 [.01, .04]	.89
No depression (<i>n</i> = 346)				
Positive emotionality	549.61	324	.04 [.04, .05]	.92
Negative emotionality	483.07	350	.03 [.03, .04]	.96
Disinhibition	562.46	432	.03 [.02, .04]	.92
Depression (<i>n</i> = 387)				
Positive emotionality	771.16	324	.06 [.05, .06]	.91
Negative emotionality	613.92	350	.04 [.04, .05]	.96
Disinhibition	614.16	432	.03 [.03, .04]	.87
No SUD (<i>n</i> = 459)				
Positive emotionality	711.93	324	.05 [.05, .06]	.92
Negative emotionality	625.23	350	.04 [.04, .05]	.96
Disinhibition	595.35	432	.03 [.02, .03]	.91
SUD (<i>n</i> = 274)				
Positive emotionality	566.32	324	.05 [.04, .06]	.93
Negative emotionality	517.90	350	.04 [.03, .05]	.96
Disinhibition	533.99	432	.03 [.02, .04]	.88

Note. *df* = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SUD = substance use disorder. Results for the disinhibition factor did not include four items with factor loadings less than [.32] in the initial model with all items. There were also two item pairs that shared content *and* were both reverse scored; we included residual covariance parameters between these items in all models.

Table 2

Comparisons of Configural, Metric, Scalar, and Strict Invariance Across Individuals With and Without Depressive Disorders on Positive Emotionality, Negative Emotionality, and Disinhibition.

	χ^2	df	RMSEA [90% CI]	CFI	Model comparison	χ^2	df	p	RMSEA	CFI
<i>Negative emotionality</i>										
1. Configural	1092.28	700	.039 [.035, .044]	.967						
2. Weak (Metric)	998.76	727	.032 [.027, .037]	.977	2 vs. 1	16.39	27	.94	.007	.010
3. Strong (Metric and Scalar)	1021.86	754	.031 [.026, .036]	.977	3 vs. 2	22.27	27	.72	.001	.000
4. Strict (Strong and Uniquenesses)	1103.77	726	.038 [.033, .042]	.968	4 vs. 3	18.02	28	.93	.007	.009
<i>Positive emotionality</i>										
1. Configural	1316.92	648	.053 [.049, .057]	.916						
2. Weak (Metric)	1180.65	675	.045 [.041, .049]	.936	2 vs. 1	28.59	27	.38	.008	.020
3. Strong (Metric and Scalar)	1227.22	700	.045 [.041, .050]	.933	3 vs. 2	44.68	25	.01	.000	.003
4. Strict (Strong and Uniquenesses)	1297.64	673	.050 [.046, .054]	.921	4 vs. 3	36.32	27	.11	.005	.012
<i>Disinhibition</i>										
1. Configural	1178.98	864	.032 [.027, .036]	.895						
2. Weak (Metric)	1165.08	895	.029 [.024, .033]	.910	2 vs. 1	31.85	31	.42	.003	.015
3. Strong (Metric and Scalar)	1199.78	924	.029 [.024, .033]	.908	3 vs. 2	37.75	29	.13	.000	.002
4. Strict (Strong and Uniquenesses)	1185.48	893	.030 [.025, .034]	.903	4 vs. 3	36.86	31	.22	.001	.005

Note. *df* = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index. Test of weak invariance imposes equality constraints on factor loadings and residual variance terms across groups. Test of strong invariance imposes equality constraints on factor loadings, thresholds, and residual variance terms across groups. Test of strict invariance imposes equality constraints on factor loadings and thresholds, but frees constraints on residual variance terms across groups.

Table 3

Comparisons of Configural, Metric, Scalar, and Strict Invariance Across Individuals With and Without Anxiety Disorders on Positive Emotionality, Negative Emotionality, and Disinhibition.

	χ^2	df	RMSEA [90% CI]	CFI	Model comparison	χ^2	df	p	RMSEA	CFI
<i>Negative emotionality</i>										
1. Configural	1092.28	700	.039 [.035, .044]	.967						
2. Weak (Metric)	998.76	727	.032 [.027, .037]	.977	2 vs. 1	16.39	27	.94	.007	.010
3. Strong (Metric and Scalar)	1021.86	754	.031 [.026, .036]	.977	3 vs. 2	22.27	27	.72	.001	.000
4. Strict (Strong and Uniquenesses)	1103.77	726	.038 [.033, .042]	.968	4 vs. 3	18.02	28	.93	.007	.009
<i>Positive emotionality</i>										
1. Configural	1202.73	648	.048 [.044, .053]	.929						
2. Weak (Metric)	1117.19	675	.042 [.038, .047]	.944	2 vs. 1	31.55	27	.25	.006	.015
3. Strong (Metric and Scalar)	1158.98	700	.042 [.038, .047]	.942	3 vs. 2	54.91	25	.000	.000	.002
4. Strict (Strong and Uniquenesses)	1211.74	673	.047 [.042, .051]	.931	4 vs. 3	38.16	27	.08	.005	.011
<i>Disinhibition</i>										
1. Configural	1141.20	864	.030 [.025, .034]	.905						
2. Weak (Metric)	1139.77	895	.027 [.022, .032]	.916	2 vs. 1	39.77	31	.13	.003	.011
3. Strong (Metric and Scalar)	1177.82	924	.027 [.022, .032]	.913	3 vs. 2	52.52	29	.005	.000	.003
4. Strict (Strong and Uniquenesses)	1160.99	893	.029 [.024, .033]	.908	4 vs. 3	40.82	31	.11	.002	.005

Note. *df* = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index. Test of weak invariance imposes equality constraints on factor loadings and residual variance terms across groups. Test of strong invariance imposes equality constraints on factor loadings, thresholds, and residual variance terms across groups. Test of strict invariance imposes equality constraints on factor loadings and thresholds, but frees constraints on residual variance terms across groups.

Table 4

Comparisons of Configural, Metric, Scalar, and Strict Invariance Across Individuals With and Without Substance Use Disorders on Positive Emotionality, Negative Emotionality, and Disinhibition.

	χ^2	df	RMSEA [90% CI]	CFI	Model comparison	χ^2	df	p	RMSEA	CFI
<i>Negative emotionality</i>										
1. Configural	1126.573	700	.041 [.036, .045]	.966						
2. Weak (Metric)	1076.85	727	.036 [.032, .041]	.972	2 vs. 1	34.97	27	.14	.005	.006
3. Strong (Metric and Scalar)	1107.26	754	.036 [.031, .040]	.972	3 vs. 2	35.72	27	.12	.000	.000
4. Strict (Strong and Uniquenesses)	1138.74	726	.039 [.035, .044]	.967	4 vs. 3	39.89	28	.07	.003	.005
<i>Positive emotionality</i>										
1. Configural	1260.47	648	.051 [.047, .055]	.927						
2. Weak (Metric)	1185.96	675	.045 [.041, .050]	.939	2 vs. 1	46.17	27	.01	.006	.012
3. Strong (Metric and Scalar)	1230.59	700	.045 [.041, .050]	.937	3 vs. 2	52.85	25	.001	.001	.001
4. Strict (Strong and Uniquenesses)	1258.57	673	.049 [.045, .053]	.930	4 vs. 3	53.67	27	.002	.004	.007
<i>Disinhibition</i>										
1. Configural	1125.31	864	.029 [.024, .033]	.900						
2. Weak (Metric)	1145.45	895	.028 [.023, .032]	.904	2 vs. 1	45.10	31	.048	.001	.004
3. Strong (Metric and Scalar)	1195.27	924	.028 [.023, .033]	.896	3 vs. 2	71.39	29	.000	.000	.008
4. Strict (Strong and Uniquenesses)	1169.26	893	.029 [.024, .034]	.894	4 vs. 3	43.90	31	.06	.001	.002

Note. df = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index. Test of weak invariance imposes equality constraints on factor loadings and residual variance terms across groups. Test of strong invariance imposes equality constraints on factor loadings, thresholds, and residual variance terms across groups. Test of strict invariance imposes equality constraints on factor loadings and thresholds, but frees constraints on residual variance terms across groups.