Published in final edited form as:
Personal Disord. 2012 April ; 3(2): 107–126. doi:10.1037/a0025905.

Integrating Competing Dimensional Models of Personality: Linking the SNAP, TCI, and NEO Using Item Response Theory

Stephanie D. Stepp,
Department of Psychiatry, University of Pittsburgh School of Medicine

Lan Yu,
Department of Psychiatry, University of Pittsburgh School of Medicine

Joshua D. Miller,
Department of Psychology, University of Georgia

Michael N. Hallquist,
Department of Psychiatry, University of Pittsburgh School of Medicine

Timothy J. Trull, and
Department of Psychological Sciences, University of Missouri

Paul A. Pilkonis
Department of Psychiatry, University of Pittsburgh School of Medicine

Abstract

Mounting evidence suggests that several inventories assessing both normal personality and personality disorders measure common dimensional personality traits (i.e., Antagonism, Constraint, Emotional Instability, Extraversion, and Unconventionality), albeit providing unique information along the underlying trait continuum. We used Widiger and Simonsen's (2005) pantheoretical integrative model of dimensional personality assessment as a guide to create item pools. We then used Item Response Theory (IRT) to compare the assessment of these five personality traits across three established dimensional measures of personality: the Schedule for Nonadaptive and Adaptive Personality (SNAP), the Temperament and Character Inventory (TCI), and the Revised NEO Personality Inventory (NEO PI-R). We found that items from each inventory map onto these five common personality traits in predictable ways. The IRT analyses, however, documented considerable variability in the item and test information derived from each inventory. Our findings support the notion that the integration of multiple perspectives will provide greater information about personality while minimizing the weaknesses of any single instrument.

Keywords

personality measurement; Item Response Theory

A variety of dimensional personality inventories have been advanced by several research groups and available data do not clearly support one proposal over another (Clark, 2007). Moreover, many empirical articles attempting to map the structure of personality originate from a particular theory or instrument, with relatively little cross-talk among theoretical

Correspondence concerning this article should be addressed to Stephanie D. Stepp, Department of Psychiatry, University of Pittsburgh School of Medicine, Western Psychiatric Institute and Clinic, 3811 O'Hara St., Pittsburgh, PA 15213. steppsd@upmc.edu.

perspectives (but see Clark & Livesley, 2002; Widiger, Livesley, & Clark, 2009, for examples). There is much evidence to suggest that dimensional personality inventories measure five common underlying traits, namely Antagonism, Constraint, Emotional Instability, Extraversion, and Unconventionality (e.g., see Widiger, 2011a). Widiger and Simonsen (2005) provide a theoretical framework toward an integrative dimensional model of personality, which highlights this common hierarchical structure found across 18 different dimensional personality inventories. We propose to use Widiger and Simonsen's (2005) theoretical work as a heuristic to examine the links between three common dimensional personality inventories: the Schedule for Nonadaptive and Adaptive Personality-2nd edition (*SNAP-2*; Clark, Simms, Wu, & Casillas, in press), the Temperament and Character Inventory (*TCI*; Cloninger, Przybeck, Svrakic, & Wetzel, 1994), and the Revised NEO Personality Inventory (*NEO PI-R*; Costa & McCrae, 1992a).

Common Dimensional Personality Traits

Theoretical and empirical reviews of the latent structure of personality provide increasing evidence for the salience of five dimensional personality traits that cut across theoretical perspectives and inventories (Clark, 2007; Krueger, 2005; Widiger, 2011a). Available evidence also supports the notion that abnormal and normal personality share a common hierarchical structure, with maladaptive traits representing extreme levels of normal traits (Markon, Krueger, & Watson, 2005; O'Connor, 2002). In addition, the redundancy across inventories observed in several incremental validity studies (e.g., Reynolds & Clark, 2001; Stepp, Trull, Burr, Wolfenstein, & Vieth, 2005) suggests that integrating items from competing inventories onto a common scale may be the most fruitful path for uncovering information about underlying personality traits.

Based on a thorough review of the empirical literature, Widiger and Simonsen (2005) provided a schematic that maps most of the 18 proposals for dimensional personality assessment onto five broad traits of Antagonism, Constraint, Emotional Instability, Extraversion, and Unconventionality. For example, the Mistrust scale from the SNAP and the Agreeableness scale from the NEO PI-R are hypothesized to measure an underlying Antagonism factor. Although evidence generally supports the notion of a common latent structure across personality inventories (Clark & Livesley, 2002; Markon et al., 2005), further study is required to validate the assertion that a shared factor structure underlies most of the 18 dimensional personality inventories. Moreover, the essential notion that scales developed from seemingly different theoretical perspectives are adequately represented by a single dimension needs to be validated. For example, does the psychometric evidence support the assertion that items from the Negative Temperament scale from the SNAP are isomorphic with items from the Neuroticism scale from the NEO PI-R, or are these scales better conceptualized as two related, but distinct, factors?

In addition to grappling with issues of factor structure and conceptual isomorphism across theoretical perspectives, an integrative view of dimensional personality assessment must address the pragmatic issue of comparing information provided by each item from competing inventories to enable researchers and clinicians to assess personality traits more efficiently, precisely, and flexibly. In view of the shared latent structure of abnormal and normal personality traits, the integration of multiple inventories will be best served by a final product that provides information across the range of trait levels (i.e., trait information at normal, subclinical, and clinical levels). Thus, it may be sensible to combine items from self-report inventories developed to assess maladaptive traits (e.g., the SNAP), with items gleaned from normal personality instruments (e.g., the NEO PI-R), as clinical inventories may provide better information about extremely high or low trait levels. That said, some have suggested that measures of maladaptive personality are redundant with normal

personality inventories (Costa & Mc-Crae, 1992b) and some recent evidence supports this assertion (e.g., Walton, Roberts, Krueger, Blonigen, & Hicks, 2008).

Previous research has utilized joint factor analyses of normal and abnormal personality instruments to argue that a common factor structure underlies different personality inventories and that normal and abnormal personality fall on a common continuum (Markon et al., 2005; Watson, Clark, & Chmielewski, 2008). Although factor analysis is a useful tool for understanding whether common traits are shared across personality inventories, it does not provide any information about whether abnormal and normal inventories yield information at different points along the trait continuum (i.e., ranging from normal to clinical), nor does it yield detailed feedback about which inventories (or items) provide the most information about the latent traits.

Item Response Theory

Item Response Theory (IRT) represents a class of modern psychometric techniques that model levels of a putative latent trait (e.g., Neuroticism) as a function of item characteristics, in which the probability of correct item response is modeled as a function of latent trait theta (θ) and one or more item parameters (Embretson & Reise, 2000; Lord, 1980). Of particular import to our study, IRT methods provide specific feedback about the position along θ, where each item or inventory provides the most psychometric information about the trait. For example, an item tapping intense expressions of anger (e.g., throwing objects) would likely provide more information about Emotional Instability at high levels of θ than an item about occasional arguments with romantic partners. One advantage of IRT is that individuals' θ estimates are independent of the number of items and the specific items used in the population for calibration. Thus, even when individuals take different sets of items with different response options, (resulting in different patterns of missingness), the data can be combined and *concurrently calibrated* or *linked*, estimating item parameters across three measures within a single latent trait model (i.e., on one single computer run; Lord, 1980).

Item- and test-information functions in IRT are estimated on the same latent trait scale, yielding psychometric information that is directly comparable across inventories (Reise & Henson, 2003). For example, if one were attempting to develop an IRT-informed integrative scale of Extraversion, it would be important to know which items provide maximum information about the latent trait across the broadest range of θ levels, from extremely shy to extremely outgoing. Because item characteristic estimates are not tied to particular inventories or theories per se, it is likely that items from several inventories would provide the most valid trait scale, and inclusion of items from both normal and abnormal inventories may be crucial to ensure that average and extreme trait levels are represented. In summary, IRT models are ideal for the present purposes because (1) they yield detailed information about the position along the trait continuum accounted for by particular inventories (or items from those inventories), (2) latent trait estimates for items and inventories are directly comparable across scales, and (3) they provide specific feedback about which tests provide the most valid information about the underlying traits.

Integrating Dimensional Personality Inventories

To identify the items that optimally measure five common dimensional personality traits from competing dimensional personality assessment inventories, we selected three measures that represent different approaches to test construction: the SNAP-2, which was derived from a bottom-up approach to measure personality pathology; the TCI, which was rationally derived from Cloninger's (1987) psychobiological theory of personality; and the NEO PI-R, which was derived from factor analytic work on normal personality.

Widiger and Simonsen (2005) developed a conceptual framework for creating a pantheoretical scale for assessing five common personality traits. This study sought to extend their conceptual work by examining how three widely used personality inventories map on to these five common personality traits. We sought to demonstrate that items drawn from different inventories map onto the five personality traits as predicted by Widiger and Simonsen (2005) by linking measures using IRT models. To demonstrate this, we refined and reorganized the scales to include only the items that provided the most information about the underlying trait and examined the proportion of items contributed by each inventory in the measurement of the underlying trait to determine which inventory provided the most information regarding the underlying trait. For example, does the NEO PI-R provide more information about Conscientiousness compared to the SNAP-2 and TCI? Second, we linked these refined item pools from the individual inventories to create five pantheoretical scales that optimally measure Antagonism, Constraint, Emotion Instability, Extra-version, and Unconventionality. To develop these item pools, we used IRT to identify the best performing items for each personality trait and linked the SNAP-2, TCI, and NEO PI-R.

Method

Participants and procedures—Participants were recruited from undergraduate psychology courses at two universities, psychiatric inpatient, psychiatric outpatient, medical, and community settings. Written informed consent was obtained from all participants before administration of the questionnaires. Undergraduate student participants were compensated with credit toward their psychology course grade and patient and community participants were compensated monetarily for their participation.

In the student sample (n = 1,517), participants completed items from the NEO PI-R, SNAP-2, and TCI. Given the high participant burden involved in completing these three personality batteries (870 items total), we used a balanced incomplete block design (BIBD; Campbell, Sengupta, Santos, & Lorig, 1995; Cochran & Cox, 1957; Van de Linden, Veldkamp, & Carlson, 2004). This design allowed participants to complete a reduced number of items from the three measures while allowing us to calibrate our IRT models. We created blocks so that each participant would complete 217 items. The resulting BIBD ensured that each item was presented with each other item an equal number of times within blocks. To create the BIBD, we first deleted two items from the SNAP-2 validity scales (so that the total number of items was divisible by 28). Second, we partitioned the total item pool into 28 sets of 31 items each. Third, we assigned 7 sets to each participant according to BIBD plan 11.39 (p. 482, Cochran & Cox, 1957). There were 36 blocks (b) and each block had 7 sets (k) of items in this design. And there were total of 28 sets (t) of items. Therefore, each set of item had 9 replicates (r) within the 36 blocks. Each participant received one booklet with a total of 217 items. Each block was administered 42 times (42 administrations × 36 blocks = 1, 512 participants) and the first five blocks were administered an additional time each (1 administration × 5 blocks = 5 participants), which yielded 1,517 participants. The optimal number of participants for this design is 5–10 times the number of items any one participant takes, resulting in an optimal sample size of 1,000–2,000 for the current study, which allowed us to be confident in our parameter estimate with this level of planned missingness in our design. This BIBD method is commonly used in the field of educational testing (cf. National Assessment of Educational Progress [NAEP]; Rock & Nelson, 1992; Yamamoto & Mazzeo, 1992) and is an appropriate method for using planned missingness to reduce the number of items that any one participant is administered (cf. Campbell et al., 1995; Rock & Nelson, 1992; Yamamoto & Mazzeo, 1992).

To augment our student sample, we included archival data from three clinical samples consisting of psychiatric patients, medical patients, and community participants (n = 400). The personality measures were administered as part of a larger battery of questionnaires and interviews regarding interpersonal functioning in PDs. These participants were included to enrich our sample with scores at the higher end of the personality dimensions of interest, increasing the generalizability of our findings. They were not included in the BIBD. Alternatively, they took only one of the measures. The inclusion of the patient-community participants resulted in missing data patterns that varied from the planned missingness in the student sample. However, the large sample allows adequate coverage of the overall data matrix which enabled us to estimate the parameters reliably. Specifically, the first sample consisted of psychiatric patients that were administered 167 items from the NEO PI-R (n = 134),[1] the second sample consisted of psychiatric and medical patients and community participants that were administered the TCI (n = 130), and the third sample consisted of psychiatric outpatients that were administered the SNAP-2 (n = 136).

There were 1,917 participants in the total sample, collapsed across the student and patient-community samples. The majority of the sample was female (64%) and White (85%). Eight percent of the sample identified themselves as African American and 4.3% identified as Asian. Two percent of the sample identified their ethnicity as Hispanic. The mean participant age in the total sample was 21.9 years (SD = 8.2, range = 18–60). The mean participant age for the three clinical samples was 34.47 (SD = 9.37), 38.35 (SD = 10.67), and 37.97 (SD = 10.63), who took the NEO PI, TCI, and SNAP-2, respectively. The mean participant age for the student sample was 18.73 (SD = 1.15).

Measures

Schedule for Nonadaptive and Adaptive Personality-2nd Edition—The SNAP-2 is a multidimensional self-report measure of 12 pathological personality traits (e.g., Aggressiveness, Dependency, Exhibitionism, Workaholism, Eccentric Perceptions), as well as the Big Three personality factors (i.e., Positive Affectivity, Negative Temperament, and Disinhibition) that contains 390, true-false self-report items. Using items drawn from *DSM–III* PD criteria, personality pathology concepts from other research programs (e.g., psychopathy; Cleckley, 1964), and trait-like symptoms of Axis I conditions (e.g., dysthymia), raters derived 22 item clusters, which were subsequently factor analyzed. Results from exploratory factor analysis indicated that 12 dimensions of pathological personality traits characterized the SNAP items, and the best indicators of these dimensions were retained to form personality pathology scales. Subsequent validation of the SNAP indicates the measure has strong psychometric properties, is correlated with *DSM–IV* PDs in predicted ways, and successfully distinguishes among distinct forms of personality pathology (Morey et al., 2003).

Temperament and Character Inventory—The TCI is a 240-item, true-false self-report inventory. It is a broadband personality assessment instrument developed a priori based on Cloninger's (1987) seven-factor psychobiological theory of personality, which was strongly influenced by genetic and family studies of personality; longitudinal studies of personality stability and change; humanistic and transpersonal notions of personality development; and basic conditioning/ learning studies in animals and humans (Cloninger, 1987; Cloninger, Svrakic, & Przybeck, 1993). The TCI measures four dimensions of temperament (Harm Avoidance, Novelty Seeking, Persistence, and Reward Dependence) and three dimensions of character (Cooperativeness, Self-Directedness, and Self-Transcendence). The seven main

[1]As a result of using archival data, the earlier version of the NEO PI-R, the NEO PI (Costa & McCrae, 1985), was administered to the patients in this sample. The 167 items from the NEO PI that are consistent with those from the NEO PI-R were included in analyses.

TCI dimensions comprise 25 facets. The TCI has generated a large and influential body of literature, spanning topics including the genetic heritability of personality (Ando et al., 2002), personality variability within Axis I diagnoses (e.g., Fassino et al., 2002), and the impact of personality on psychotherapy outcome (Joyce, Mulder, McKenzie, Luty, & Cloninger, 2004).

Revised NEO Personality Inventory—The NEO PI-R is self-report inventory developed to measure the Big Five personality domains: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. Each of the five broad domains is divided into six facets and each facet is assessed by eight items. It consists of 240 statements for which participants rate their level of agreement on a 5-point scale, with 1 indicating *strongly agree* and 5 indicating *strongly disagree*. Exploratory factor analyses of large sets of trait adjective self-ratings (e.g., "friendly," "courageous") or short trait descriptions (e.g., "I am not a worrier.") consistently yield the Big Five personality domains. Several independent studies have replicated the essential factor structure of this inventory (e.g., Savla, Davey, Costa, & Whitfield, 2007; Wu, Lindsted, Tsai, & Lee, 2008). The NEO PI-R has been used extensively in empirical studies of normal and abnormal personality (e.g., Markon et al., 2005; Yamagata et al., 2006), and has been accepted among many as the dominant Big Five personality assessment instrument (Clark, 2007).

Creating item pools—We used Widiger and Simonsen's (2005) pantheoretical integrative model of personality disorder classification to develop item pools for Antagonism, Constraint, Emotional Instability, Extraversion, and Unconventionality from the SNAP-2, TCI, and NEO PI-R. Widiger and Simonsen (2005), classified items from these three instruments into the five scales of interest. Table 1 lists the scales from each personality measure that were used to create the five item pools. The Entitlement SNAP-2 scale was included in both Antagonism and Extraversion (Widiger & Simonsen, 2005). Table 1 lists the scales from each personality inventory as well as the number of items from each of the scales used in the analyses. The initial item pools consisted of a large set of items for each personality dimension: Antagonism (214 items: 48 NEO, 93 SNAP, and 73 TCI items), Constraint (232 items: 48 NEO, 92 SNAP, and 92 TCI items), Emotional Instability (190 items: 48 NEO, 62 SNAP, and 80 TCI items), Extraversion (186 items: 48 NEO, 77 SNAP, and 61 TCI items), and Unconventionality (96 items: 48 NEO, 15 SNAP, and 33 TCI items). Response frequencies for each item were inspected before data analysis to ensure that all scale values were endorsed by at least 1.0% of the sample.[2]

Analytic Approach

The data analytic approach consisted of two major components. First, we conducted factor analyses to ensure sufficient unidimensionality for the proposed models. Next, we used concurrent IRT calibrations to link the items from the three personality inventories.

Dimensionality—Before examining dimensionality, the sample was randomly divided into two groups of about equal size: a development ($n = 979$) and a validation ($n = 938$) sample. To ensure that the assumption of unidimensionality was met for IRT analysis, we conducted exploratory factor analysis (EFA) of each item pool representative of Antagonism, Constraint, Emotional Instability, Extraversion, and Unconventionality with the development sample. This was followed by a confirmatory factor analysis (CFA) for each personality dimension with the validation sample. All factor analyses were run using a mean and variance-adjusted weighted least squares estimator (WLSMV) in Mplus 5.1

[2]Additional tables providing item content and response frequencies for the Antagonism, Constraint, Emotional Instability, Extraversion, and Unconventionality initial item pools are available upon request from the corresponding author.

(Muthén & Muthén, 2007). The factor loading pattern matrix was examined to determine whether individual items loaded on a single factor. The strength of item loadings was considered poor if they did not reach a value of .35. Items not reaching this threshold were discarded, and the EFA was rerun. This process was repeated to ensure all item loadings were greater than or equal to .35. We chose to use .35 instead of a stricter cutoff at this stage to retain more items and the most comprehensive item pools. After the final item pool for the EFA was determined, we examined the scree plot and eigenvalues to evaluate dimensionality. This process yielded a refined pool of items, indicating the content overlap for the SNAP-2, TCI, and NEO PI-R regarding each personality dimension.

Following the EFA, the items were then submitted to a single-factor CFA using the validation sample. We assessed absolute fit of the confirmatory models using global fit indices, including the chi-square statistic (χ^2), comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). The conventional cutoff values for the CFI and TLI, are .90 or greater for acceptable fit, and .95 or greater for good fit. Additionally, RMSEA values between .05 and .08 represent an acceptable fit, while values less than .05 indicate a good fit (McDonald & Ho, 2002). Given these established standards of CFA fit statistics, we also noted the caution of mechanical use of CFA fit criteria as a “permission slip” for modeling data using IRT, because CFA fit results can be affected dramatically by large number of items and skewed data distributions (Cook, Kallen, & Amtamnn, 2009), which are common characteristics of personality data. For example, Hays and his colleagues (2007) reported CFI = 0.95 and RMSEA = 0.12 as sufficiently unidimensional for IRT analysis on their Physical Functioning item bank development. Similarly, Revicki and his colleagues (2009) reported CFI = 0.902, TLI = 0.991, and RMSEA = 0.156 as sufficiently unidimensional for further IRT analysis. Buysse and his colleagues (2010) reported RMSEA = 0.140, TLI = 0.957, and CFI = 0.843 for Sleep Disturbance item bank and RMSEA = 0.157, TLI = 0.955, and CFI = 0.812 for Sleep-Related Impairment item bank as sufficiently unidimensional for further IRT analysis. Therefore, we also checked the scree plot of eigenvalues and the ratio of the first two eigenvalues from EFA in judging unidimensionality. Alternatives to the basic one-factor model were considered to improve fit (cf. McHorney & Cohen, 2000).

IRT calibration—Following the factor analyses to determine which items across the three personality measures were sufficiently unidimensional for IRT analysis, we further examined item response distributions because item parameter estimates may be biased for items with sparse cells (Thissen, 2003). Although it is common that the item response distributions are skewed for questionnaire data, the item response categories with few observations (less than 5% of total frequencies) have to be combined with their adjacent categories to achieve reliable estimates. We removed those items with at least one response category having less than 5% of total frequencies to keep the original response scales. Because we had such large item pools, we were able to delete these items; there was no need to retain items with sparse cells. Then we concurrently calibrated all items from each personality dimension of interest using the GRM for the NEO PI-R and 2PL for the SNAP-2 and TCI in Multilog 7.03 (Thissen, 2003). The advantages of this concurrent calibration include: (1) it retains the integrity of the original scale, (2) it allows individuals to take different sets of items, and (3) it tolerates inventories of differing lengths and rating scales (cf. McHorney & Cohen, 2000; Reise & Waller, 2009). The Multilog program for the GRM estimates a slope (*a*) parameter and four location (*b*) parameters for each five-category NEO PI-R item. The Multilog program 2PL model estimates a slope (*a*) and one location (*b*) parameter for each two-category SNAP-2 and TCI item.

After the initial concurrent calibration, we examined items in terms of item information and item content. Items with low item information were considered to be poor items in IRT

calibration. Item discrimination parameter estimates affect an item's total information function. The higher the discrimination parameter, the more peaked the item information function. Thus, items with discrimination parameter estimates less than 1.00 provide little information and were removed from the item pool. The reduced item pool was then recalibrated.

Results

Descriptive Statistics

Summed scores of the item pools were computed and examined for each personality dimension. The mean of the summed scores for the 214 Antagonism items was 41.51 (*SD* = 14.59) for the total sample, 43.31 (*SD* = 13.66) for the student sample, and 34.42 (*SD* = 15.97) for the nonstudent sample. The summed score distributions for Antagonism were not significantly skewed in the total sample or either of the subsamples when considered separately (total sample, skew = .27; student sample = .56; nonstudent sample = .08).

The mean of the summed 232 Constraint items for the total sample was 62.68 (*SD* = 14.51), 64.83 (*SD* = 14.25) for the student sample, and 54.35 (*SD* = 12.39) for the nonstudent sample. The summed score distributions for Constraint were also not significantly skewed (total sample, skew = .30; student sample = .38; nonstudent sample = −.20).

For the total sample, the mean of the summed scores for Emotional Instability was 52.61 (*SD* = 31.69), and 47.15 (*SD* = 14.82) and 73.77 (*SD* = 59.10) for the student and nonstudent samples, respectively. The summed score distribution in the total sample for Emotional Instability was positively skewed (total sample, skew = 2.43); however, the distributions were not significantly skewed when examining the subsamples separately (student sample = .73; nonstudent sample = .74).

The mean of the summed scores for the 186 Extraversion items was 66.13 (*SD* = 25.74) for the total sample, 66.40 (*SD* = 15.90) for the student sample, and 65.06 (*SD* = 47.40) for the nonstudent sample. The summed score distributions for Extraversion were not significantly skewed (total sample, skew = .86; student sample = .14; nonstudent sample = .72).

Finally, the mean of the summed 96 Unconventionality items was 48.76 (*SD* = 33.85) for the total sample, 46.44 (*SD* = 12.10) for the student sample, and 57.74 (*SD* = 70.29) for the nonstudent sample. The summed score distribution in the total sample for Unconventionality was positively skewed (total sample, skew = 2.13); however, the distributions were not significantly skewed when examining the subsamples separately (student sample = .64; nonstudent sample = .78).

Because of the BIBD, internal consistency reliability coefficients could not be calculated for the student sample (i.e., because of planned missingness, too few student cases completed all the items from a scale or instrument for α to be calculated). However, reliability coefficients were calculated for the nonstudent sample by instrument within each of the five item banks. For the Antagonism item bank, α = .80 for the NEO items, α = .95 for the SNAP items, and α = .84 for the TCI items. The internal consistency coefficients (α) for the Constraint item pool were .83 for the NEO items, .88 for the SNAP items, and .68 for the TCI items. For the Emotional Instability item bank, α = .92 for the NEO items, α = .94 for the SNAP items, and α = .93 for the TCI items. For the Extra-version item bank, α = .87 for the NEO items, α = .94 for the SNAP items, and α = .78 for the TCI items. Lastly, the internal consistency coefficients (α) for the Unconventionality item bank were .90 for the NEO items, .86 for the SNAP items, and .85 for the TCI items.

Assessing Dimensionality

For each of the personality dimensions, the scree plot of eigenvalues from the EFA in the development sample was suggestive of a single factor, with the first value larger than the others. Specifically, the ratio of the first to the second eigenvalue was greater than 2.0 for all traits: Antagonism (3.10), Constraint (3.54), Emotional Instability (3.92), Extraversion (5.90), and Unconventionality (2.03). Although the ratio of the first two factors for Unconventionality was less than 3, all item loadings were above .35. Specifically, in the final single factor solution, item loadings ranged from .35 to .79 for Antagonism, .35 to .93 for Constraint, .38 to .88 for Emotional Instability, .43 to .92 for Extra-version, and .38 to . 76 for Unconventionality.

The basic 1-factor CFA model for Extraversion fit well to the validation sample data (CFI = .901, TLI = .930, RMSEA = .035). The RMSEA index indicated at least acceptable fit for the remaining four dimensions. However, the CFI/TLI global fit indices indicated less than adequate fit for the basic 1-factor CFA model: Antagonism (CFI = .821, TLI = .838, RMSEA = .031), Constraint (1.85, CFI = .773, TLI = .781, RMSEA = .030), Emotional Instability (2.00, CFI = .861, TLI = .871, RMSEA = .030), and Unconventionality (3.01, CFI = .684, TLI = .696, RMSEA > .046).

Similar to the method used by McHorney and Cohen (2000), we next tried alternatives to the basic 1-factor CFA model. Correlated errors among indicators from the same measure were specified to reflect that some of the covariance among the items from the same inventory because of measurement error. The addition of correlations among the residuals of the items from the same measure (cf. McHorney & Cohen, 2000) improved the CFI and TLI indexes to acceptable levels: Antagonism (CFI = .900, TLI = .907, RMSEA = .024), Emotional Instability (CFI = .910, TLI = .909, RMSEA = .027), and Unconventionality (CFI = .957, TLI = .948, RMSEA = .019); with the exception of Constraint (CFI = .884, TLI = .867, RMSEA = .024), where the CFI and TLI indexes remained slightly outside the acceptable range. Although the CFA fit indices were slightly outside the traditional acceptable range, CFA fit values were found to be sensitive to data distribution and number of items (Cook, Kallen, & Amtmann, 2009). As they suggested, using traditional cutoffs and standards for CFA fit statistics is not recommended for establishing unidimensionality of item banks because the impact of distribution and item number was quite large in some cases. We also examined alternative models that posited additional factors (e.g., a model that allowed items from different measures to load on method factors). These solutions did not yield superior fit relative to the single-factor model with correlated errors. Further, we found in the literature about robustness of item parameter estimation to assumptions of unidimensionality: Studies using multidimensional data generated by a factor analytic approach tend to show that a unidimensional IRT model is robust to moderate degrees of multidimensionality (Harrison, 1986; Kirisci, Hsu, & Yu, 2001; Reckase, 1979). Given the EFA results and eigenvalues as well as CFA cutoff values used in previous IRT studies, we judged these results overall provide sufficient evidence of unidimensionality.

Based on these results, we determined 92 Antagonism, 108 Constraint, 114 Emotional Instability, 91 Extraversion, and 49 Unconventionality items were sufficiently unidimensional for IRT analysis (Reeve et al., 2007; Tate, 2003; Zwick & Velicer, 1986). Table 1 delineates the name of the scales as well as the number of items retained from each of these scales following the factor analyses. These item pools reflected the overlapping content in the SNAP-2, TCI, and NEO PI-R. Specifically, the three measures overlapped in measuring anger and verbal and physical aggression for Antagonism; premeditation and perseverance for Constraint; stress susceptibility, negative affectivity, and impulsiveness for Emotional Instability; high activity, positive affectivity, and sociability for Extraversion; and curiosity, unusual experiences, and connectedness for Unconventionality.

IRT Calibration

For item response frequency distribution examinations, all items having at least one response category with less than 5% of total frequencies were NEO items. This was expected since NEO had 5 response categories while TCI and SNAP had only 2 response categories. Thirty-six NEO items were removed accordingly: 4 items from Antagonism, 13 items from Constraint, 4 items from Emotional Instability, 14 items from Extraversion, and 1 item from Unconventionality. Using the items banks selected from the factor analyses for each personality dimension, the slope estimates from the initial concurrent calibration indicated a wide range of item discrimination for Antagonism (*a*s = .48–2.51), Constraint (*a*s = .42–2.05), Emotional Instability (*a*s = .55–2.75), Extra-version (*a*s = .64–2.93), and Unconventionality (*a*s = .38–1.58), with many items providing poor discrimination (i.e., *a*s < 1.0). Based on these item parameters, we further reduced the number of items to yield a final number of items for each trait: Antagonism (24 items; 1 NEO, 15 SNAP, and 8 TCI items), Constraint (19 items; 7 NEO, 8 SNAP, and 4 TCI items), Emotional Instability (56 items; 9 NEO, 27 SNAP, and 20 TCI items), Extraversion (59 items; 9 NEO, 45 SNAP, and 5 TCI items), and Unconventionality (18 items; 7 NEO, 4 SNAP, and 7 TCI items). For illustrative purposes, Table 2 provides the parameter estimates and their standard errors from the final concurrent calibrations in rank order of their slope parameter estimates for the Constraint domain. The parameter estimates for the remaining domains are available upon request from the first author. The items retained for each scale maintained a reasonable balance of content and provided adequate coverage for the dimensions of interest. Interestingly, the final calibration for each of the personality dimensions contains items from all three measures, which suggests that each measure provides some utility in measuring the underlying trait. The slope estimates of the final item pools for Antagonism (*a*s = 1.11–3.93), Constraint (*a*s = 1.09–1.99), Emotional Instability (*a*s = 1.00–2.59), Extraversion (*a*s = 1.04–3.03), and Unconventionality (*a*s = 1.07–2.45) indicated considerable variation in item discrimination. The location parameters for the Antagonism (*b*s = −1.70–2.04), Constraint (*b*s = −2.81–2.49), Emotional Instability (*b*s = −2.75–2.57), Extraversion (*b*s = − 2.70–1.90), and Unconventionality (*b*s = −2.50–1.61) reflect a sizable range of the underlying personality dimension of interest.

Next, we compared the psychometric information at the test level. One of the advantages of concurrent calibration is that all three measures are on the same metric. For each of the personality dimensions, the test information curves (see Figure 1) were plotted for the three measures separately and combined. Panel 1a displays the test information curves for Antagonism. For these three measures, the TCI provided the most information, followed by the SNAP-2. However, the SNAP-2 covered a broader θ range relative to the TCI. The NEO PI-R covered a broad range; albeit providing relatively little information to the test information curve (it contributed only one item). Panel 1b displays the test information curve for Constraint. The NEO PI-R provided the most information and covered the widest θ range followed by the SNAP-2 and TCI. Panel 1c displays the test information curves for Emotional Instability. The SNAP-2 provided more information at a narrower θ range of −1.5 to 2.5, while the NEO PI-R provided more information at the low tail (i.e., less than −1.5). The TCI covered a slightly narrower θ range for Emotional Instability than the SNAP-2 and provided less information. Panel 1d displays the test information curves for Extraversion. Similar to Emotional Instability, the SNAP-2 provided more information at a θ range of −3.0 to 2.5, while the NEO PI-R provided more information at the two tails (i.e., less than −3.0 and larger than 2.5). The TCI provided the least information. Finally, Panel 1e displays the test information curves for Unconventionality. The NEO PI-R covered the widest θ range and provided the most information, followed by the TCI and SNAP-2. Across all five personality domains, combining items from the three individual scales provides the maximum amount of information with the most precision across the widest range of θ when

compared to the information provided by the subset of items retained from each individual scale.

Discussion

Our goal was to map the scales from the SNAP-2, NEO, and TCI onto five common personality traits (i.e., Antagonism, Constraint, Emotional Instability, Extraversion, and Unconventionality) using Widiger and Simonsen's (2005) model as a guide. Thus, we specified the factors we sought to measure in advance. The results demonstrated that items from the SNAP-2, TCI, and NEO PI-R overlap in their measurement of Antagonism, Emotional Instability, Extraversion, Constraint, and Unconventionality in predictable ways. For example, items from the Negative Temperament, Self-harm, and Dependency SNAP-2 scales, Harm Avoidance and Self-Directedness TCI scales, and Neuroticism NEO PI-R scale all overlap in measuring Emotional Instability. Our results are consistent with recent meta-analytic and empirical evidence demonstrating that five personality dimensions are shared among dimensional measures of abnormal and normal personality (Markon et al., 2005; O'Connor, 2005; Samuel, Simms, Clark, Livesley, & Widiger, 2010). Our findings are consistent with Samuel and colleagues (2010) demonstrating that common latent personality dimensions cut across the NEO PI-R as well as personality measures intended to assess more extreme variants of personality pathology (SNAP and Dimensional Assessment of Personality Pathology-Basic Questionnaire; Livesley & Jackson, 2011). Additionally, items from the NEO PI-R provided more information at the lower range of the latent trait compared to the measures of maladaptive personality.

The integration of multiple perspectives, specifically the SNAP-2, TCI, and NEO PI-R, provides the most information about the underlying trait, thereby minimizing the weaknesses of any single perspective. Our final item banks integrated information functions across personality scales from competing inventories, which provides the most information about the underlying trait compared with the subset of items retained from each individual inventory. Integrating multiple information functions always leads to an increase in the amount of information provided. For example, for Extraversion, the NEO PI-R provides the most information at the high and low ends of trait, whereas the SNAP-2 provides more information in the middle range of the trait. Each measure contributed items to the final calibrated item pool, suggesting that each measure provides some utility in measuring the underlying domains of interest. Even though factor analyses pruned unrelated items, we further pruned items if they provided little information to the construct (i.e., items with slope parameters <1.00 were eliminated). Thus, items from any measure could have been eliminated at this stage and it is important to note that all three scales were represented in the final item pool. These results are also consistent with past reports that two-, three-, and four-factor models of personality, which have all been proposed as alternative accounts for normal and abnormal personality (e.g., Eysenck & Eysenck, 1976; O'Connor & Dyce, 1998; Tellegen, 2000), are well-represented within a five-factor model hierarchy (Digman, 1997; Markon et al., 2005).

Comparing the SNAP, TCI, and NEO

Our data analytic strategy enabled us to directly compare scales from the SNAP-2, TCI, and NEO PI-R by linking the scales to the same metric. We retained items from each of the inventories that provided the best psychometric information to represent the five personality traits. By putting items across inventories on the same underlying latent trait scale, we were able to provide information to researchers and clinicians regarding the "best" functioning items from each of the three inventories. We did not intend to create a new measure based on these three inventories, rather to provide information across them.

The results demonstrated the relative strengths and weaknesses of the SNAP-2, TCI, and NEO PI-R when measuring Antagonism, Constraint, Emotional Instability, Extraversion, and Unconventionality as separate personality dimensions. Specifically, the SNAP-2 and TCI provided information at narrower bands of θ for all personality traits relative to the NEO PI-R with the exception of Antagonism. Moreover, the SNAP-2 provided *more* information at *narrow* bands of Antagonism, Emotional Instability, and Extraversion relative to the NEO PI-R. Additionally, the NEO PI-R provided the *most* information across all bands of Constraint and Unconventionality relative to the SNAP-2 and TCI. Lastly, the TCI provided the most information for Antagonism but provided the least information for Constraint, Emotion Instability, and Extraversion compared to the SNAP-2 and NEO PI-R. The TCI provided more information and measured a wider range of Unconventionality relative to the SNAP-2 but not the NEO PI-R.

A comparison of item parameters also demonstrated that the NEO PI-R items covered a wider θ range than SNAP-2 and TCI items across four personality traits. This can be partly attributed to the structure of the measure. Poly-tomous response items generally cover a wider theta range than dichotomous items since each polytomous response item can be treated as a series of dichotomous items. However, this structural difference does not fully account for our findings because increasing the number of categorical responses does not uniformly increase the information for θ levels over the entire θ range (Muraki, 1993). Moreover, information from different items, even if indicators have different response options, when combined in an item pool allows us to directly compare the information provided by the different scales. Each NEO item is spread out across four decision points (Strongly Disagree vs. Disagree; Disagree vs. Neutral; Neutral vs. Agree; and Agree vs. Strongly Agree). Thus, each NEO item can be thought of as 4 binary discriminators (*k*-1 response options). When NEO, SNAP, and TCI indicators comprise an item pool, we are able to compare the aggregate information provided by all the items from the scale (test information curve). Concurrent calibration ensured that items from different measures could be compared on the same metric. Although the test information is a sum of individual item information, which means that the height of the information curve is often affected by the number of items included in the item pool, this does not inevitably mean that the more items included, the more information the corresponding “test” will have. With IRT, a longer test does not provide more information than a shorter test (Embretson & Reise, 2000).

By combining items from different scales, we are able to provide more information about the underlying construct than any one scale. This observation is important because many personality assessments are created with little crosstalk between them. This has resulted in a plethora of personality inventories often viewed as competing for “favored” status. However, we have demonstrated that, by using IRT calibrations across different instruments, we can equate them on the same metric and measure the broadest range of theta with the most precision when compared with any one instrument in isolation.

Constructs of Antagonism, Constraint, Emotional Instability, Extraversion, and Unconventionality

In addition to demonstrating the relative strengths and weaknesses of the SNAP-2, TCI, and NEO PI-R, latent trait models inform our understanding of the constructs of interest. The factor analyses and concurrent calibrations culled items from all inventories that did not sufficiently overlap, which ensured the final item pool was representative of the unidimensional latent trait. Our results distilled core features of the construct, while culling peripheral aspects.

Antagonism—The majority of the Antagonism items measured argumentativeness (e.g., "I often get into arguments with my family and friends"), physical aggression (e.g., "When I get angry, I am often ready to hit someone"), animosity (e.g., "I enjoy getting revenge on people who hurt me"), and self-centeredness (e.g., "Some people think I am selfish and egotistical"). The original set of items were from the Aggressiveness and Manipulativeness SNAP-2 scales, the Cooperativeness TCI scale, and the Agreeableness NEO PI-R scale (i.e., NEO PI-R facets: Trust, Straightforwardness, Altruism, Compliance, and Modesty). However, the final calibrated scale only contained one item from the NEO PI-R Antagonism scale. Contrary to Widiger and Simonsen's (2005) predictions, no items from the Reward Dependence TCI scale, Entitlement and Dependency SNAP-2 scales, and the Tender-mindedness facet of the Agreeableness NEO PI-R scale were represented in the final item pool. Interestingly, a relatively large number of items were dropped from the Antagonism item pool because items factor loadings were <.35, indicating that the Antagonism construct does not appear to be as unidimensional when compared with the other personality domains across the NEO, SNAP, and TCI.

Constraint—The majority of the Constraint items measured premeditation (e.g., "I plan ahead carefully when I go on a trip"), perseverance (e.g., "When I start a task, I am determined to finish it"), and diligence (e.g., "I enjoy working hard"). Items from the final pool originated from the Workaholism, Impulsivity, Propriety, and Disinhibition SNAP-2 scales, Novelty Seeking and Self-Directedness TCI scales, and Conscientiousness NEO PI-R scale (i.e., all facets: Competence, Order, Dutifulness, Achievement Striving, Self-Discipline, and Deliberation). At least one item from all proposed overlapping scales were represented in the final pool.

Emotional Instability—The preponderance of Emotional Instability items measure stress susceptibility (e.g., "I sometimes get too upset my minor setbacks"), anxiety (e.g., "I often feel nervous and stressed"), depression (e.g., "I rarely feel lonely or blue" [reverse-scored]), anger (e.g., "My anger frequently gets the better of me"), impulsivity (e.g., "Sometimes I get so upset, I feel like hurting myself"), helplessness (e.g., "I often feel that I am the victim of circumstance"), and self-consciousness (e.g., "I usually stay away from social situations"). The final item pool contained items from the Negative Temperament and Self-harm SNAP-2 scales, Harm Avoidance and Self-Directedness TCI scales, and the Neuroticism NEO PI-R scale (i.e., all facets: Anxiety, Hostility, Depression, Self-Consciousness, Impulsiveness, and Vulnerability). The Dependency SNAP-2 scale was not represented in the final item pool, which suggests that indecisiveness may not be best understood as part of the Emotional Instability construct.

Extraversion—The majority of Extraversion items measured high activity (e.g., "Most days I have a lot of 'pep' or vigor"), positive affectivity (e.g., "I laugh easily"), and sociability (e.g., "I go out of my way to meet people"). At least one item was represented from the Positive Affectivity, Exhibitionism, Entitlement, and Detachment SNAP-2 scales, the Reward Dependence, Exploratory Excitability, and Shyness TCI scales, and the Extra-version NEO PI-R scale (i.e., all facets: Warmth, Gregariousness, Assertiveness, Activity, Excitement-Seeking, and Positive Emotions). Contrary to Widiger and Simons-en's (2005) predictions, the Extravagance TCI scale was not represented in the final item pool. This finding suggests that the ease with which one spends money may not be an important aspect of the Extraversion construct.

Unconventionality—The preponderance of Unconventionality items measured intellectual curiosity (e.g., "I am intrigued by the patterns I find in art"), unusual experiences, (e.g., "Sometimes I have this strange experience in which things seem "more

real' than usual") and connectedness (e.g., "I sometimes feel so connected to nature that everything seems to be part of one living organism"). As predicted by Widiger and Simonsen (2005) items in our final pool originated from the Eccentric Perceptions SNAP-2 scale, Self-Transcendence TCI scale, and Openness NEO PI-R scale (i.e., NEO PI-R facets: Ideas, Aesthetics, and Feelings). Three Openness facets were not represented in the final item pool: Values, Fantasy, and Actions. Including Unconventionality in a dimensional model of personality pathology is somewhat controversial because it is not clear that this construct is relevant to personality pathology (O'Connor & Dyce, 1998; Watson et al., 2008; Watson, Clark, & Harkness, 1994). However, our findings are consistent with Widiger (2011b) and suggest that certain facets from the Openness NEO PI-R scale (i.e., Ideas, Aesthetics, and Feelings) can be linked with more extreme unusual experiences as measured with the Eccentric Perceptions SNAP-2 scale. Future research with this scale is required to determine its validity for psychotic features that may be important markers for particular manifestations of personality pathology (e.g., Schizotypal PD).

Additional information regarding the nature of the underlying traits can be found by examining the individual item distributions and test information curves. For Constraint, Emotional Instability, Extraversion, and Unconventionality, the frequency distributions for individual items were approximately normal and resulting test information curves were also approximately normally distributed (peaks of the distributions range from approximately −1 for Constraint to 1 for Extraversion and Unconventionality). However, the frequency distributions for individual items in our final Antagonism item bank were positively skewed despite our attempt to enrich the sample at the ceiling with clinical patients. Thus, the test information curves for the SNAP-2, TCI, NEO PI-R, and combined test were displaced to the right, with more information and precision provided in the moderate to marked ranges (approximately 0 to +2 *SD*). It might seem as though it is appropriate to identify low threshold Antagonism items; however, it is unclear if such items would measure the same construct as problematic, higher levels of Antagonism. Reis and Waller (2009) observe that peaked and (most often positively) skewed information functions for clinical scales are indicative of an underlying "quasi-trait." As a result, the construct may be less informative at the low end of the scale. Thus, low antagonism may not reflect cooperativeness/amicability but something entirely different, such as flexibility to engage in a wide range of interpersonal behavior. Reis and Waller note that quasi-traits have implications for many IRT applications. Specifically, finding items that provide information across the entire range of Antagonism could prove difficult, which poses complications for computer adaptive testing. This status also has implications for measuring change as extremely different precision for individuals exists at different ranges of Antagonism.

Nosological Implications

The best classification scheme for personality disorders (PDs) is the topic of considerable debate (Clark, 2007). Overwhelming evidence indicates that the dominant psychiatric nosology, the *DSM–IV–TR* (American Psychiatric Association, 2000), which divides personality pathology into 10 separate diagnoses, fails to align with empirical classification research (e.g., Krueger, 2005; Livesley, 2001; O'Connor, 2005), and few clinicians or researchers maintain that the *DSM–IV* PD taxonomy adequately captures the range of personality pathology (Westen, 1997; Zimmerman et al., 2005).

Dimensional classification of PDs has been proposed as an attractive alternative approach because it addresses most of the limitations of the current categorical system (Widiger & Trull, 2007). Conceptualizing an individual's personality as a multidimensional profile composed of distinct traits explains the co-occurrence among PDs as a function of shared trait liabilities, and heterogeneity within a disorder reflects differential interactions among traits (Krueger & Markon, 2006). Additionally, dimensional models preserve information

about subclinical manifestations of personality pathology that may have significant functional consequences, such as excessive alcohol use and social maladjustment (Bagge et al., 2004; Stepp, Trull, & Sher, 2005). Several dimensional models have been explicitly developed to assess a wide range of maladaptive personality traits. Morey and colleagues (2007) demonstrated the incremental validity of these approaches over the extant diagnostic system.

Our current work demonstrates the advantages of linking competing personality inventories into an integrated framework. By linking inventories from different perspectives, we can develop a comprehensive classification system that capitalizes on the strengths of different inventories. For example, because the NEO PI-R provided information at the low end of Emotional Instability and the SNAP-2 provided more information in more moderate and high ranges of Emotional Instability, integrating items from both of these models provides the most information along the personality trait continuum when compared with the subset of items from any one inventory. This finding is consistent with previous work demonstrating equivocal results when pitting one model against another (Harkness & McNulty, 1994; Morey et al., 2007; Reynolds & Clark, 2001; Stepp et al., 2005). Thus, it seems that selecting a single inventory to serve as our future taxonomy would result in a classification system that leaves out meaningful aspects of personality.

Our data analytic strategy illustrates IRT as one tool that can be used for linking personality inventories to develop an improved measurement system for personality traits. IRT models yield information about the position on the personality trait continuum where each item and inventory provides maximum psychometric information about the trait. This could enable us to develop an empirically informed measurement system that contains items that tap the low, middle, and high ranges of each trait. For example, a comprehensive inventory for Extra-version might include items that assess low (e.g., “I am a ‘people person’”), middle (e.g., “I prefer to start conversations, rather than waiting for others to talk to me”), and high (e.g., “I often feel as if I’m bursting with energy”) ranges of the trait. Future research can develop cut-points along the trait continuum to aid in clinical decision-making.

Limitations

We were only able to link a small subset of the 18 competing dimensional personality inventories in the current study. However, given that we have now integrated the SNAP-2, TCI, and NEO PI-R, we will be able to forge ahead. As participants in future studies complete at least one of the measures we have already concurrently calibrated in addition to another personality inventory (e.g., MPQ, DAPP-BQ), we will be able to link the additional personality inventory to the current scales. As more inventories are linked and we are able to directly compare each inventory, we will learn about the best path for a future integrative measurement tool and refine our understanding of the underlying traits.

Although this linking approach based on concurrent IRT calibration can be used to create a new integrated inventory, we did not intend to create a new inventory based on these three commercially available inventories. Rather, we intended to provide researchers information on the item performance form each of these three inventories with advantages and disadvantages of each inventories.

Potential biases can sometimes be introduced by combining samples (cf. Waller, 2008). We wanted to bolster our student sample with psychiatric patient samples to expand the potential range of scores that would be endorsed. For IRT purposes, we felt that the potential increase in range of scores by also using patients outweighed the concerns of commingling the samples.

Future Directions for Personality Assessment

Our next set of objectives includes validating the integrated item banks by demonstrating their utility in predicting social functioning and treatment response compared with already existing measures of personality, including the extant classification system for personality disorders. One of the advantages to our approach is that we can further refine these item pools. The predictive utility of these refined item banks should also be tested against existing measures.

In summary, we encourage researchers to continue to investigate the utility of integrative personality inventories. We believe this approach provides the most information along the entire personality trait continuum and will yield the most comprehensive, flexible, and precise inventory. This approach brings together theories and inventories that are distinct but contain significant overlap. For this reason we feel that an integrative dimensional personality inventory will generate novel empirical studies and refine our understanding of underlying personality traits.

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: American Psychiatric Association; 2000. revised

Ando J, Ono Y, Yoshimura K, Onoda N, Shinohara M, Kanba S, Asai M. The genetic structure of Cloninger's seven-factor model of temperament and character in a Japanese sample. Journal of Personality. 2002; 70:583–609.10.1111/1467-6494.05018 [PubMed: 12322854]

Bagge C, Nickell A, Stepp S, Durrett C, Jackson K, Trull TJ. Borderline personality disorder features predict negative outcomes 2 years later. Journal of Abnormal Psychology. 2004; 113:279–288.10.1037/0021-843X.113.2.279 [PubMed: 15122948]

Buysse DJ, Yu L, Moul DE, Germain A, Stover A, Dodds NE, Pilkonis PA. Development and validation of Patient-Reported Outcome Measures for sleep disturbance and sleep-related impairments. Sleep. 2010; 33:781–792. [PubMed: 20550019]

Campbell BF, Sengupta S, Santos C, Lorig KR. Balanced incomplete block design: Description, case study and implications for practice. Health Education Quarterly. 1995; 22:201–210. [PubMed: 7622388]

Clark LA. Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. Annual Review of Psychology. 2007; 58:227–257.10.1146/annurev-.psych.57.102904.190200

Clark, LA.; Livesley, WJ. Two approaches to identifying the dimensions of personality disorder: Convergence on the five-factor model. In: Costa, PT.; Widiger, TA., editors. Personality disorders and the five-factor model of personality. 2. Washington, DC: American Psychological Association; 2002. p. 161-176.

Clark, LA.; Simms, LJ.; Wu, KD.; Casillas, A. Schedule for Nonadaptive and Adaptive Personality: Manual for administration, scoring, and interpretation. 2. Minneapolis, MN: University of Minnesota Press; in press

Cleckley, H. The mask of sanity. 4. St. Louis, MO: Mosby; 1964.

Cloninger CR. A unified biosocial theory of personality and its role in the development of anxiety states. Psychiatric Developments. 1986; 4:167–226. [PubMed: 3809156]

Cloninger CR. A systematic method for clinical description and classification of personality variants: A proposal. Archives of General Psychiatry. 1987; 44:573–588. [PubMed: 3579504]

Cloninger, CR.; Przybeck, TR.; Svrakic, DM. The Temperament and Character Inventory (TCI): A guide to its development and use. St. Louis, MO: Centre for Psychobiology of Personality; 1994.

Cloninger CR, Svrakic DM, Przybeck TR. A psychobiological model of temperament and character. Archives of General Psychiatry. 1993; 50:975–990. [PubMed: 8250684]

Cochran, WG.; Cox, GM. Experimental designs. New York, NY: Wiley; 1957.

Cook KF, Kallen MA, Amtmann D. Having a fit: Fit statistics for simulated and actual data fit statistics for IRT models. Quality of Life Research. 2009; 18:447–460.10.1007/s11136-009-9464-4 [PubMed: 19294529]

Costa, PT.; McCrae, RR. The NEO Personality Inventory manual. Odessa, FL: Psychological Assessment Resources; 1985.

Costa, PT.; McCrae, RR. NEO PI-R professional manual. Odessa, FL: Psychological Assessment Resources, Inc; 1992a.

Costa PT, McCrae RR. 'Normal' personality inventories in clinical assessment: General requirements and the potential for using the NEO Personality Inventory": Reply. Psychological Assessment. 1992b; 4:20–22.10.1037/1040-3590.4.1.20

Digman JM. Higher-order factors of the Big Five. Journal of Personality and Social Psychology. 1997; 73:1246–1256.10.1037/0022-3514.73.6.1246 [PubMed: 9418278]

Embretson, SE.; Reise, SP. Item Response Theory for psychologists. Mahwah, NJ: Erlbaum; 2000.

Eysenck, HJ.; Eysenck, SBG. Psychoticism as a dimension of personality. New York, NY: Crane, Russak, & Company; 1976.

Fassino S, Abbate-Daga G, Amianto F, Leombruni P, Boggio S, Rovera GG. Temperament and character profile of eating disorders: A controlled study with the Temperament and Character Inventory. The International Journal of Eating Disorders. 2002; 32:412–425.10.1002/eat.10099 [PubMed: 12386906]

Harkness, AR.; McNulty, JL. The Personality Psychopathology Five (PSY-5): Issues from the pp. of a diagnostic manual instead of a dictionary. In: Strack, S.; Lorr, M., editors. Differentiating normal and abnormal personality. 1. New York, NY: Springer; 1994. p. 291-315.

Harrison DA. Robustness of IRT parameter estimation to violations of the unidimensionality assumption. Journal of Educational Statistics. 1986; 11:91–115.10.2307/1164972

Hays RD, Liu H, Spritzer K, Cella D. Item response theory analyses of physical functioning items in the medical outcomes study. Medical Care. 2007; 45:S32–S38.10.1097/01.mlr.0000246649.43232.82 [PubMed: 17443117]

Joyce PR, Mulder RT, McKenzie JM, Luty SE, Cloninger CR. Atypical depression, atypical temperament and a differential anti-depressant response to fluoxetine and nortriptyline. Depression and Anxiety. 2004; 19:180–186.10.1002/da.20001 [PubMed: 15129420]

Kirisci L, Hsu T, Yu L. Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. Applied Psychological Measurement. 2001; 25:146–162.10.1177/01466210122031975

Krueger RF. Continuity of axes I and II: Toward a unified model of personality, personality disorders, and clinical disorders. Journal of Personality Disorders. 2005; 19:233–261. 10.1521/ pedi.2005.19.3.233. [PubMed: 16175735]

Krueger RF, Markon KE. Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. Annual Review of Clinical Psychology. 2006; 2:111–133.10.1146/annurev.clinpsy.2.022305.095213

Livesley, WJ. Conceptual and taxonomic issues. In: Livesley, WJ., editor. Handbook of personality disorders. New York, NY: Guilford Press; 2001. p. 3-38.

Livesley, WJ.; Jackson, DN. Dimensional Assessment of Personality Pathology-Basic Questionnaire. Port Huron, MI: Research Psychologists Press; 2009.

Lord, FM. Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum; 1980.

Markon KE, Krueger RF, Watson D. Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. Journal of Personality and Social Psychology. 2005; 88:139–157.10.1037/0022-3514.88.1.139 [PubMed: 15631580]

McDonald RP, Ho MH. Principles and practice in reporting structural equation analyses. Psychological Methods. 2002; 7:64–82. 10.1037/ 1082-989X.7.1.64. [PubMed: 11928891]

McHorney CA, Cohen AS. Equating health status measures with item response theory: Illustration with functional status items. Medical Care. 2000; 38:11–45.10.1097/00005650-200009002-00008

Morey LC, Hopwood CJ, Gunderson JG, Skodol AE, Shea MT, Yen S, McGlashan TH. Comparison of alternative models for personality disorders. Psychological Medicine. 2007; 37:983–994. 10.1017/ S0033291706009482. [PubMed: 17121690]

Morey LC, Warner MB, Shea MT, Gunderson JG, Sanislow CA, Grilo C, McGlashan TH. The representation of four personality disorders by the Schedule for Nonadaptive and Adaptive Personality dimensional model of personality. Psychological Assessment. 2003; 15:326–332. 10.1037/1040-3590 .15.3.326. [PubMed: 14593832]

Muraki E. Information functions of the generalized partial credit model. Applied Psychological Measurement. 1993; 17:351–363. 10.1177/ 014662169301700403.

Muthén, B.; du Toit, SHC.; Spisic, D. Unpublished Technical Report. 1997. Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes.

O'Connor BP. 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. 2002; 83:962–982.10.1037/0022-3514.83.4.962 [PubMed: 12374447]

O'Connor BP. A search for consensus on the dimensional structure of personality disorders. Journal of Clinical Psychology. 2005; 61:323–345.10.1002/jclp.20017 [PubMed: 15468325]

O'Connor BP, Dyce JA. A test of models of personality disorder configuration. Journal of Abnormal Psychology. 1998; 107:3–16.10.1037/0021-843X.107.1.3 [PubMed: 9505034]

Reckase, MD. Multidimensional item response theory. New York, NY: Srpinger; 2009.

Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Cella D. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the patient-reported outcomes measurement information system (PROMIS). Medical Care. 2007; 45:s22–s31.10.1097/01.mlr. 0000250483.85507.04 [PubMed: 17443115]

Reise SP, Henson JM. A discussion of modern versus traditional psychometrics as applied to personality assessment scales. Journal of Personality Assessment. 2003; 81:93–103. 10.1207/ S15327752JPA8102_01. [PubMed: 12946916]

Reise SP, Waller NG. Item response theory and clinical measurement. Annual Review of Clinical Psychology. 2009; 5:25–46.

Revicki DA, Chen WH, Harnam N, Cook K, Amtmann D, Callahan LF, Keefe FJ. Development and psychometric analysis of the PROMIS pain behavior item bank. Pain. 2009; 146:158–169.10.1016/j.pain.2009.07.029 [PubMed: 19683873]

Reynolds SK, Clark LA. Predicting dimensions of personality disorder from domains and facets of the Five-Factor Model. Journal of Personality. 2001; 69:199–222.10.1111/1467-6494.00142 [PubMed: 11339796]

Rock DA, Nelson J. Applications and extensions of NAEP concepts and technology. Journal of Educational Statistics. 1992; 17:219–232.10.2307/1165171

Samuel DB, Simms LJ, Clark LE, Livesley WJ, Widiger TA. An item response theory integration of normal and abnormal personality scales. Personality Disorders: Theory, Research, and Treatment. 2010; 1:5–21. 10.1037/ a0018136.

Savla J, Davey A, Costa PT, Whitfield KE. Replicating the NEO PI-R factor structure in African-American older adults. Personality and Individual Differences. 2007; 43:1279–1288.10.1016/ j.paid.2007.03.019

Stepp SD, Trull TJ, Burr RM, Wolfenstein M, Vieth AZ. Incremental validity of the Structured Interview for the Five-Factor Model of personality (SIFFM). European Journal of Personality. 2005; 19:343–357.10.1002/per.565

Stepp SD, Trull TJ, Sher KJ. Borderline personality features predict alcohol use problems. Journal of Personality Disorders. 2005; 19:711–722.10.1521/pedi.2005.19.6.711 [PubMed: 16553564]

Tate R. A comparison of selected empirical methods for assessing the structure of responses to test items. Applied Psychological Measurement. 2003; 27:159–203. 10.1177/ 0146621603027003001.

Tellegen, A. Manual of the Multidimensional Personality Questionnaire. Minneapolis, MN: University of Minnesota Press; 2000.

Thissen, D. MULTILOG 7: Multiple categorical item analysis and test scoring using item response theory [computer program]. Chicago, IL: Scientific Software; 2003.

Thissen, D.; Wainer, H., editors. Test scoring. Mahwah, NJ: Erlbaum; 2001.

Van de Linden WJ, Veldkamp BP, Carlson JE. Optimizing balanced incomplete block designs for educational assessments. Applied Psychological Measurement. 2004; 28:317–331.10.1177/0146621604264870

Waller NG. Commingled Samples: A Neglected Source of Bias in Reliability Analysis. Applied Psychological Measurement. 2008; 32:211–223.10.1177/0146621607300860

Walton KE, Roberts BW, Krueger RF, Blonigen DM, Hicks BM. Capturing abnormal personality with normal personality inventories: An item response theory approach. Journal of Personality. 2008; 76:1623–1647.10.1111/j.1467-6494.2008.00533.x [PubMed: 19012660]

Watson D, Clark LA, Chmielewski M. Structures of personality and their relevance to psychopathology: II. Further articulation of a comprehensive unified trait structure. Journal of Personality. 2008; 76:1545–1585.10.1111/j.1467-6494.2008.00531.x [PubMed: 19012658]

Watson D, Clark LA, Harkness AR. Structures of personality and their relevance to psychopathology. Journal of Abnormal Psychology. 1994; 103:18–31. 10.1037/0021-843X .103.1.18. [PubMed: 8040477]

Westen D. Divergences between clinical and research methods for assessing personality disorders: Implications for research and the evolution of Axis II. American Journal of Psychiatry. 1997; 154:895–903. [PubMed: 9210738]

Widiger TA. Integrating normal and abnormal personality structure: A proposal for DSM-V. Journal of Personality Disorders. 2011a; 25:338–363.10.1521/pedi.2011.25.3.338 [PubMed: 21699396]

Widiger TA. The DSM-5 dimensional model of personality disorder: Rationale and empirical support. Journal of Personality Disorders. 2011b; 25:224–234.10.1521/pedi.2011.25.2.222

Widiger TA, Livesley WJ, Clark LEA. An integrative dimensional classification of personality disorder. Psychological Assessment. 2009; 21:243–255.10.1037/a0016606 [PubMed: 19719338]

Widiger TA, Simonsen E. Alternative dimensional models of personality disorder: Finding a common ground. Journal of Personality Disorders. 2005; 19:110–130. 10.1521/pedi.19.2 .110.62628. [PubMed: 15899712]

Widiger TA, Trull TJ. Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. American Psychologist. 2007; 62:71–83.10.1037/0003-066X.62.2.71 [PubMed: 17324033]

Wu K, Lindsted KD, Tsai S, Lee JW. Chinese NEO PI-R in Taiwanese adolescents. Personality and Individual Differences. 2008; 44:656–667.10.1016/j.paid.2007.09.025

Yamagata S, Suzuki A, Ando J, Ono Y, Kijima N, Yoshimura K, Jang KL. Is the genetic structure of human personality universal? A cross-cultural twin study from North America, Europe, and Asia. Journal of Personality and Social Psychology. 2006; 90:987–998.10.1037/0022-3514.90.6.987 [PubMed: 16784347]

Yamamoto K, Mazzeo J. Item response theory scale linking in NAEP. Journal of Educational Statistics. 1992; 17:155–173. 10.2307/ 1165167.

Zimmerman M, Rothschild L, Chelminski I. The prevalence of *DSM–IV* personality disorders in psychiatric outpatients. American Journal of Psychiatry. 2005; 162:1911–1918.10.1176/appi.ajp.162.10.1911 [PubMed: 16199838]

Zwick WR, Velicer WF. Comparison of five rules for determining the number of components to retain. Psychological Bulletin. 1986; 99:432–442.10.1037/0033-2909.99.3.432

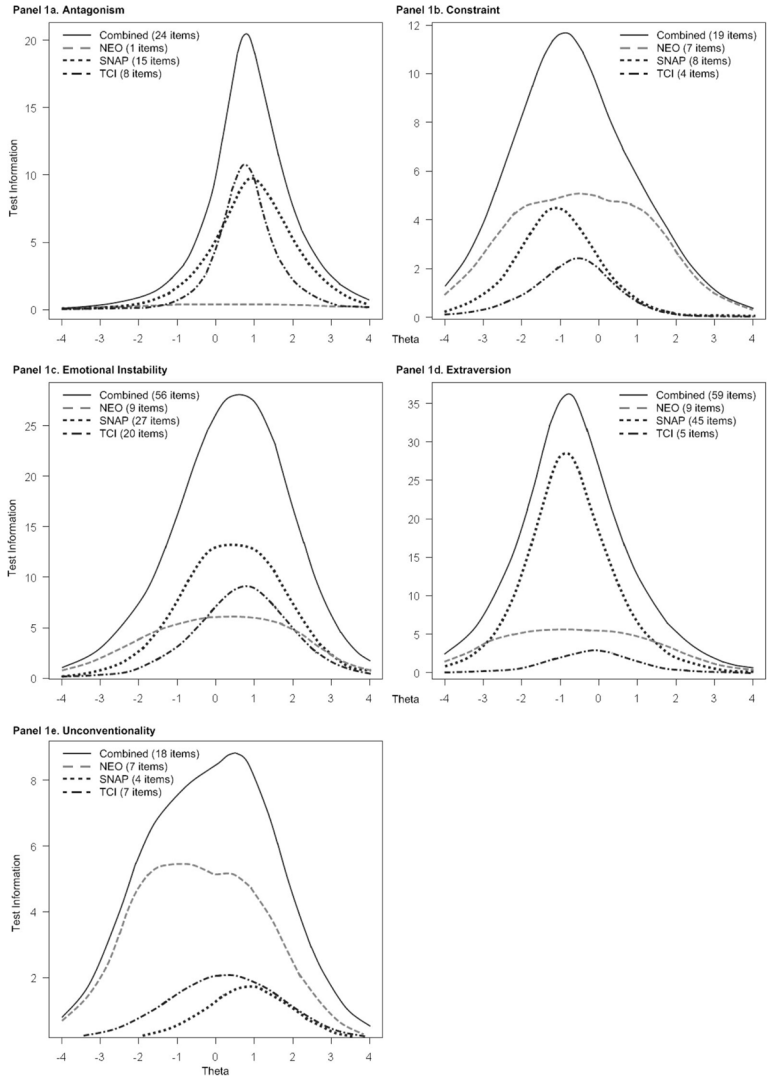

Figure 1.

Test information curves for the SNAP-2, TCI, NEO PI-R, and combined test for Antagonism (Panel 1a), Constraint (Panel 1b), Emotional Instability (Panel 1c), Extraversion (Panel 1d), and Unconventionality (Panel 1e).

Personal Disord. Author manuscript; available in PMC 2013 October 10.

Table 1

NEO, SNAP, and TCI Scales to Create Item Pools for Antagonism, Constraint, Emotional Instability, Extraversion, and Unconventionality Personality Dimensions

	Scales and initial item pools			Items retained following psychometric analyses (Steps 1[b], 2[c], and 3[d])
	SNAP	**TCI**	**NEO**	
Antagonism • Initial item pool (*N* = 214) • Post-factor analysis item pool (*N* = 92) • Final item pool (*N* = 24)	Mistrust (19)[a] Manipulativeness (20) Aggressiveness (20) Entitlement (16) Dependency *(18)	Cooperativeness *(40) Reward dependence *(33)	Agreeableness *(48)	Mistrust (1 = 9, 2 = 9, 3 = 0) Manipulativeness (1 = 13, 2 = 13, 3 = 1) Aggressiveness (1 = 19, 2 = 19, 3 = 14) Cooperativeness (1 = 23, 2 = 23, 3 = 8) Reward dependence (1 = 6, 2 = 6, 3 = 0) Agreeableness (1 = 22, 2 = 18, 3 = 1)
Emotional instability • Initial item pool (*N* = 190) • Post-factor analysis item pool (*N* = 114) • Final item pool (*N* = 56)	Negative temperament (28) Self-harm (16) Dependency (18)	Harm avoidance (36) Self-directedness *(44)	Neuroticism (48)	Negative temperament (1 = 26, 2 = 26, 3 = 19) Self-harm (1 = 13, 2 = 13, 3 = 8) Dependency (1 = 4, 2 = 4. 3 = 0) Harm avoidance (1 = 29, 2 = 29, 3 = 10) Self-directedness (1 = 23, 2 = 23, 3 = 10) Neuroticism (1 = 29, 2 = 25, 3 = 9)
Extraversion • Initial item pool (*N* = 186) • Postfactor analysis item pool (*N* = 91) • Final item pool (*N* = 59)	Positive affectivity (27) Exhibitionism (16) Entitlement (16) Detachment *(18)	Reward dependence (33) Extravagance (9) Exploratory excitability (11) Shyness *(8) Attachment (9)	Extraversion (48)	Positive affectivity (1 = 22, 2 = 22, 3 = 18) Exhibitionism (1 = 13, 2 = 13, 3 = 10) Entitlement (1 = 6, 2 = 6. 3 = 2) Detachment (1 = 17, 2 = 17. 3 = 15) Reward dependence (1 = 1, 2 = 1, 3 = 1) Extravagance (1 = 0, 2 = 0, 3 = 0) Exploratory excitability (1 = 1, 2 = 1, 3 = 1) Shyness (1 = 2, 2 = 2, 3 = 2) Attachment (1 = 2, 2 = 2, 3 = 1) Extraversion (1 = 28, 2 = 14, 3 = 9)
Constraint • Initial item pool (*N* = 232) • Postfactor analysis item pool (*N* = 108) • Final item pool (*N* = 19)	Disinhibition *(35) Workaholism (18) Propriety (20) Impulsivity *(19)	Self-directedness (44) Novelty seeking *(40) Persistence (8)	Conscientiousness (48)	Disinhibition (1 = 23, 2 = 23. 3 = 4) Workaholism (1 = 8, 2 = 8, 3 = 2) Propriety (1 = 4, 2 = 4, 3 = 1) Impulsivity (1 = 9, 2 = 9, 3 = 1) Self-directedness (1 = 11, 2 = 11, 3 = 0) Novelty seeking (1 = 11, 2 = 11, 3 = 1) Persistence (1 = 6, 2 = 6, 3 = 3) Conscientiousness (1 = 36, 2 = 23, 3 = 7)
Unconventionality • Initial item pool (*N* = 96) • Postfactor analysis item pool (*N* = 49)	Eccentric perceptions (15)	Self-transcendence (33)	Openness (48)	Eccentric perceptions (1 = 10, 2 = 10, 3 = 4) Self-transcendence (1 = 16, 2 = 16, 3 = 7) Openness (1 = 23, 2 = 22, 3 = 7)

Scales and initial item pools			Items retained following psychometric analyses (Steps 1[b], 2[c], and 3[d])
SNAP	TCI	NEO	
• Final item pool (N = 18)			

[a] Total number of items.

[b] Number of items retained following combined factor analysis.

[c] Number of items retained following removal of items with response frequencies less than 5% (only impacted NEO scales).

[d] Number of items retained following IRT calibration.

[*] Scale was reverse-scored so higher scores indicated higher levels of personality dimension.

Table 2

Concurrent GRM and 2PLM Item Parameter Estimates and SEs for the 19-Item Constraint Scale From the NEO, SNAP, and TCI

Item	Original scale	Stem	*a*	*b1*	*b2*	*b3*	*b4*
T103	Persistence	I usually push myself harder than most…	1.99 (0.22)	−0.58 (0.08)			
N130	Conscientiousness	I never seem to be able to get organized (R)	1.96 (0.16)	−1.71 (0.14)	−0.67 (0.08)	−0.16 (0.07)	0.89 (0.10)
S29	Workaholism	When I start a task, I am determined to finish…	1.94 (0.23)	−0.96 (0.10)			
N40	Conscientiousness	I keep my belongings neat and clean	1.90 (0.15)	−2.17 (0.19)	−0.80 (0.09)	−0.16 (0.08)	0.97 (0.10)
S89	Impulsivity	I tend to value and follow a rational, …	1.85 (0.26)	−1.36 (0.13)			
T148	Novelty seeking	I like to pay close attention to details in…	1.73 (0.22)	−0.47 (0.09)			
N100	Conscientiousness	I like to keep everything in its place so I…	1.60 (0.14)	−2.30 (0.22)	−0.98 (0.11)	−0.15 (0.08)	1.16 (0.12)
S214	Workaholism	I enjoy working hard	1.57 (0.20)	−1.28 (0.13)			
N25	Conscientiousness	I'm pretty good about pacing myself so as to…	1.55 (0.13)	−2.21 (0.20)	−0.83 (0.11)	−0.27 (0.09)	1.43 (0.14)
S173	Disinhibition	I usually use careful reasoning when…	1.50 (0.20)	−1.08 (0.13)			
S154	Disinhibition	I always try to be fully prepared before I…	1.45 (0.18)	−1.04 (0.13)			
T62	Persistence	I am more hard-working than most people	1.37 (0.20)	−0.36 (0.10)			
S261	Disinhibition	I work just hard enough to get by (R)	1.34 (0.16)	−0.72 (0.12)			
S202	Propriety	When I'm working on something, I'm not…	1.22 (0.33)	−1.05 (0.16)			
N210	Conscientiousness	I plan ahead carefully when I go on a trip	1.22 (0.14)	−2.81 (0.37)	−1.37 (0.19)	−0.49 (0.13)	1.11 (0.17)
N205	Conscientiousness	There are so many little jobs that need… (R)	1.22 (0.15)	−2.63 (0.37)	−1.10 (0.18)	−0.28 (0.14)	1.45 (0.21)
T166	Persistence	I often give up a job if it takes much… (R)	1.21 (0.21)	−1.64 (0.24)			
S254	Disinhibition	Before making a decision, I carefully…	1.10 (0.17)	−1.13 (0.17)			
N230	Conscientiousness	I'm something of a "workaholic"	1.09 (0.13)	−1.79 (0.26)	−0.05 (0.14)	0.90 (0.16)	2.49 (0.32)

Note. Letters before the item number indicate the measure to which the item originated (N = NEO-PI-R, S = SNAP-2, and T = TCI). Some items had to be shortened (…) because of space limitations. "R" indicates the item was reverse scored. The "*a*" parameter represents slope and "*b*" parameter(s) represents location.

Personal Disord. Author manuscript; available in PMC 2013 October 10.