

Evaluating the Criterion Validity of Hierarchical Psychopathology Dimensions Across Models: Familial Aggregation and Associations with Research Domain Criteria (RDoC) (Sub)constructs

Carter J. Funkhouser
University of Illinois at Chicago

Eugene M. Cozza
Northwestern University

Kelly A. Correa
University of Illinois at Chicago

Ryne Estabrook
University of Illinois at Chicago

Allison M. Letkiewicz
Northwestern University

Stewart A. Shankman
Northwestern University

The Hierarchical Taxonomy of Psychopathology (HiTOP) posits that psychopathology is a hierarchy of correlated dimensions. Numerous studies have examined the validity of these dimensions using bifactor models, in which each disorder loads onto both a general and specific factor (e.g., internalizing, externalizing). Although bifactor models tend to fit better than alternative models, concerns have been raised about bifactor model selection, factor reliability, and interpretability. Therefore, we compared the reliability and validity of several higher-order HiTOP dimensions between bifactor and correlated factor models using familial aggregation and associations with Research Domain Criteria (RDoC) (sub)constructs as validators. Lifetime psychopathology was assessed in a community sample ($N=504$) using dimensional disorder severity scales calculated from semi-structured interview data. A series of unidimensional, correlated factor, and bifactor models were fit to model several HiTOP dimensions. A bifactor model with two specific factors (internalizing and disinhibited externalizing) and a correlated two-factor model provided the best fit to the data. HiTOP dimensions had adequate reliability in the correlated factor model, but suboptimal reliability in the bifactor model. The disinhibited externalizing dimension was highly correlated across the two models and was familial, yet largely unrelated to RDoC (sub)constructs in both models. The internalizing dimension in the correlated factor model and the general factor in the bifactor model were highly correlated and had similar validity patterns, suggesting the general factor was largely redundant with the internalizing dimension in the correlated factor model. These findings support concerns about the interpretability of psychopathology dimensions in bifactor models.

This manuscript has been accepted for publication in *Journal of Abnormal Psychology*.

It will undergo copyediting, typesetting, and review of the resulting proof before being published in its final form.

The Hierarchical Taxonomy of Psychopathology (HiTOP) is a structural organization of psychopathology based on findings from quantitative studies of psychopathology structure (Kotov et al., 2017). In contrast to traditional diagnostic

classification systems, HiTOP conceptualizes psychopathological phenotypes as dimensional in nature and organizes phenotypes into a hierarchy of five levels based on observed patterns of covariation. First, individual symptoms and signs are combined into homogeneous symptom components (e.g., insomnia) and maladaptive traits (e.g., emotional lability). In turn, symptom components and maladaptive traits are grouped into syndromes, which correspond to diagnostic constructs in the Diagnostic and Statistical Manual (DSM). Syndromes are combined into subfactors, which capture the shared variance among constituent DSM diagnoses. For example, depression, generalized anxiety, and posttraumatic stress are combined into a distress subfactor, and circumscribed anxiety (social anxiety,

Preprint template created by Brenton Wiernik (<https://osf.io/hsv6a/>).

Ethical approval for this study was obtained from the UIC Office for the Protection of Research Subjects (protocol 2012-0646). This work was supported by National Institute of Mental Health Grants R01 MH098093 (PI: Shankman), F31 MH123042 (PI: Funkhouser), and F31 MH121003 (PI: Correa).

Correspondence concerning this article should be addressed to Carter J. Funkhouser, Department of Psychology, University of Illinois at Chicago, Chicago, IL, 60607. Email: carterfunkhouser@gmail.com

specific phobia, agoraphobia, panic disorder) and Obsessive Compulsive Disorder make up a fear subfactor. Subfactors in turn form spectra (e.g., the distress and fear subfactors form a higher-order internalizing factor). Finally, spectra can be aggregated into a superspectrum representing a general factor of psychopathology, or p-factor (Caspi et al., 2014). While HiTOP is a promising framework for understanding the ontology, causes, and consequences of psychopathology, it is critical to test the validity of HiTOP constructs and compare validity across different statistical models that are commonly used to estimate HiTOP dimensions.

Several statistical models have been used to model the HiTOP structure. Initial psychopathology structural studies used a correlated factor model, in which two or more distinct but correlated factors (e.g., internalizing, externalizing) summarize the shared variance among their indicators. Correlated factor models have consistently identified either two (internalizing, externalizing) or three (distress, fear, externalizing) transdiagnostic factors (Achenbach, 1966; Krueger & Markon, 2006; Watson, 2005). A separate thought disorder spectrum has also emerged in some studies that modeled measures of psychosis, mania, and/or certain personality disorders (e.g., Caspi et al., 2014; Kotov et al., 2011). More recently, HiTOP dimensions have been predominantly modeled using a bifactor model, which extracts a general factor that is orthogonal to narrower specific factors (e.g., spectra). Studies comparing bifactor and correlated factor models have typically adjudicated these models using fit indices and found that the bifactor model fit better (Watts et al., 2020). However, simulation studies demonstrate that the superior fit of the bifactor model may be due to overfitting (e.g., Greene et al., 2019). The bifactor model of psychopathology has also been critiqued for yielding factors that are unreliable and difficult to interpret due to weak and inconsistent factor loadings (Levin-Aspenson et al., 2020; Watts et al., 2019) and the orthogonality of the general and specific factors (Bonifay et al., 2017). These concerns, in conjunction with recent findings that the validity of HiTOP dimensions can differ substantially across models (Conway, Mansolf, et al., 2019; Hyland et al., 2020; Moore et al., 2020), have prompted recommendations that comparisons between bifactor models and alternative models consider external validity in addition to model fit (e.g., Bonifay et al., 2017; Bornovalova et al., 2020; Forbes et al., 2020; Watts et al., 2019).

Examining validity of HiTOP dimensions through tests of familial aggregation, clinical and laboratory correlates, specificity (i.e., discriminant validity), and longitudinal course is crucial for advancing this novel approach (Feighner et al., 1972; Robins & Guze, 1970). Twin and family studies suggest that the p-factor and internalizing and externalizing spectra are weakly to moderately familial (Lahey et al., 2011; Martel et al., 2017; Michelini et al., 2019; Pesenti-Gritti et al., 2008; Waldman et al., 2016). However, these investigations have primarily studied children and adolescents, and the extent to which familial aggregation estimates in youth generalize to older populations is unclear. Additionally, prior studies of the familial aggregation of HiTOP dimensions have not compared familial aggregation across different structural psychopathology models (e.g., Lahey et al., 2011; Martel et al., 2017; Waldman et al., 2016), highlighting the novel contribution of the present report.

In addition to familial aggregation, studies have linked HiTOP dimensions to a wide range of theoretically relevant clinical and laboratory correlates (see Conway, Forbes, et al., 2019; Kotov et al., 2017 for reviews). As described in a recent review (Michelini et al., 2020), some of these studies have examined relationships between HiTOP dimensions and validators from NIMH's Research Domain Criteria (RDoC) matrix. Like HiTOP, RDoC was developed to address limitations of traditional categorical nosologies, and it aims to identify biobehavioral processes underlying psychopathological phenotypes (Insel et al., 2010; Zalta & Shankman, 2016). These biobehavioral processes are organized into six domains (Positive Valence Systems [PVS], Negative Valence Systems [NVS], Cognitive Systems, Social Processes, Arousal and Regulatory Systems, and Sensorimotor Systems), each of which contains constructs and constituent subconstructs that can be studied at various units of analysis such as genes, circuits, physiology, behavior, and self-report. RDoC's conceptualization of psychiatric phenotypes is explicitly agnostic to DSM diagnoses, and, like HiTOP, it conceptualizes psychopathology as dimensional rather than categorical. Unlike HiTOP, which is a descriptive model of clinical symptoms and is agnostic toward etiology (Kotov et al., 2017), RDoC specifies a set of constructs thought to be involved in the etiology of psychopathology. Although RDoC has had a substantial impact on psychopathology research, it has been criticized for providing insufficient guidance on

how clinical problems should be conceptualized and measured (Patrick & Hajcak, 2016). Given the complementarity of the two frameworks, studies bridging HiTOP dimensions and RDoC (sub)constructs are needed (Conway, Forbes, et al., 2019; Michelini et al., 2020).

The present study aimed to (a) examine the criterion and discriminant validity of HiTOP dimensions using familial aggregation and RDoC (sub)constructs as validators, and (b) compare criterion validity across bifactor and correlated factor models. Due to sample exclusion criteria and the psychopathology data available, we focused our analyses on the internalizing and disinhibited externalizing spectra, the distress and fear subfactors, and the p-factor. Additionally, as there are many RDoC (sub)constructs, we focused on certain (sub)constructs that have been consistently linked to at least one of the HiTOP dimensions under examination and for which data were available. Specifically, we examined Reward Anticipation and Initial Response to Reward (subconstructs within the Reward Responsiveness construct in the PVS), the Potential Threat construct in the NVS, and two subconstructs within the Cognitive Control construct in the Cognitive Systems domain (switching and inhibition).

We hypothesized that both correlated factor models and bifactor models would fit the data adequately, with the bifactor models having superior fit. We also expected that analyses of familial aggregation would replicate results from several prior studies that reported moderate familial aggregations of HiTOP dimensions derived from bifactor models (Martel et al., 2017; Waldman et al., 2016) or exploratory factor analysis (Michelini et al., 2019). Based on a recent review of studies examining connections between RDoC (sub)constructs and HiTOP dimensions (Michelini et al., 2020) and other recent relevant studies (Conway, Li, et al., 2019; Correa et al., 2019; Crane et al., 2021; Radoman et al., 2019; Stevens et al., 2019), we hypothesized that (a) disinhibited externalizing would be positively related to Reward Anticipation, (b) disinhibited externalizing and the distress subfactor would be negatively associated with Initial Response to Reward, (c) internalizing and the fear subfactor would be positively associated with Potential Threat, (d) the disinhibited externalizing spectrum would be negatively related to Switching and Inhibition, and (e) internalizing spectra and the fear subfactor would be negatively associated with Inhibition. Based on several studies that compared the validity of bifactor

and correlated factor psychopathology models using other validators and in other populations (Conway, Mansolf, et al., 2019; Forbes et al., 2020; Watts et al., 2019), we hypothesized that the best-fitting bifactor model may have a less consistent and interpretable validity pattern than the correlated factor model.

Methods

Participants

Young adults ($N=504$) were recruited from the local community via fliers, Internet postings, and area mental health clinics as part of a family study funded under the RDoC initiative (Correa et al., 2019; Funkhouser et al., 2019; Stevens et al., 2019). Consistent with recommendations for both HiTOP (Stanton et al., 2020) and RDoC (Cuthbert & Insel, 2013), the study had no diagnostic-based inclusion criteria, but oversampled for internalizing symptoms and substance use disorders to increase clinical relevance (e.g., specific and separate advertising for individuals with depression, anxiety, substance use, etc.). Participants were required to be between the ages of 18 and 30 years to ensure they were still in the peak risk window for psychopathology (Kessler et al., 2005) and to have at least one full biological sibling within the same age range who was eligible to enroll. Although most participants had a sibling who also participated, 30 participants did not. Participants were excluded from the study if they had a personal or family history of psychosis or mania; were unable to read or write English; had a history of head trauma with loss of consciousness; or were left-handed. These exclusion criteria served to ensure participants were able to provide consent and to mitigate confounds to psychophysiological data (reported elsewhere). Sample demographics and clinical characteristics are presented in Table 1.

Table 1. Demographic and clinical characteristics.

Characteristic	No. (%)
Age (<i>M, SD</i>)	22.40 (3.18)
Sex (Female)	321 (63.7)
Ethnicity	
Caucasian	214 (42.5)
Hispanic	106 (21.0)
African American	79 (15.7)
Asian	59 (11.7)
Middle Eastern	15 (3.0)
Mixed Race	29 (5.8)
Other	2 (0.4)
Education	
Some high school	16 (3.2)
Graduated high school or equivalent	27 (5.4)
Some college	254 (50.5)
Graduated 2-year college	26 (5.2)
Graduated 4-year college	180 (35.8)
Employment	
Full-time	117 (23.3)
Part-time	114 (22.7)
Student	245 (48.7)
Unemployed	27 (5.4)
Lifetime Diagnoses	
Major Depressive Disorder	182 (36.1)
Generalized Anxiety Disorder	54 (10.7)
Posttraumatic Stress Disorder	36 (7.1)
Social Anxiety Disorder	101 (20.0)
Specific Phobia	105 (20.8)
Panic Disorder	42 (8.3)
Obsessive Compulsive Disorder	33 (6.6)
Alcohol Use Disorder	149 (29.6)
Cannabis Use Disorder	101 (20.0)
Stimulant Use Disorder	20 (4.0)

Psychopathology Assessment

Lifetime psychopathologies were assessed by interviewers using the research version of the Structured Clinical Interview for DSM-5 (SCID; First et al., 2015). The following disorders were assessed: major depressive disorder (MDD), generalized anxiety disorder (GAD), posttraumatic stress disorder (PTSD), panic disorder (PD), social anxiety disorder (SAD), specific phobia (SP), agoraphobia, obsessive-compulsive disorder (OCD), anorexia nervosa, bulim-

ia nervosa, binge eating disorder, and substance use disorders for all of the DSM-5 substance classes except caffeine and tobacco. However, agoraphobia, eating disorders, and all substance use disorders except for alcohol use disorder (ALC), cannabis use disorder (CANN), and stimulant use disorder (STIM) were excluded from analyses due to low frequency (<20 cases meeting lifetime diagnostic criteria).

The diagnostic interview used in the present study was modified in several ways to improve its utility for modeling psychopathology structure. First, skip-outs for MDD and GAD were ignored, allowing for the assessment of each GAD and MDD symptom regardless of the presence of the cardinal symptom(s). For all other disorders, DSM hierarchical exclusion rules were followed – thus, non-cardinal symptoms were not assessed if the cardinal symptom(s) were not endorsed because non-cardinal symptoms for these disorders are dependent on the endorsement of the cardinal symptoms(s). For example, for PTSD, it would not make sense to assess criteria B, C, D, and/or E (i.e., trauma-related symptoms) if a participant did not endorse criterion A (exposure to a trauma). Second, if a participant *did* endorse the cardinal symptom of a disorder, all subsequent skip-outs were ignored and all non-cardinal symptoms were assessed. For example, if a participant endorsed PTSD criterion A, interviewers assessed all lifetime PTSD symptoms even if the participant did not fully meet criteria B, C, D, and/or E. The suspension of skip outs when appropriate is especially important in structural psychopathology studies because adhering to the DSM-based skip structure can influence the covariance between disorders, which in turn can bias estimates of higher-order factors (Conway & Brown, 2018). Third, in contrast to the DSM's requirement that a diagnosis should not be assigned if its features are better accounted for by another disorder, symptoms were permitted to count toward multiple diagnoses. Interviewers were trained to criterion by viewing the SCID-101 training videos and completing three SCID interviews observed by an advanced interviewer in which diagnoses were in full agreement with those of the observer.

Consistent with HiTOP's and RDoC's dimensional conceptualization of psychopathology and recommendations for assessing HiTOP constructs (Stanton et al., 2020), we computed dimensional lifetime disorder severity scales by summing each symptom within a disorder, with each symptom rated as either 1 (absent), 2 (subthreshold), or 3 (full threshold). In a previous study of a subset of this sample, these lifetime

disorder severity scales were more reliable and valid than categorical diagnoses and exhibited strong internal consistency (α s > .88) and test-retest reliability (ICCs > .61; Shankman et al., 2018). Associations between disorder severity scales are presented in Table S1.

RDoC Measures

Temporal Experience of Pleasure Scale. The anticipatory and consummatory subscales of the Temporal Experiences of Pleasure Scale (TEPS; Gard et al., 2006) were used to measure the PVS Reward Responsiveness subconstructs of Reward Anticipation and Initial Response to Reward, respectively. The anticipatory subscale contains 10 items (e.g., “I look forward to a lot of things in my life”) and the consummatory subscale contains 8 items (e.g., “I enjoy taking a deep breath of fresh air when I walk outside”). Each item is rated on a 6-point Likert scale ranging from 1 (“very false for me”) to 6 (“very true for me”). Cronbach’s alphas for the anticipatory ($\alpha = .78$) and consummatory ($\alpha = .78$) subscales were adequate in the present study.

Intolerance of Uncertainty Scale-12. The Intolerance of Uncertainty Scale-12 (IUS-12; Carleton et al., 2007) assesses responses to uncertainty or ambiguity, and was used as a measure of the Potential Threat construct in the NVS. The IUS-12 contains 12 items (e.g., “Unforeseen events upset me greatly,” “The smallest doubt can stop me from acting”) rated on a 5-point Likert scale ranging from 1 (“not at all characteristic of me”) to 5 (“entirely characteristic of me”). The IUS-12 yields a total score and two subscales (prospective and inhibitory intolerance of uncertainty), but as the RDoC matrix does not distinguish between prospective and inhibitory intolerance of uncertainty, we focused on total IUS-12 scores ($\alpha = .90$).

Anxiety Sensitivity Index-3. The NVS Potential Threat construct was also assessed using the Anxiety Sensitivity Index-3 (ASI-3; Taylor et al., 2007), an 18-item self-report measure of fear of specific anxiety symptoms and associated consequences in physical, cognitive, and social domains. Items are rated on a Likert scale from 1 (“very little”) to 5 (“very much”) and were summed to form a measure of anxiety sensitivity ($\alpha = .90$). Sample items include “It scares me when my heart beats rapidly,” “I worry that other people will notice my anxiety,” and “It scares me when I am unable to keep my mind on a task.”

Delis-Kaplan Executive Function System. Subconstructs of Cognitive Control within the Cognitive Systems domain were assessed at the behavioral unit of analysis using subtests from the Delis-Kaplan Executive Function System (D-KEFS; Delis et al., 2001). Switching was assessed using the Number-Letter Sequencing condition of the Trail Making test, the Category Switching condition of the Verbal Fluency subtest, and the Switching condition of the Design Fluency subtest. As performance in these three conditions were moderately correlated (r s = .18-.31, p s < .001), a switching composite score was calculated by reverse-scoring raw scores reflecting completion time (so that higher scores reflected better performance for all variables) and averaging the z-scores of these three conditions. Inhibition was assessed using the Empty Dots Only condition from Design Fluency and the Inhibition condition from the Color-Word Interference Test, which is analogous to the classic Stroop test (Stroop, 1935). Consistent with prior studies (Rey-Mermet et al., 2018), these inhibition measures were only weakly correlated, $r = .09$, $p = .050$, and were thus examined as separate validators.

Data Analysis

Bivariate associations between disorder severity scales and RDoC (sub)constructs were calculated using multilevel correlations implemented in the *correlations* R package (Makowski et al., 2020), which adjusted for nonindependence of observations (i.e., siblings nested within families). The familial aggregations of disorder severity scales and RDoC (sub)constructs were estimated using intraclass correlations (ICCs).

A series of measurement models were fit using confirmatory factor analysis (CFA). The following models were tested: (1) a unidimensional model with one general factor, (2) a correlated two-factor model with internalizing and disinhibited externalizing dimensions, (3) a correlated three-factor model with distress, fear, and disinhibited externalizing dimensions, (4) a bifactor model with a general factor orthogonal to two specific factors (internalizing and disinhibited externalizing), and (5) a bifactor model with a general factor orthogonal to three specific factors (distress, fear, and disinhibited externalizing). The internalizing factor featured loadings from MDD, GAD, PTSD, SAD, SP, PD, and OCD, and the disinhibited externalizing factor was indicated by ALC, CANN, and STIM. In the models in which internaliz-

ing was separated into distress and fear, distress was indicated by MDD, GAD, and PTSD and fear was indicated by SAD, SP, PD, and OCD. Higher-order models with two or three first-order factors were not tested because at least four first-order factors are required to statistically compare the fit of a higher-order model and correlated factor model (Brown, 2015).

Participants were randomly assigned to be either sibling one or two within their sibling pair, making them indistinguishable dyads. CFAs were estimated in a dyadic analytic framework following Olsen and Kenny (2006; see Figure S1). Means, loadings, variances, and intra-sibling covariances were estimated for both sets of siblings, and were constrained to be equal across siblings. All models also included cross-sibling covariances for all manifest and latent variables, representing their familial aggregation.

Model fit was evaluated using traditional fit indices, which were adjusted as appropriate to account for model misfit attributable to the random assignment of siblings within each sibling pair (Olsen & Kenny, 2006). Indices of absolute fit included the comparative fit index (CFI), Tucker Lewis index (TLI), and root mean square error of approximation (RMSEA). CFI and TLI $> .90$ and RMSEA $< .06$ were considered indicative of acceptable fit (Hu & Bentler, 1999). We also report the Akaike information criterion (AIC) and Bayesian information criterion (BIC), for which lower values indicate better fit (e.g., differences of 10 points strongly favor the model with a lower value; Raftery, 1993).

Due to concerns about factor reliability in the bifactor model based on prior studies (Conway, Mansolf, et al., 2019; Levin-Aspenson et al., 2020; Martel et al., 2017; Watts et al., 2019), we assessed factor reliability for the best-fitting models. Specifically, construct replicability of the latent factors in the best-fitting correlated factor model and bifactor model was measured by Hancock's H (Hancock & Mueller, 2001), which quantifies the extent to which a factor is well-defined by its indicators and replicable. H values range from 0-1, with $> .70$ were considered acceptable (Hancock & Mueller, 2001; Rodriguez et al., 2016). For the best-fitting bifactor model, coefficient omega (ω) and omega subscale (ω_s) were used to measure the proportion of common variance explained by the general and specific factors (McDonald, 1999). Omega hierarchical (ω_h ; McDonald, 1999; Reise et al., 2013) and omega hierarchical subscale (ω_{hs}) were calculated to measure the reliability of the general and specific factors after partialling out variance attributable to the specific or general factor(s), respectively. In contrast

to H , which represents the correlation between a factor and an optimally-weighted item composite, ω_h and ω_{hs} reflect the correlation between a factor and a unit-weighted composite (Rodriguez et al., 2016). A large ω_h estimate in conjunction with relatively small ω_{hs} estimates provides evidence that the general factor is more reliable than the specific factors. Explained common variance (ECV; Reise et al., 2010), or the proportion of indicator variance explained by each factor, was used to assess the relative strengths of the general and specific factors. Omega indices and ECV range from 0-1, with values $> .80$ indicating acceptable reliability (Rodriguez et al., 2016).

Estimated factor scores for the HiTOP dimensions were extracted from the best fitting correlated factor model and bifactor model, and multilevel correlations between these factor scores were used to assess factor similarity between the two models.

Structural equation modeling (SEM) was used to predict each RDoC validator from the HiTOP dimensions. The factor structure for the SEM models was fixed to be identical to the estimated factor structure in the CFA models. As in the measurement models, the means and variances of RDoC validators and regression coefficients were constrained to equality across siblings, and cross-sibling covariances were modeled for each validator. Discriminant validity was evaluated using likelihood ratio tests to assess whether cross-sibling factor correlations or structural regression coefficients from latent factors to each RDoC validator could be equated without a decrease in model fit. All variables were standardized prior to analysis. The vast majority (98.2%) of the sample had complete disorder severity scale data. There was also minimal missing data for the RDoC validators ($\leq 6\%$), with the exception of the TEPS (21.3%), which was added to the study's questionnaire battery after data collection had commenced. Missing data were handled using full-information maximum likelihood. Analyses were performed using the *lavaan* package in R (Rosseel, 2012).

Results

Measurement Model Fit

Fit indices for the CFA models are presented in Table 2. Whereas the unidimensional model fit the data poorly, the correlated two-factors model had acceptable fit to the data. The correlated three-factors model also fit acceptably, but did not improve fit relative to the correlated two-factor model. Additionally, the distress

and fear dimensions in the correlated three-factor model were strongly correlated ($r = .89$), meaning they could not be meaningfully differentiated in these data. Consistent with prior findings, the bifactor model with two specific factors had superior fit to the correlated two-factor model as measured by the CFI, TLI, and RMSEA. The bifactor model with three specific factors was inadmissible due to a negative residual variance for PTSD, and is thus not reported. As simulation studies

have demonstrated the superior fit of the bifactor model to any data relative to correlated factor models (Greene et al., 2019), we focused our validity analyses on the correlated two-factor model and bifactor model with two specific factors, which were the two best-fitting measurement models.

Table 2. Fit indices for confirmatory factor analyses.

Model	df	χ^2	CFI	TLI	RMSEA	BIC	AIC
Unidimensional	79	271.129	.744	.676	.070	13909	13761
Correlated factor (2 factors)	77	125.451	.935	.916	.035	13774	13620
Correlated factor (3 factors)	74	127.77	.928	.903	.038	13793	13628
Bifactor (2 specific factors)	67	87.887	.972	.958	.025	13792	13602

Note. df = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; BIC = Bayesian Information Criterion; AIC = Akaike Information Criterion.

Table 3. Correlations between HiTOP dimension factor scores from the bifactor and correlated factor models.

Bifactor Dimensions			
	Internalizing	Disinhibited Externalizing	General
Correlated Factor Dimensions			
Internalizing	.31	-.18	.96
Disinhibited Externalizing	-.27	.90	.61

Note. All correlations are statistically significant at $p < .001$.

Standardized factor loadings for these two models are presented in Figure 1. All loadings in the correlated factor model were positive and significant (mean $\lambda = .51$; $ps < .001$), ranging from .25 (OCD) to .72 (ALC). Loadings were somewhat weaker for the internalizing dimension (mean $\lambda = .46$) relative to the disinhibited externalizing dimension (mean $\lambda = .64$). In the bifactor model, loadings on the specific factors were generally weaker (mean $\lambda = .34$) and had larger standard errors compared to the correlated factor model. In particular, internalizing specific factor loadings within the bifactor model (mean $\lambda = .24$) were mostly nonsignificant and

were weaker than the disinhibited externalizing specific factor loadings (mean $\lambda = .57$). Loadings on the general factor were weak to moderate (mean $\lambda = .38$; range = .13 [SP] - .69 [MDD]), and were somewhat stronger for internalizing disorders (mean $\lambda = .41$) than disinhibited externalizing disorders (mean $\lambda = .30$).

Reliability of HiTOP Factors

Reliability results are presented in Table S2. In the correlated factor model, Hancock's H values (i.e., construct replicability) were acceptable for both the inter-

nalizing (.71) and disinhibited externalizing factors (.70). In the bifactor model, H statistics were suboptimal for the internalizing (.45) and disinhibited externalizing (.62) specific factors, but were sufficient for the general factor (.71). ECV values for the general (.52) and internalizing specific factor (.29) were below acceptable cutoffs, although ECV for the disinhibited externalizing specific factor (.79) approached the cutoff for acceptability. ω/ω_s and ω_h/ω_{hs} values for all factors in the bifactor model did not meet benchmarks for acceptable reliability.

bifactor and correlated factor models are presented in Table 3. These correlations indicated that the internalizing ($r = .31$) and disinhibited externalizing ($r = .90$) dimensions were moderately and strongly similar across the two models, respectively. Additionally, the correlation between the (a) internalizing dimension in the correlated factor model, and the (b) general factor in the bifactor model was near perfect ($r = .96$). The general factor in the bifactor model was also associated with the disinhibited externalizing factor in the correlated factor model ($r = .61$).

Associations Between Factor Scores Across Models

Multilevel correlations between estimated factor scores for each estimated HiTOP dimension in the

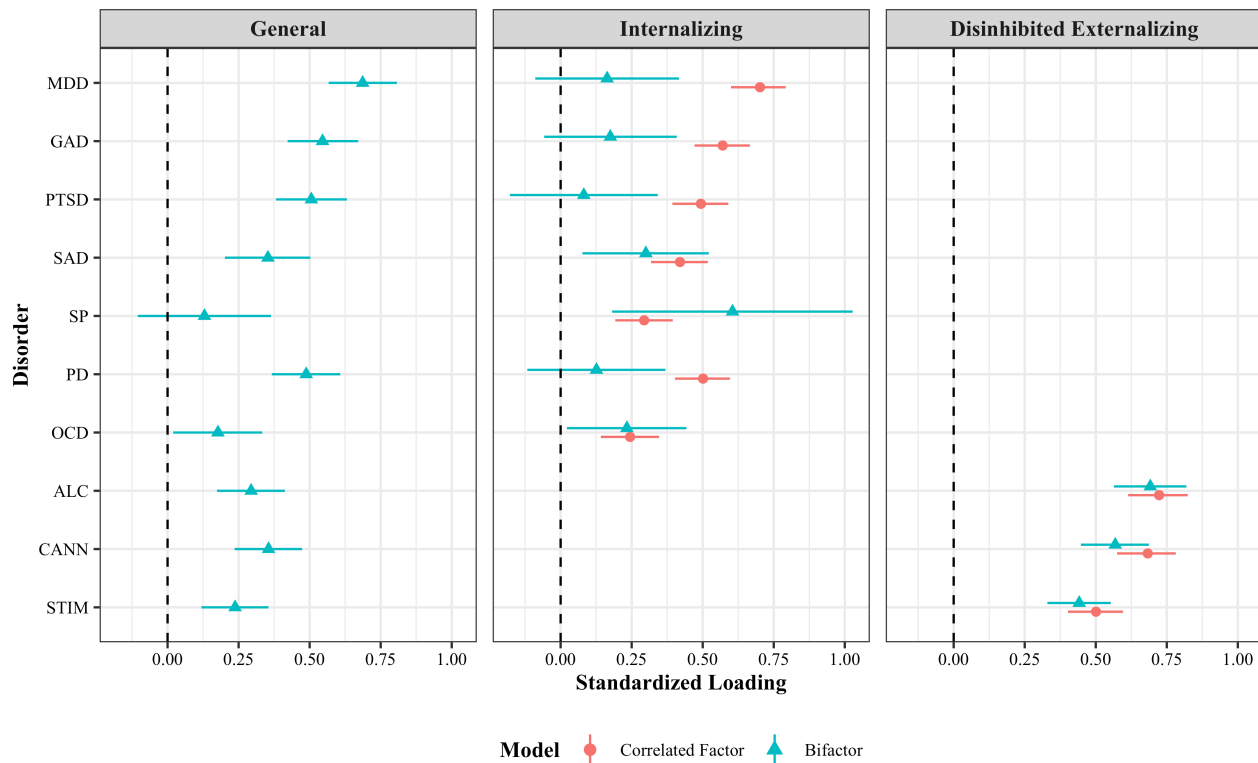


Figure 1. Standardized factor loadings for the two best-fitting measurement models.

Note. MDD = Major Depressive Disorder. GAD = Generalized Anxiety Disorder. PTSD = Post-traumatic Stress Disorder. SAD = Social Anxiety Disorder. SP = Specific Phobia. PD = Panic Disorder. OCD = Obsessive-Compulsive Disorder. ALC = Alcohol Use Disorder. CANN = Cannabis Use Disorder. STIM = Stimulant Use Disorder. Error bars represent 95% confidence intervals.

Familial Aggregation of HiTOP Factors

In the correlated factor model, the internalizing, $r = .28$, $p = .002$, and disinhibited externalizing, $r = .40$, p

$< .001$, dimensions were moderately and significantly familial, and their familial aggregations did not significantly differ in magnitude, $\Delta\chi^2(1) = .817$, $p = .366$. In the bifactor model, the general factor, $r = .51$, $p < .001$,

and the externalizing specific factor, $r = .45$, $p < .001$, were also familial. However, the familial aggregation of the internalizing dimension in the bifactor model did not significantly differ from zero, $r = -.06$, $p = .837$, and was significantly weaker than the familial aggregations of disinhibited externalizing and the general factor, $\Delta\chi^2(1) > 4$, $ps < .043$, indicating that the familial aggregation of the internalizing dimension may be absent or minimal when constrained to be orthogonal to a general psychopathology factor. The familial aggregations of disorder severity scales and RDoC (sub)construct measures are reported in Table S3.

Associations with RDoC (Sub)constructs

Standardized regression coefficients from the structural models predicting each RDoC (sub)construct from HiTOP dimensions are presented in Figure 2. For the PVS validators, disinhibited externalizing was associated with greater Initial Response to Reward in the correlated factor model. Initial Response to Reward was more strongly associated with disinhibited externalizing than internalizing, $\Delta\chi^2(1) = 5.19$, $p = .023$, which was not significantly associated with Initial Response to Reward. In the bifactor model, internalizing was negatively associated with Initial Response to Reward, whereas associations between Initial Response to Reward and disinhibited externalizing or the general factor were weakly positive and nonsignificant. Discriminant validity tests indicated that the magnitudes of these associations significantly differed such that Initial Response to Reward was significantly more related to internalizing than disinhibited externalizing, $\Delta\chi^2(1) = 15.27$, $p < .001$, and the general factor, $\Delta\chi^2(1) = 7.37$, $p = .007$. No HiTOP dimensions were associated with Reward Anticipation in either model.

For the NVS validators, both the correlated factor model and bifactor model suggested that internalizing was positively associated with both measures of Potential Threat (intolerance of uncertainty and anxiety sensitivity). In both models, the Potential Threat measures were more strongly associated with internalizing than with disinhibited externalizing, $\Delta\chi^2(1) > 33$, $ps < .001$, which was weakly and nonsignificantly related to the Potential Threat measures in both models. In the bifactor model, the general factor was also positively associated with the two Potential Threat measures. Discriminant validity tests suggested that anxiety sensitivity was more strongly associated with the general factor than with internalizing, $\Delta\chi^2(1) = 10.74$, $p = .001$, but intolerance of uncertainty was not differentially associated with the two dimensions, $\Delta\chi^2(1) = 1.11$, $p = .291$. Notably, the effects of internalizing on the Potential Threat measures were attenuated by 39-57% in the bifactor model compared to the correlated factor model due to the inclusion of the general factor.

Regarding the Cognitive Control subconstructs within the Cognitive Systems domain, the HiTOP dimensions in both the correlated factor model and bifactor model were unrelated to switching and inhibition as measured by the Color-Word Interference test. Disinhibited externalizing was unrelated to inhibition as assessed by the Design Fluency subtest in both models. Internalizing in the correlated factor model was associated with poorer inhibition as assessed by the Design Fluency subtest. In contrast, in the bifactor model, the general factor (but not internalizing) was significantly associated with poorer inhibition, although the magnitude of these two associations did not significantly differ, $\Delta\chi^2(1) = 1.44$, $p = .229$.

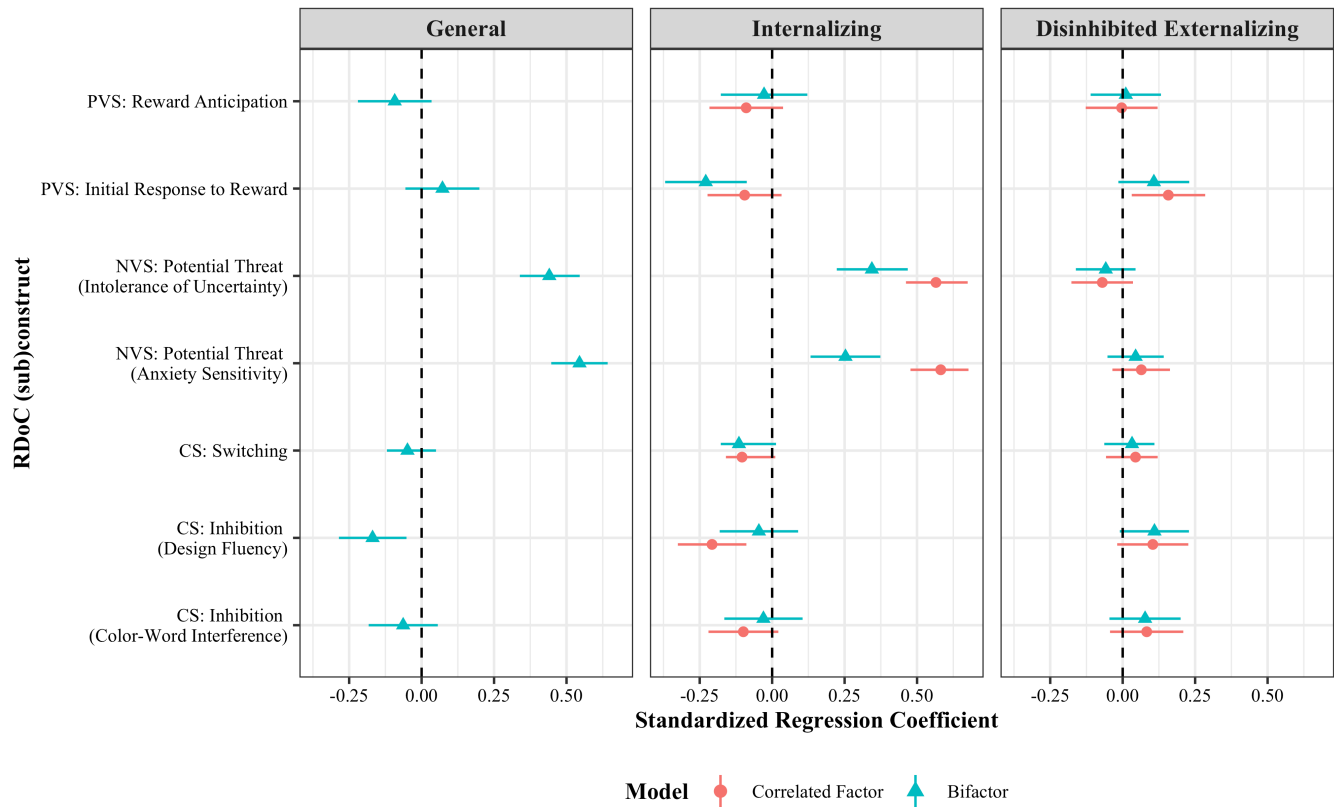


Figure 2. Structural regression coefficients predicting RDoC (sub)constructs from HiTOP dimensions in the correlated two-factor model and corresponding bifactor model. PVS = positive valence system. NVS = negative valence system. CS = cognitive systems. Error bars represent 95% confidence intervals.

Discussion

Studies examining the structure and validity of HiTOP dimensions have predominantly used bifactor models, which tend to overfit, may be unreliable, and may yield different validity patterns than alternative models (Greene et al., 2019; Hyland et al., 2020; Watts et al., 2019). The present study tested and compared the reliability and validity of higher-order HiTOP dimensions across bifactor and correlated factor models of psychopathology using familial aggregation and associations with RDoC (sub)constructs as validators.

The data were best fit by a correlated two-factor model and a bifactor model with two specific factors, with the bifactor model having slightly superior fit. The internalizing and disinhibited externalizing dimensions in the correlated factor model featured moderate loadings and demonstrated acceptable reliability. When specified as specific factors in a bifactor model, however, the internalizing and disinhibited externalizing dimensions exhibited suboptimal reliability and

had weaker factor loadings with larger standard errors. Specific factor loadings were particularly weak for internalizing disorders, which were the strongest indicators of the general factor. The general factor had acceptable construct replicability as assessed by Hancock's H , but did not meet acceptable reliability thresholds as assessed by ω , ω_h , or ECV. Cross-model correlations of factor scores revealed additional interpretative issues regarding the bifactor model. While the disinhibited externalizing factor was highly similar across the correlated factor and bifactor models ($r = .90$), the internalizing dimension in the correlated factor model was only weakly correlated with its counterpart in the bifactor model ($r = .31$) and instead was nearly identical to the general factor ($r = .96$).

Validity analyses indicated that the internalizing and disinhibited externalizing dimensions in the correlated factor model were moderately familial, and the magnitude of their familial aggregations did not significantly differ. In the bifactor model, the familial aggregations of the general factor and disinhibited externalizing dimension were both moderate to strong,

which is consistent with prior findings in youth (Martel et al., 2017; Michelini et al., 2019; Pesenti-Gritti et al., 2008; Waldman et al., 2016). However, the internalizing factor was no longer familial after partialling out variance attributable to the general factor in the bifactor model. This might suggest that familial influences on the internalizing spectrum may be largely mediated by genetic and shared environmental influences on the general factor (Lahey et al., 2011; Waldman et al., 2016). However, considering the general factor was disproportionately represented by internalizing disorders and the internalizing specific factor in the bifactor model was poorly defined and did not reach benchmarks for acceptable reliability, this finding may be due to there being relatively little reliable residual variance in the internalizing factor after removing variance accounted for by the general factor.

Cross-model comparisons of associations between HiTOP dimensions and RDoC (sub)constructs yielded a similar – albeit more nuanced – pattern. Initial Response to Reward was positively associated with the disinhibited externalizing dimension in the correlated factor model, and was weakly, nonsignificantly, and negatively associated with its counterpart specific factor in the bifactor model. Initial Response to Reward also was more strongly negatively associated with the internalizing dimension in the bifactor model than in the correlated factor model. This was one of the few instances in which a HiTOP spectrum was more strongly related to a validator in the bifactor model than in the correlated factor model. However, this effect was surprising considering that MDD – the internalizing disorder most typically associated with Initial Response to Reward (Michelini et al., 2020) – loaded weakly onto the internalizing specific factor. Instead, the internalizing specific factor was most strongly defined by specific phobia. Although Initial Response to Reward is not widely studied in relation to specific phobia (Michelini et al., 2020), several studies have found no association between Initial Response to Reward and fear-based disorders (Burkhouse et al., 2016, 2017; Kujawa et al., 2017; however, see Lieberman et al., 2017), suggesting that the association between Initial Response to Reward and the internalizing specific factor may be a spurious effect generated from the bifactor model.

In both the correlated factor and bifactor models, the NVS construct Potential Threat was positively associated with the internalizing spectrum, but was unrelated to the disinhibited externalizing spectrum. This effect was robust across two measures of Poten-

tial Threat (intolerance of uncertainty and anxiety sensitivity), and is consistent with prior studies of this construct (Correa et al., 2019; Naragon-Gainey, 2010). The general factor in the bifactor model was moderately associated with greater Potential Threat. The pattern of significant associations between HiTOP spectra and Potential Threat measures was identical between the correlated factor and bifactor models, although the effects of the internalizing dimension on intolerance of uncertainty and anxiety sensitivity were weakened by 39% and 57%, respectively, in the bifactor model compared to the correlated factor model.

Compared to the NVS and PVS (sub)constructs, the Cognitive Systems subconstructs had weaker associations with HiTOP dimensions. This is unsurprising given that the Cognitive Systems subconstructs were assessed behaviorally while the NVS and PVS (sub)constructs were assessed using self-report and thus had overlapping method variance (Campbell & Fiske, 1959) with the HiTOP assessment (although HiTOP dimensions were assessed via interview not questionnaire). Nevertheless, Design Fluency inhibition was associated with the internalizing dimension in the correlated factor model and with the general factor in the bifactor model, although the other two Cognitive Systems measures (switching and Color-Word Interference inhibition) were not related to any psychopathology dimensions in either model.

In sum, associations between HiTOP dimensions and RDoC (sub)constructs were strongest between NVS measures and the internalizing dimension in the correlated factor model and the general factor in the bifactor model, which were largely redundant with each other.

More broadly, the factor loadings, reliabilities, cross-model factor score correlations, and validity evidence suggest that the bifactor model had a vanishing internalizing specific factor, meaning that the general factor served as a referent marker of the indicators contained in the internalizing specific factor (Eid et al., 2017). This explains why the general factor was isomorphic with the internalizing dimension in the correlated factor model, and supports the hypothesis that the addition of the general factor in the bifactor model may merely serve to redistribute variance described by the correlated factor model rather than capture previously unexplained variance (Watts et al., 2019).

Strengths, Limitations, and Constraints on Generalizability

This study was well suited for modeling HiTOP dimensions and their relationships with RDoC (sub)constructs. The lack of diagnostic inclusion criteria, modifications to the interview-based psychopathology assessment (e.g., suspended certain skip-outs), and dimensional operationalization of disorders (which has better psychometric properties than diagnoses; Shankman et al., 2018) were all in line with recommendations for HiTOP and RDoC studies (Cuthbert & Insel, 2013; Stanton et al., 2020; Walther et al., 2019). Other strengths included the examination of numerous RDoC validators in the same sample, tests of the discriminant validity of HiTOP dimensions, and the use of a dyadic analytic framework to examine the familial aggregation of latent HiTOP dimensions.

However, there are a number of limitations and constraints on the generalizability of these findings, which we believe are also applicable to many other structural psychopathology studies. Cross-study differences in sample characteristics such as age, recruitment setting, and severity of psychopathology may affect the relationships between disorders upon which factor analytic models are based (Conway & Brown, 2018; Levin-Aspenson et al., 2020). Additionally, differences in psychopathology assessments such as mode of administration (e.g., interview versus questionnaire), operationalization of psychopathology (e.g., diagnoses versus continuous measures), and time period assessed (e.g., lifetime versus current) may impact generalizability. This study used interview-based continuous measures of worst-in-lifetime psychopathology in a sample of young adult community members. Although these measures have demonstrated adequate reliability in this sample (Shankman et al., 2018), results may differ as a function of sample and/or assessment characteristics.

Additionally, the extent to which these findings would generalize to studies of other HiTOP dimensions – or even studies modeling the same dimensions using different indicators – is unclear. Recent work focusing on the general psychopathology factor in bifactor models has demonstrated that (a) there is wide variability in the sets of indicators included across studies (Watts et al., 2020), and (b) differences in the set of indicators examined can have a substantial effect on the composition of the general factor (Watts et al., 2019). Future research characterizing the extent to which these issues apply to other HiTOP

dimensions (e.g., spectra) and/or models is needed. For example, the disinhibited externalizing factor was comprised solely of substance use disorders in the present study, and the extent to which the omission of other disinhibited externalizing indicators might affect generalizability is unclear. In addition, the relatively narrow measurement of disinhibited externalizing disorders and the lack of measures of other HiTOP spectra (e.g., thought disorder) meant that the general factor in the bifactor model was disproportionately defined by internalizing indicators, which may have increased its correlation with the internalizing dimension in the correlated factor model and contributed to the weaker reliability and validity of the internalizing specific factor in the bifactor model (Bornovalova et al., 2020). This also may have contributed to the associations between the general factor and NVS validators, which may alternatively be interpreted as consistent with the p-factor reflecting negative emotionality (e.g., Brandes et al., 2019). Vanishing and/or unreliable specific factors have appeared in previously reported bifactor models that included other HiTOP spectra (e.g., thought disorder) or syndromes/disorders (e.g., eating pathology) instead of, or in addition to, those examined in the present study (e.g., Caspi et al., 2014; Conway, Mansolf, et al., 2019; Hyland et al., 2020). Interestingly, these issues have also emerged in studies that, unlike the present study, had roughly equal numbers of indicators per specific factor (e.g., Caspi et al., 2014; Martel et al., 2017), suggesting that our results may not be solely attributable to the omission of a particular HiTOP dimension or the disproportionate number of indicators for internalizing relative to disinhibited externalizing.

The present study also focused on select RDoC (sub)constructs and units of analysis. HiTOP dimensions extracted from self-report measures are likely more strongly related to self-report and behavioral RDoC measures than more distal units of analysis due to common method variance, equifinality, and multifinality (Patrick & Hajcak, 2016). Future studies could extend this work by incorporating other RDoC (sub)constructs and/or units of analysis, which may help to elucidate biological processes involved in the pathophysiology of psychopathology dimensions (e.g., Perkins et al., 2020; Shankman & Gorka, 2015).

Lastly, it is important to consider the possibility that psychopathological comorbidity may be better explained by other models not examined in the present study. For example, the HiTOP structure has also been modeled using a higher-order model in which spectra are nested under the p-factor. The present study did

not include higher-order models because there were not enough first-order factors to make a higher-order model over-identified. While there is evidence that the general factor is nearly identical across the bifactor and higher-order models (Forbes et al., 2020; Hyland et al., 2020; Moore et al., 2020), spectra may be only moderately correlated between the bifactor and higher-order models (Forbes et al., 2020; Hyland et al., 2020; Moore et al., 2020). The generalizability of the bifactor criterion validity results to the higher-order model may therefore be weaker for spectra than for the general factor. Additionally, comorbidity may be explained by direct causal relationships between symptoms (Borsboom, 2017), and research examining HiTOP dimensions from this perspective (Funkhouser et al., 2020, 2021; McElroy et al., 2018) may be beneficial.

Conclusions

HiTOP is a highly promising organizational framework and has had a rapid and substantial impact on psychopathology research, but there is evidence that the reliability and validity of HiTOP dimensions may differ across different models. The weak loadings, suboptimal reliability, and weaker validity of the bifactor internalizing dimension relative to its counterpart in the correlated factor model support concerns about the interpretability of psychopathology dimensions when modeled using a bifactor model and demonstrate the importance of evaluating reliability and validity when comparing structural psychopathology models.

References

- Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: a factor-analytic study. *Psychological Monographs*, 80(7), 1–37. <https://doi.org/10.1037/h0093906>
- Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three Concerns With Applying a Bifactor Model as a Structure of Psychopathology. *Clinical Psychological Science*, 5(1), 184–186. <https://doi.org/10.1177/2167702616657069>
- Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate use of bifactor analysis in psychopathology research: Appreciating benefits and limitations. *Biological Psychiatry*, 88(1), 18–27. <https://doi.org/10.1016/j.biopsych.2020.01.013>
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5–13. <https://doi.org/10.1002/wps.20375>
- Brandes, C. M., Herzhoff, K., Smack, A. J., & Tackett, J. L. (2019). The p Factor and the n Factor: Associations Between the General Factors of Psychopathology and Neuroticism in Children. *Clinical Psychological Science*, 7(6), 1266–1284. <https://doi.org/10.1177/2167702619859332>
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research, Second Edition*. Guilford Press.
- Burkhouse, K. L., Gorka, S. M., Afshar, K., & Phan, K. L. (2017). Neural reactivity to reward and internalizing symptom dimensions. *Journal of Affective Disorders*, 217, 73–79. <https://doi.org/10.1016/j.jad.2017.03.061>
- Burkhouse, K. L., Kujawa, A., Kennedy, A. E., Shankman, S. A., Langenecker, S. A., Phan, K. L., & Klumpp, H. (2016). Neural Reactivity to Reward as a Predictor of Cognitive Behavioral Therapy Response in Anxiety and Depression. *Depression and Anxiety*, 33(4), 281–288. <https://doi.org/10.1002/da.22482>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <https://doi.org/10.1037/h0046016>
- Carleton, R. N., Norton, M. A. P. J., & Asmundson, G. J. G. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*, 21(1), 105–117. <https://doi.org/10.1016/j.janxdis.2006.03.014>
- Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, 2(2), 119–137. <https://doi.org/10.1177/2167702613497473>
- Conway, C. C., & Brown, T. A. (2018). Evaluating dimensional models of psychopathology in outpatients diagnosed with emotional disorders: A cautionary tale. *Depression and Anxiety*, 35(9), 898–902. <https://doi.org/10.1002/da.22740>
- Conway, C. C., Forbes, M. K., Forbush, K. T., Fried, E. I., Hallquist, M. N., Kotov, R., Mullins-Sweatt, S. N., Shackman, A. J., Skodol, A. E., South, S. C., Sunderland, M., Waszczuk, M. A., Zald, D. H., Afzali, M. H., Bornovalova, M. A., Caragher, N., Docherty, A. R., Jonas, K. G., Krueger, R. F., ... Eaton, N. R. (2019). A Hierarchical Taxonomy of Psychopathology Can Transform Mental Health Research. *Perspectives on Psychological Science*, 14(3), 419–436. <https://doi.org/10.1177/1745691618810696>
- Conway, C. C., Li, Y. I., & Starr, L. R. (2019). Trait anhedonia is a transdiagnostic correlate of internalizing problems during adolescence. *Journal of Research in Personality*, 81, 56–63. <https://doi.org/10.1016/j.jrp.2019.05.004>
- Conway, C. C., Mansolf, M., & Reise, S. P. (2019). Ecological validity of a quantitative classification system for mental illness in treatment-seeking adults. *Psychological Assessment*, 31(6), 730–740. <https://doi.org/10.1037/pas0000695>
- Correa, K. A., Liu, H., & Shankman, S. A. (2019). The role of intolerance of uncertainty in current and remitted internalizing and externalizing psychopathology. *Journal of Anxiety Disorders*, 62, 68–76. <https://doi.org/10.1016/J.JANXDIS.2019.01.001>
- Crane, N. A., Funkhouser, C. J., Burkhouse, K. L., Klumpp, H., Phan, K. L., & Shankman, S. A. (2021). Cannabis users demonstrate enhanced neural reactivity to reward: An event-related potential and time-frequency EEG study. *Addictive Behaviors*, 113, 106669. <https://doi.org/10.1016/j.addbeh.2020.106669>
- Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, 11(1), 126. <https://doi.org/10.1186/1741-7015-11-126>
- Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS): Examiner's Manual*. The Psychological Corporation.

- Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G -factor models: Explanations and alternatives. *Psychological Methods*, 22(3), 541–562. <https://doi.org/10.1037/met0000083>
- Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Winkur, G., & Munoz, R. (1972). Diagnostic Criteria for Use in Psychiatric Research. *Archives of General Psychiatry*, 26(1), 57–63. <https://doi.org/10.1001/archpsyc.1972.01750190059011>
- First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). *Structured Clinical Interview for DSM-5–Research Version*. American Psychiatric Association. <https://doi.org/10.1002/9781118625392.wbecp351>
- Forbes, M. K., Greene, A. L., Levin-Aspenson, H. F., Watts, A. L., Hallquist, M., Lahey, B. B., Markon, K. E., Patrick, C. J., Tackett, J. L., Waldman, I. D., Wright, A. G. C., Caspi, A., Ivanova, M., Kotov, R., Samuel, D. B., Eaton, N. R., & Krueger, R. F. (2020). *Three recommendations based on a comparison of the reliability and validity of the predominant models used in research on the empirical structure of psychopathology*. <https://doi.org/10.31219/osf.io/fhp2r>
- Funkhouser, C. J., Chacko, A. C., Correa, K. A., Kaiser, A. J. E., & Shankman, S. A. (2021). Unique longitudinal relationships between symptoms of psychopathology in youth: A cross-lagged panel network analysis in the ABCD study. *Journal of Child Psychology and Psychiatry*, 62(2), 184–194. <https://doi.org/10.1111/jcpp.13256>
- Funkhouser, C. J., Correa, K. A., Carrillo, V. L., Klemballa, D. M., & Shankman, S. A. (2019). The time course of responding to aversiveness in females with a history of non-suicidal self-injury. *International Journal of Psychophysiology*, 141, 1–8. <https://doi.org/10.1016/J.IJPSYCHO.2019.04.008>
- Funkhouser, C. J., Correa, K. A., Gorka, S. M., Nelson, B. D., Phan, K. L., & Shankman, S. A. (2020). The replicability and generalizability of internalizing symptom networks across five samples. *Journal of Abnormal Psychology*, 129(2), 191–203. <https://doi.org/10.1037/abn0000496>
- Gard, D. E., Gard, M. G., Kring, A. M., & John, O. P. (2006). Anticipatory and consummatory components of the experience of pleasure: A scale development study. *Journal of Research in Personality*, 40(6), 1086–1102. <https://doi.org/10.1016/J.JRP.2005.11.001>
- Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., Waldman, I. D., Cicero, D. C., Conway, C. C., Docherty, A. R., Fried, E. I., Ivanova, M. Y., Jonas, K. G., Latzman, R. D., Patrick, C. J., Reinighaus, U., Tackett, J. L., Wright, A. G. C., & Kotov, R. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *Journal of Abnormal Psychology*, 128(7), 740–764. <https://doi.org/10.1037/abn0000434>
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In *Structural equation modeling: Present and future - a festschrift in honor of Karl Joreskog* (pp. 195–216). Scientific Software International.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hyland, P., Murphy, J., Shevlin, M., Bentall, R. P., Karatzias, T., Ho, G. W. K., Boduszek, D., & McElroy, E. (2020). On top or underneath: Where does the general factor of psychopathology fit within a dimensional model of psychopathology? *Psychological Medicine*. <https://doi.org/10.1017/S003329172000104X>
- Insel, T. R., Cuthbert, B. N., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 167(7), 748–751. <https://doi.org/10.1176/appi.ajp.2010.09091379>
- Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 617. <https://doi.org/10.1001/archpsyc.62.6.617>
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., ... Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454–477. <https://doi.org/10.1037/abn0000258>
- Kotov, R., Ruggero, C. J., Krueger, R. F., Watson, D., Yuan, Q., & Zimmerman, M. (2011). New dimensions in the quantitative classification of mental illness. *Archives of General Psychiatry*, 68(10), 1003. <https://doi.org/10.1001/archgenpsychiatry.2011.107>
- Krueger, R. F., & Markon, K. E. (2006). Reinterpreting Comorbidity: A Model-Based Approach to Understanding and Classifying Psychopathology. *Annual Review of Clinical Psychology*, 2(1), 111–133. <https://doi.org/10.1146/annurev.clinpsy.2.022305.095213>
- Kujawa, A., Kessel, E. M., Carroll, A., Arfer, K. B., & Klein, D. N. (2017). Social processing in early adolescence: Associations between neurophysiological, self-report, and behavioral measures. *Biological Psychology*, 128, 55–62. <https://doi.org/10.1016/j.biopsycho.2017.07.001>
- Lahey, B. B., Van Hulle, C. A., Singh, A. L., Waldman, I. D., & Rathouz, P. J. (2011). Higher-order genetic and environmental structure of prevalent forms of child and adolescent psychopathology. *Archives of General Psychiatry*, 68(2), 181–189. <https://doi.org/10.1001/archgenpsychiatry.2010.192>
- Levin-Aspenson, H. F., Watson, D., Clark, L. A., & Zimmerman, M. (2020). What Is the General Factor of Psychopathology? Consistency of the p Factor Across Samples. *Assessment*. <https://doi.org/10.1177/1073191120954921>
- Lieberman, L., Gorka, S. M., Funkhouser, C. J., Shankman, S. A., & Phan, K. L. (2017). Impact of posttraumatic stress symptom dimensions on psychophysiological reactivity to threat and reward. *Journal of Psychiatric Research*, 92, 55–63. <https://doi.org/10.1016/j.jpsychires.2017.04.002>
- Makowski, D., Ben-Shachar, M., Patil, I., & Lüdtke, D. (2020). Methods and Algorithms for Correlation Analysis in R. *Journal of Open Source Software*, 5(51), 2306. <https://doi.org/10.21105/joss.02306>
- Martel, M. M., Pan, P. M., Hoffmann, M. S., Gadelha, A., do Rosário, M. C., Mari, J. J., Manfro, G. G., Miguel, E. C., Paus, T., Bressan, R. A., Rohde, L. A., & Salum, G. A. (2017). A general psychopathology factor (P Factor) in children: Structural model analysis and external validation through familial risk and child global executive function. *Journal of Abnormal Psychology*, 126(1), 137–148. <https://doi.org/10.1037/abn0000205>

- McDonald, R. P. (1999). *Test theory: A unified approach*. Earlbaum.
- McElroy, E., Belsky, J., Carragher, N., Fearon, P., & Patalay, P. (2018). Developmental stability of general and specific factors of psychopathology from early childhood to adolescence: Dynamic mutualism or p-differentiation? *Journal of Child Psychology and Psychiatry*, 59(6), 667–675. <https://doi.org/10.1111/jcpp.12849>
- Micheline, G., Barch, D. M., Tian, Y., Watson, D., Klein, D. N., & Kotov, R. (2019). Delineating and validating higher-order dimensions of psychopathology in the Adolescent Brain Cognitive Development (ABCD) study. *Translational Psychiatry*, 9(1). <https://doi.org/10.1038/s41398-019-0593-4>
- Micheline, G., Palumbo, I. M., DeYoung, C. G., Latzman, R. D., & Kotov, R. (2020). *Linking RDoC and HiTOP: A new interface for advancing psychiatric nosology and neuroscience*. <https://doi.org/10.31234/osf.io/ps7tc>
- Moore, T. M., Kaczurkin, A. N., Durham, E. L., Jeong, H. J., McDowell, M. G., Dupont, R. M., Applegate, B., Tackett, J. L., Cardenas-Iniguez, C., Kardan, O., Akcelik, G. N., Stier, A. J., Rosenberg, M. D., Hedeker, D., Bertram, M. G., & Lahey, B. B. (2020). Criterion validity and relationships between alternative hierarchical dimensional models of general and specific psychopathology. *Journal of Abnormal Psychology*, 129(7), 677–688. <https://doi.org/10.1037/abn0000601>
- Naragon-Gainey, K. (2010). Meta-Analysis of the Relations of Anxiety Sensitivity to the Depressive and Anxiety Disorders. *Psychological Bulletin*, 136(1), 128–150. <https://doi.org/10.1037/a0018055>
- Olsen, J. A., & Kenny, D. A. (2006). Structural equation modeling with interchangeable dyads. *Psychological Methods*, 11(2), 127–141. <https://doi.org/10.1037/1082-989X.11.2.127>
- Patrick, C. J., & Hajcak, G. (2016). RDoC: Translating promise into progress. *Psychophysiology*, 53(3), 415–424. <https://doi.org/10.1111/psyp.12612>
- Perkins, E. R., Latzman, R. D., & Patrick, C. J. (2020). Interfacing neural constructs with the Hierarchical Taxonomy of Psychopathology: “Why” and “how.” *Personality and Mental Health*, 14(1), 106–122. <https://doi.org/10.1002/pmh.1460>
- Pesenti-Gritti, P., Spatola, C. A. M., Fagnani, C., Oglia, A., Patriarca, V., Stazi, M. A., & Battaglia, M. (2008). The co-occurrence between internalizing and externalizing behaviors: A general population twin study. *European Child and Adolescent Psychiatry*, 17(2), 82–92. <https://doi.org/10.1007/s00787-007-0639-7>
- Radoman, M., Phan, K. L., & Gorka, S. M. (2019). Neural correlates of predictable and unpredictable threat in internalizing psychopathology. *Neuroscience Letters*, 701, 193–201. <https://doi.org/10.1016/j.neulet.2019.02.031>
- Raftery, A. E. (1993). Bayesian model selection in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 163–180). Sage.
- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95(2), 129–140. <https://doi.org/10.1080/00223891.2012.725437>
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544–559. <https://doi.org/10.1080/00223891.2010.496477>
- Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning Memory and Cognition*, 44(4), 501–526. <https://doi.org/10.1037/xlm0000450>
- Robins, E., & Guze, S. B. (1970). Establishment of Diagnostic Validity in Psychiatric Illness: Its Application to Schizophrenia. *American Journal of Psychiatry*, 126(7), 983–987. <https://doi.org/10.1176/ajp.126.7.983>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137–150. <https://doi.org/10.1037/met0000045>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48. <https://doi.org/10.18637/jss.v048.i02>
- Shankman, S. A., Funkhouser, C. J., Klein, D. N., Davila, J., Lerner, D., & Hee, D. (2018). Reliability and validity of severity dimensions of psychopathology assessed using the Structured Clinical Interview for DSM-5 (SCID). *International Journal of Methods in Psychiatric Research*, 27(1), e1590. <https://doi.org/10.1002/mpr.1590>
- Shankman, S. A., & Gorka, S. M. (2015). Psychopathology research in the RDoC era: Unanswered questions and the importance of the psychophysiological unit of analysis. *International Journal of Psychophysiology*, 98(2), 330–337. <https://doi.org/10.1016/j.ijpsycho.2015.01.001>
- Stanton, K., McDonnell, C. G., Hayden, E. P., & Watson, D. (2020). Transdiagnostic approaches to psychopathology measurement: Recommendations for measure selection, data analysis, and participant recruitment. *Journal of Abnormal Psychology*, 129(1), 21–28. <https://doi.org/10.1037/abn0000464>
- Stevens, E. S., Lieberman, L., Funkhouser, C. J., Correa, K. A., & Shankman, S. A. (2019). Startle during threat longitudinally predicts functional impairment independent of DSM diagnoses. *Psychiatry Research*, 279, 207–215. <https://doi.org/10.1016/j.psychres.2019.02.072>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662. <https://doi.org/10.1037/h0054651>
- Taylor, S., Zvolensky, M. J., Cox, B. J., Deacon, B., Heimberg, R. G., Ledley, D. R., Abramowitz, J. S., Holaway, R. M., Sandin, B., Stewart, S. H., Coles, M., Eng, W., Daly, E. S., Arrindell, W. A., Bouvard, M., & Cardenas, S. J. (2007). Robust Dimensions of Anxiety Sensitivity: Development and Initial Validation of the Anxiety Sensitivity Index-3. *Psychological Assessment*, 19(2), 176–188. <https://doi.org/10.1037/1040-3590.19.2.176>
- Waldman, I. D., Poore, H. E., van Hulle, C., Rathouz, P. J., & Lahey, B. B. (2016). External validity of a hierarchical dimensional model of child and adolescent psychopathology: Tests using confirmatory factor analyses and multivariate behavior genetic analyses. *Journal of Abnormal Psychology*, 125(8), 1053–1066. <https://doi.org/10.1037/abn0000183>
- Walther, S., Bernard, J. A., Mittal, V. A., & Shankman, S. A. (2019). The utility of an RDoC motor domain to understand psychomotor symptoms in depression. *Psychological Medicine*, 49(2), 212–216. <https://doi.org/10.1017/S0033291718003033>
- Watson, D. (2005). Rethinking the mood and anxiety disorders: A quantitative hierarchical model for DSM-V. *Journal of Abnormal Psychology*, 114(4), 522–536. <https://doi.org/10.1037/0021-843X.114.4.522>
- Watts, A. L., Lane, S. P., Bonifay, W., Steinley, D., & Meyer, F. A. C. (2020). Building theories on top of, and not independent of, statistical models: The case of the p-factor. *Psychological*

- Inquiry*, 31(4), 310–320.
<https://doi.org/10.1080/1047840X.2020.1853476>
- Watts, A. L., Poore, H. E., & Waldman, I. D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science*, 7(6), 1285–1303.
<https://doi.org/10.1177/2167702619855035>
- Zalta, A. K., & Shankman, S. A. (2016). Conducting Psychopathology Prevention Research in the RDoC Era. *Clinical Psychology: Science and Practice*, 23(1), 94–104.
<https://doi.org/10.1111/cpsp.12144>