

















Recommendations for Adjudicating Among Alternative Structural Models of Psychopathology

Clinical Psychological Science
2023, Vol. 11(4) 616–640
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/21677026221144256
www.psychologicalscience.org/CPS



Irwin D. Waldman¹ , **Christopher D. King¹**, **Holly E. Poore²**,
Justin M. Lunningham³, **Richard M. Zinbarg⁴** ,
Robert F. Krueger⁵, **Kristian E. Markon⁶** ,
Marina Bornovalova⁷, **Michael Chmielewski⁸**,
Christopher Conway⁹ , **Michael Dretsch¹⁰**,
Nicholas R. Eaton¹¹, **Miriam K. Forbes¹²** , **Kelsie Forbush¹³** ,
Kristin Naragon-Gainey¹⁴ , **Ashley Lauren Greene¹⁵**,
J. D. Haltigan¹⁶, **Masha Ivanova¹⁷**, **Keanan Joyner¹⁸** ,
Katherine M. Keyes¹⁹, **Kevin M. King²⁰** , **Roman Kotov²¹**,
Holly Levin-Aspenson²² , **Thomas Olino²³** , **Jason A. Oliver²⁴**,
Christopher J. Patrick²⁵ , **David Preece²⁶**, **Lauren A. Rutter²⁷**,
Martin Sellbom²⁸, **Susan South²⁹** , **Nicholas J. Wagner³⁰**,
Ashley L. Watts³¹ , **Sylia Wilson³²** , **Aidan G.C. Wright³³** ,
and **David Zald²**

¹Department of Psychology, Emory University; ²Department of Psychiatry, Robert Wood Johnson Medical School, Rutgers University; ³Department of Biostatistics and Epidemiology, School of Public Health, University of North Texas Health Sciences Center; ⁴Department of Psychology, Northwestern University; ⁵Department of Psychology, University of Minnesota; ⁶Department of Psychological and Brain Sciences, University of Iowa; ⁷Department of Psychology, University of South Florida; ⁸Department of Psychology, Southern Methodist University; ⁹Department of Psychology, Fordham University; ¹⁰U.S. Army Medical Research Directorate-West, Walter Reed Army Institute of Research, Silver Spring, Maryland; ¹¹Department of Psychology, Stony Brook University; ¹²Department of Psychology, Macquarie University; ¹³Department of Psychology, University of Kansas; ¹⁴School of Psychological Science, University of Western Australia; ¹⁵Department of Psychiatry, Icahn School of Medicine at Mount Sinai; ¹⁶Center for Addiction and Mental Health, University of Toronto; ¹⁷Department of Psychiatry, University of Vermont Medical School; ¹⁸Department of Psychology, University of California, Berkeley; ¹⁹Department of Epidemiology, Mailman School of Public Health, Columbia University; ²⁰Department of Psychology, University of Washington; ²¹Department of Psychiatry, Stony Brook University Medical School; ²²Department of Psychology, University of North Texas; ²³Department of Psychology, Temple University; ²⁴Department of Psychiatry and Behavioral Sciences, Duke University Medical Center; ²⁵Department of Psychology, Florida State University; ²⁶School of Psychology, Curtin University; ²⁷Department of Psychological and Brain Sciences, Indiana University; ²⁸Department of Psychology, University of Otago; ²⁹Department of Psychological Sciences, Purdue University; ³⁰Department of Psychological and Brain Sciences, Boston University; ³¹Department of Psychology, Vanderbilt University; ³²Institute of Child Development, University of Minnesota; and ³³Department of Psychology, University of Pittsburgh

Corresponding Author:

Irwin D. Waldman, Department of Psychology, Emory University
Email: psyiw@emory.edu

Abstract

Historically, researchers have proposed higher-order factors to explicate the structure of psychopathology, including Externalizing, Internalizing, Fear, Distress, Thought Disorder, and a general factor. Despite extensive research in this domain, the underlying structure of psychopathology remains unresolved. Here, we examine several issues in adjudicating among structural models of psychopathology. Using simulations and analyses of the extant literature, we contrast the model-based reliability of alternative structural models of psychopathology and highlight shortcomings of conventional model-fit indices for such adjudication. We propose alternative criteria for evaluating and contrasting competing structural models, including various model characteristics (e.g., the magnitude and consistency of factor loadings and their precision), the consistency and sensitivity of factors to their constituent indicators, and the variance explained in and patterns of associations with relevant variables. Using these criteria as adjuncts to conventional fit indices should become standard practice and will greatly facilitate adjudication among alternative structural models of psychopathology.

Keywords

classification, comorbidity, dimensional vs categorical, psychopathology, statistical analysis

Received 4/23/22; Revision Accepted 11/22/22

Multiple attempts have been made to classify psychopathology and to grapple with the observation that individual disorders are overlapping, a phenomenon referred to as *comorbidity* (Feinstein, 1970). The prototypical way of studying comorbidity from the 1980s to the 2000s was to examine the overlap among discrete diagnoses, often two at a time. Examples of this include major depressive disorder and generalized anxiety disorder (Fava et al., 2000; Kessler et al., 2008) and attention-deficit/hyperactivity disorder (ADHD) with oppositional defiant disorder (ODD) and conduct disorder (Biederman et al., 1991).

Apropos of this special section and the place of this article therein, some of Scott Lilienfeld's earliest publications reflected his burgeoning interests in the classification of psychopathology and comorbidity (Lilienfeld, 1992; Lilienfeld et al., 1986; Lilienfeld & Waldman, 1990). For his comprehensive exam paper at the University of Minnesota, Lilienfeld reviewed the evidence across multiple domains—studies of classification and diagnostic overlap, course and outcome, familiarity and available behavior genetic studies, and psychophysiological correlates—for the validity of the “Saint Louis quartet,” a set of conditions that included psychopathy, antisocial behavior, somatization, and histrionic personality disorder. Later, while on his clinical internship at Western Psychiatric Institute and Clinic in Pittsburgh, Pennsylvania, Scott published the first article on which he was lead author, which examined the relation of histrionic personality disorder to antisocial personality and somatization disorders (Lilienfeld et al., 1986). This article proved to be a harbinger of Scott's interests in comorbidity and classification of psychopathology, which were reflected in many subsequent publications. These included a review and integration of theoretical models of the association between antisocial personality and somatization disorders (Lilienfeld, 1992) and a comprehensive review of the overlap between ADHD in childhood and later aggression and antisocial behavior

(Lilienfeld & Waldman, 1990), as well as a subsequent publication on the overlap and distinctions between ADHD and ODD (Waldman & Lilienfeld, 1991). Scott's work in this domain also included critiques of the concept and use of the term comorbidity (Lilienfeld et al., 1994; Lilienfeld & Waldman, 2004) and proposed extensions of the study of comorbidity and classification using various types of latent-variable models (Waldman & Lilienfeld, 2001; Waldman et al., 1995). Indeed, a snippet of the abstract of one of these articles (Lilienfeld et al., 1994) seems rather prescient in hindsight, as it stated that most uses of the term comorbidity

blur the distinction between latent constructs and manifest indicators . . . The authors conclude that . . . application of the term comorbidity to psychopathological syndromes encourages the premature reification of diagnostic entities and arguably has led to more confusion than clarification. (p. 71)

Paralleling Scott's work, the historical use of comorbidity was supplanted by *transdiagnostic* approaches to the classification of psychopathology beginning in the 1990s and continues today. In a transdiagnostic approach, the overlap among disorders or covariation among symptom dimensions is often captured by one or more latent dimensions. Transdiagnostic approaches recognize the fact that multiple disorders share common risk factors and correlates, show common course and outcomes, and may be ameliorated by the same treatments (Barlow, Farchione, Bullis, et al., 2017; Barlow, Farchione, Sauer-Zavala, et al., 2017). Canonical contributions to this approach include characterizing the overlap among children's symptoms using Externalizing and Internalizing dimensions (Achenbach, 1966) and among common adult psychiatric diagnoses using Externalizing, Distress, and Fear dimensions (Krueger, 1999).

More recently, there has been a shift from a *trans-diagnostic* approach to what might be termed a *trans-dimensional* approach in contemporary studies of psychopathology. The transdimensional approach differs from the transdiagnostic approach in that higher-order dimensions explain covariation among lower-order dimensions. This approach can be conceptualized as a hierarchical structure in which latent dimensions are further classified as sharing a higher-order dimension because of their substantial covariance. Examples of this approach include Distress and Fear dimensions loading on a higher-order Internalizing factor (Krueger, 1999); Antagonistic and Disinhibited Antisocial Behavior loading on a higher-order Externalizing factor (Burt, 2009, 2012; Kotov et al., 2017, 2021; Lahey et al., 2017a); and various diagnoses, symptom dimensions, or symptoms loading on a general psychopathology factor, often termed the “p” factor (Caspi et al., 2014; Caspi & Moffitt, 2018; Lahey et al., 2012, 2017a). In a transdimensional approach, the focus shifts from attempting to find common correlates of and risk factors for multiple diagnoses to finding such correlates and putative causes of multiple higher-order dimensions (e.g., Lee et al., 2021; Neumann et al., 2016; Riglin et al., 2020). Transdimensional approaches may also better avoid the content overlap between different disorders and the heterogeneity within diagnoses. The general factor of psychopathology has received particular attention as a transdimensional construct in the contemporary psychopathology literature over the past decade, as witnessed by the many studies that have used it to model the covariation among psychopathology dimensions (e.g., Caspi et al., 2014; Caspi & Moffitt, 2018; Lahey et al., 2012, 2017a, 2017b).

The transdimensional approach has been advocated and considerably facilitated by the Hierarchical Taxonomy of Psychopathology (HiTOP) Consortium and overarching comprehensive model (DeYoung et al., 2022; Kotov et al., 2017, 2021), which characterizes psychopathology dimensionally rather than categorically and is hierarchical in the sense that psychopathology is organized using a set of dimensions of increasing generality and comprehensiveness. The overarching HiTOP model is intended to reduce the heterogeneity within and comorbidity among diagnostic categories, and its components are intended to be construed as testable hypotheses that are subject to falsification and revision (DeYoung et al., 2022; Kotov et al., 2021; Krueger et al., 2018). Given its comprehensive overarching nature, the HiTOP model can best be viewed as a framework that subsumes most extant structural models of psychopathology that have been supported by a preponderance of evidence.

A number of structural representations of psychopathology have also been advanced in the literature. Examples include a two-factor model comprising

correlated Externalizing and Internalizing dimensions; a three-factor model distinguishing Distress from Fear within Internalizing; models including Thought Disorder and Neurodevelopmental Disorders factors; and models that include a general psychopathology factor that influences diagnoses, symptom dimensions, or individual symptoms (hereafter referred to as *indicators*). Despite these different approaches and a multitude of studies, there is only partial consensus on the underlying structure of psychopathology. Researchers studying the structure of psychopathology tend to emphasize substantive differences among alternative models (e.g., distinguishing Distress from Fear within Internalizing, uneven coverage of psychopathology across studies) while failing to consider methodological issues (e.g., overfitting, bias in tests of certain models) that can spuriously favor one model over alternatives. As an example, the general factor of psychopathology and the bifactor model from which it typically emerges have shown a sharp rise in usage and popularity among psychopathology researchers (Bornovalova et al., 2020; Greene et al., 2019; Levin-Aspenson et al., 2021; Smith et al., 2020). Nonetheless, statisticians have pointed out difficulties in distinguishing between bifactor and both correlated-factors and higher-order models that include a general factor (Gignac, 2008; Markon, 2019; Mulaik & Quartetti, 1997; Yung et al., 1999), as well as the tendency for common statistical fit indices to be biased in favor of the bifactor model (Bonifay & Cai, 2017; Bonifay et al., 2017; Greene et al., 2019; Murray & Johnson, 2013). Also, although statisticians have emphasized the utility of simulation studies for elucidating various issues and biases in differentiating among alternative structural models of psychopathology, simulations remain underused (cf. Greene et al., 2019). Given concerns with the overreliance on fit indices, model-based reliability indices (e.g., H , ω_H) for adjudicating among structural models of psychopathology and evaluating their factors' reliability have recently been proposed (Bornovalova et al., 2020; Forbes, Greene, et al., 2021; Martel et al., 2017; Rodriguez et al., 2016; Waldman, 2017; Watts et al., 2019).

The Current Study

Our goal in this article is to elucidate a set of concerns and issues with current methods for adjudicating among structural models of psychopathology and to propose solutions and alternative criteria for adjudicating among such models. These concerns and issues include the following: (a) Conventional fit indices are useful for comparing some models but not others, (b) model-based reliability indices have both advantages and disadvantages for adjudicating among competing alternative models, (c) the consistency of factor loadings varies across models and can be a useful index of model

validity, (d) factors are quite sensitive to their constituent indicators in some models but not others, (e) differences in the pattern and magnitude of associations with relevant criterion variables can help in adjudicating among models, and (f) psychopathology researchers need a greater awareness of statistically distinguishable versus indistinguishable models. These concerns and issues are illustrated using simulations and analyses of the results from extant studies. We propose several alternative criteria for evaluating and contrasting competing structural models, including various model characteristics (e.g., the magnitude and consistency of factor loadings and their precision), the sensitivity of factors to their constituent indicators and the consistency of factor loadings across models, and the percentage of variance explained in and patterns of associations with relevant criterion variables.

Method

None of the analyses reported in this article were pre-registered. Supplementary text, figures, and tables, as well as the *Mplus*, *R*, and *SPSS* code used in analyses, can be found in the Supplemental Material available online. We report all data inclusion and exclusion procedures, all manipulations, and all measures used. Given that this study involved analyses of existing data rather than new data collection, we did not determine sample sizes, as these were determined by the authors of the original studies reanalyzed here. In addition, we report details and results of all simulations we conducted as part of the work presented here. All of the studies that contributed data to the analyses reported here received approval from the institutional review boards at the authors' home institutions.

Samples and procedures

To better characterize current practices in the literature and to illustrate our concerns with concrete examples, we conducted a set of simulations and real-data analyses to address the specific concerns and issues raised above. First, we conducted a set of simulations of confirmatory factor analyses (CFAs) using *Mplus* (Version 7.4; Muthén & Muthén, 2012) to examine issues of overfitting and bias in commonly used fit indices. These simulations extend previous work on overfitting and fitness propensity (Bonifay & Cai, 2017; Preacher, 2006) and simulations previously used to examine bias in models of psychopathology (Greene et al., 2019). We used the factor loadings and factor correlations from the three-correlated-factors model (the best-fitting model) and the modified bifactor model with three correlated factors from Watts et al. (2019) as the true parameter values in the two simulations conducted. To

address issues of overfitting, we examined indices of model fit (root-mean-square error of approximation [RMSEA] and standardized root-mean-square residual [SRMR]; see Figs. S2a and S2b in the Supplemental Material for the Bayesian information criterion [BIC]), as well as the percentage of the replications that did not converge for each of the alternative models. We present results for a sample size of 10,000 using 10,000 replications and fitted the following alternative models: (a) three correlated factors, (b) two correlated factors, (c) one general factor, (d) bifactor model with three orthogonal factors, (e) bifactor model with two orthogonal factors, (f) modified bifactor model with three correlated factors, and (g) modified bifactor model with two correlated factors. Conventional statistical fit indices (i.e., RMSEA, SRMR, and BIC) and their variability were estimated across the replications. The *Mplus* scripts used to conduct these simulations are presented in the Supplemental Material.

Second, to address many of the issues listed above that we raise about the structure of psychopathology literature broadly, we conducted analyses of 100 studies that are representative of the extant literature in this domain (these studies are listed in a separate References section in the Supplemental Material). We conducted a systematic search for empirical studies and adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for reporting search procedures and study methods. A PRISMA flowchart for study inclusion/exclusion is shown in Figure S1 in the Supplemental Material. Studies were included in this review if they consisted of original empirical research that characterized psychopathology broadly and if they tested and presented at least one latent factor model of psychopathology. Studies were excluded if they did not test a structural model of psychopathology or if they modeled only a narrow facet of psychopathology (e.g., only various dimensions of anxiety disorders). To be included in our analyses of CFA models, studies had to have conducted one or more CFAs. We used the following keywords entered into Google Scholar via the *Publish or Perish* software (Harzing, 2016): “psychopathology factor structure dimension” or “bifactor” or “p factor” or “general factor” or “specific factor” or “correlated factors” or “hierarchical.” Literature reviews and reference sections of the identified articles were examined for relevant articles that were missed in the original search. In addition, the Google Scholar “cited by” function was used to search for relevant articles citing the studies already found. Two graduate students independently screened and read the identified studies, recording data on each study's methodology and results in a spreadsheet. Titles and abstracts of the studies were reviewed, and studies were included or excluded on the basis of the eligibility

criteria mentioned above. If there was ambiguity about a study meeting the inclusion criteria after this step, the students and the first author together reviewed the article. If they determined by consensus that an article did not model psychopathology broadly or use CFA, it was excluded from analyses. All data collected from these studies are described in the Supplemental Material. The following data were used in analyses: (a) year of publication, (b) number of models tested, (c) types of models tested (e.g., bifactor, correlated factors), (d) best-fitting model type, (e) ad hoc model features (e.g., correlation among the specific [i.e., group] factors in a bifactor model, correlated residuals), (f) concerning results (e.g., negative residual variances), (g) number of indicators used per factor, (h) factor loadings and correlations and their standard errors reported for best and alternative models, (i) correlations with external criteria reported for one or more models, and (j) types of specific factors tested in bifactor models.

Third, for some of the analyses, we relied heavily on two large, sociodemographically diverse, population-representative twin studies included in the 100 studies for which we had additional data. These were the Tennessee Twin Study (TTS; Lahey et al., 2011; Waldman et al., 2016), which includes 3,136 twins between the ages of 9 and 17 (49% male; 71% non-Hispanic European American ethnicity, 24% African American ethnicity, and 5% mixed or other ethnicity), and the Georgia Twin Study (GTS; Singh & Waldman, 2010; Watts et al., 2019), which includes 2,498 twins and their siblings between the ages of 5 and 18 (49% male; 82% non-Hispanic European American ethnicity, 11% African American ethnicity, and 7% mixed or other ethnicity). Family income for TTS and GTS participants at recruitment ranged from \$0 to \$150,000 (TTS: $M = \$58,633$, $SD = \$43,086$; GTS: $M = \$53,000$, $SD = \$28,500$). In the TTS, psychopathology was based on diagnostic interviews of both caretakers and youth using the Child and Adolescent Psychopathology Scale, whereas in the GTS, psychopathology was based on parent ratings on the Emory Combined Rating Scale (Waldman et al., 1998), a parent-report questionnaire assessing symptoms of the major *Diagnostic and Statistical Manual of Mental Disorders (DSM)* childhood psychiatric disorders (American Psychiatric Association, 2013). Further information on the participants and psychopathology measures included are presented in representative publications from these studies (Lahey et al., 2011; Singh & Waldman, 2010; Waldman et al., 2016; Watts et al., 2019).

Data analyses

For the simulations in our first set of analyses, we conducted a set of CFAs based on the results of a prior

study (Watts et al., 2019), in which alternative structural models of psychopathology were contrasted using CFA. As stated above, for the true parameter values in the simulations, we used the factor loadings, factor correlations, and residual variances from the best-fitting model (the three-correlated-factors model) as well as from the alternative modified bifactor model with three correlated specific factors. We used maximum likelihood estimation and recorded the number of nonconvergences and fit indices (RMSEA, SRMR, and BIC) and their 95% confidence intervals (CIs) across the 10,000 replications.

For the analyses of real data in our second and third sets of analyses, we relied primarily on several types of general linear models, including t tests, Pearson correlations, simple and multiple regression analyses, and one-way and multifactor analyses of variance. Effect sizes and their 95% CIs for H and the median, standard deviation, and standard errors of standardized factor loadings were presented alongside all statistical tests. In addition, for the third set of analyses, we conducted a set of additional CFAs over and above those conducted in the original publication (Waldman et al., 2016), in which we repeated the CFAs conducted in the original publication but removed one indicator at a time from the general and specific factors and correlated factors in order to examine the sensitivity of the factors to inclusion or exclusion of each of their indicators. We also conducted a parallel set of exploratory structural equation models (ESEMs), each containing three factors, that estimated models that included three correlated factors, three orthogonal factors, or one general plus two correlated specific factors. We used a robust maximum likelihood estimator to account for nonnormality and clustering of samples and either geomin, bi-geomin, or geomin (orthogonal) rotations.

Results

As mentioned above, we first conducted a set of simulations of CFAs to examine issues of overfitting and nonconvergence and to extend previous literature on fitting propensity (Bonifay & Cai, 2017; Preacher, 2006) to several commonly used structural models of psychopathology. We next conducted analyses of 100 studies that are representative of the extant literature on the structure of psychopathology to examine advantages and disadvantages of model-based reliability indices for adjudicating among alternative models, as well as the consistency and precision of factor loadings across models. Finally, we used data from two large twin studies (included in the aforementioned 100 studies) to examine the sensitivity of factors to their constituent indicators in bifactor and correlated-factors models,

differences in the pattern and magnitude of associations with relevant criterion variables in adjudicating among models, and statistically distinguishable versus indistinguishable models.

Conventional model-fit indices are useful for comparing some models but not others

Researchers have documented limitations of conventional statistical fit indices (e.g., overfitting, bias in tests of certain models) for adjudicating among alternative structural models of psychopathology and cognitive ability (Bonifay & Cai, 2017; Bonifay et al., 2017; Forbes, Greene et al., 2021; Greene et al., 2019; Morgan et al., 2015; Murray & Johnson, 2013; Waldman, 2017; Watts et al., 2019) and have suggested alternative criteria (Bonifay & Cai, 2017; Bonifay et al., 2017; Forbes, Greene, et al., 2021; Waldman, 2017; Watts et al., 2019). Unfortunately, this may lead to a sentiment that fit indices are never useful for adjudicating among alternative models, which is untrue (McNeish & Wolf, 2021). Rather, fit indices may be useful for contrasting some models but not others; thus, it is hard to know in which scenarios these are meaningful and unbiased and in which they are misleading and biased. Nonetheless, these two types of scenarios and how to tell them apart remain unclear. To illustrate this, we conducted the simulations described in the Method section, extending previous simulations (Bonifay & Cai, 2017; Greene et al., 2019) to cover a wider variety of correlated-factors and bifactor models commonly used in contemporary research on the structure of psychopathology.

As shown in Figure 1a, when the true parameter values were generated by the three-correlated-factors model, model fit represented by the RMSEA very successfully discriminated between the true three-correlated-factors model and the incorrect two-correlated-factors and one-general-factor models—as indicated by their nonoverlapping 95% CIs. In contrast, each of the misspecified bifactor models fitted as well or better than the true three-correlated-factors model. Similar results were found for the SRMR (as shown in Fig. 1a) but not for the BIC (as shown in Fig. S2a), as the three-correlated-factors and two-correlated-factors models could not be reliably discriminated from each other using the BIC, given their overlapping 95% CIs. This means that, in practice, researchers relying solely on these fit indices would likely choose an incorrect model of the structure of psychopathology. In contrast, as shown in Figure 1b, when the true parameter values were generated by the model with one general plus three correlated factors, the superiority of this model over all competing alternative models was clear according to the RMSEA. Similar results were found for the SRMR (as shown by the

nonoverlapping 95% CIs in Fig. 1b) but not for the BIC (as shown in Fig. S2b), as the model with one general plus three correlated factors could not be reliably discriminated from other models using the BIC, given their overlapping 95% CIs. The results shown in Figures 1a and 1b suggest an asymmetry in the ability of fit indices to adjudicate among some models but not others. Similar to previous studies, these results also demonstrate the potential for considerable overfitting in bifactor models (Bonifay & Cai, 2017; Bonifay et al., 2017; Forbes, Greene et al., 2021; Greene et al., 2019; Preacher, 2006; Watts et al., 2019).

We also examined the nonconvergence rate for each of the models in these simulations. Although most of the models converged in each of the 10,000 replications when the true parameter values were generated by the three-correlated-factors model, three of the four bifactor models showed appreciable nonconvergence rates (12% for the model with one general plus three correlated factors, 27% for the model with one general plus two correlated factors, and 63% for the model with one general plus three orthogonal factors). In contrast, when the true parameter values were generated by the model with one general plus three correlated factors, very few nonconvergences (33 of 10,000 replications) were observed only for the model with one general plus three orthogonal factors and for no other models.

The importance of these findings is highlighted by the entries in Figure 2, which shows the ad hoc model specifications and concerning results for the best-fitting bifactor and correlated-factors models that are very commonly used in this literature. Researchers will sometimes make ad hoc modifications to model specifications simply to improve model fit, even if the modifications make the model more difficult to interpret or do not align with theory. As Figure 2 shows, these model modifications and concerning results are more frequent in the best-fitting bifactor models than in the best-fitting correlated-factors models. Specifically, ad hoc model modifications were used in, and concerning results occurred in, 62% and 61% of the best-fitting bifactor models and only 18% and 5% of the best-fitting correlated-factors models (odds ratio = 7.27, 95% CI = [2.01, 26.29], Fisher's exact test: $p = .0014$, and odds ratio = 29.23, 95% CI = [3.48, 245.64], Fisher's exact test: $p = .000037$, respectively). Given the percentage of these ad hoc modifications in the best-fitting bifactor models, these model respecifications appear to be included either to improve model fit or to modify a model that did not converge so that it would run successfully. The substantial rate of nonconvergence for three of the four bifactor models in our first set of simulations in which the true parameter values were generated by the three-correlated-factors model suggests that

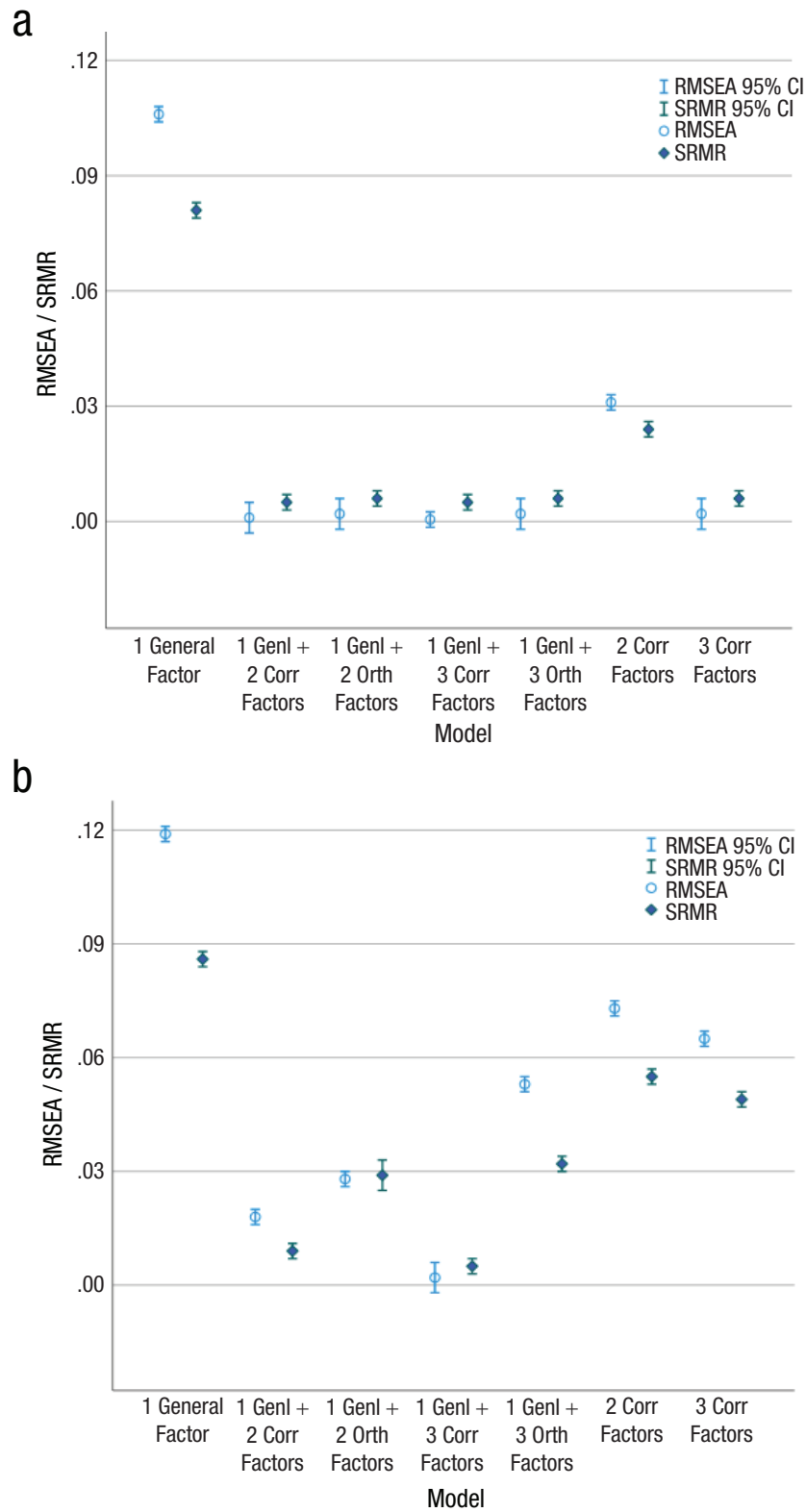


Fig. 1. Root-mean-square error of approximation (RMSEA) and standardized root-mean-square residual (SRMR) for the seven models investigated in the present study, with (a) showing results for the three-correlated-factors model as the true generating model and (b) showing results for the modified bifactor model with one general and three correlated factors as the true generating model. Error bars indicate 95% confidence intervals (CIs).

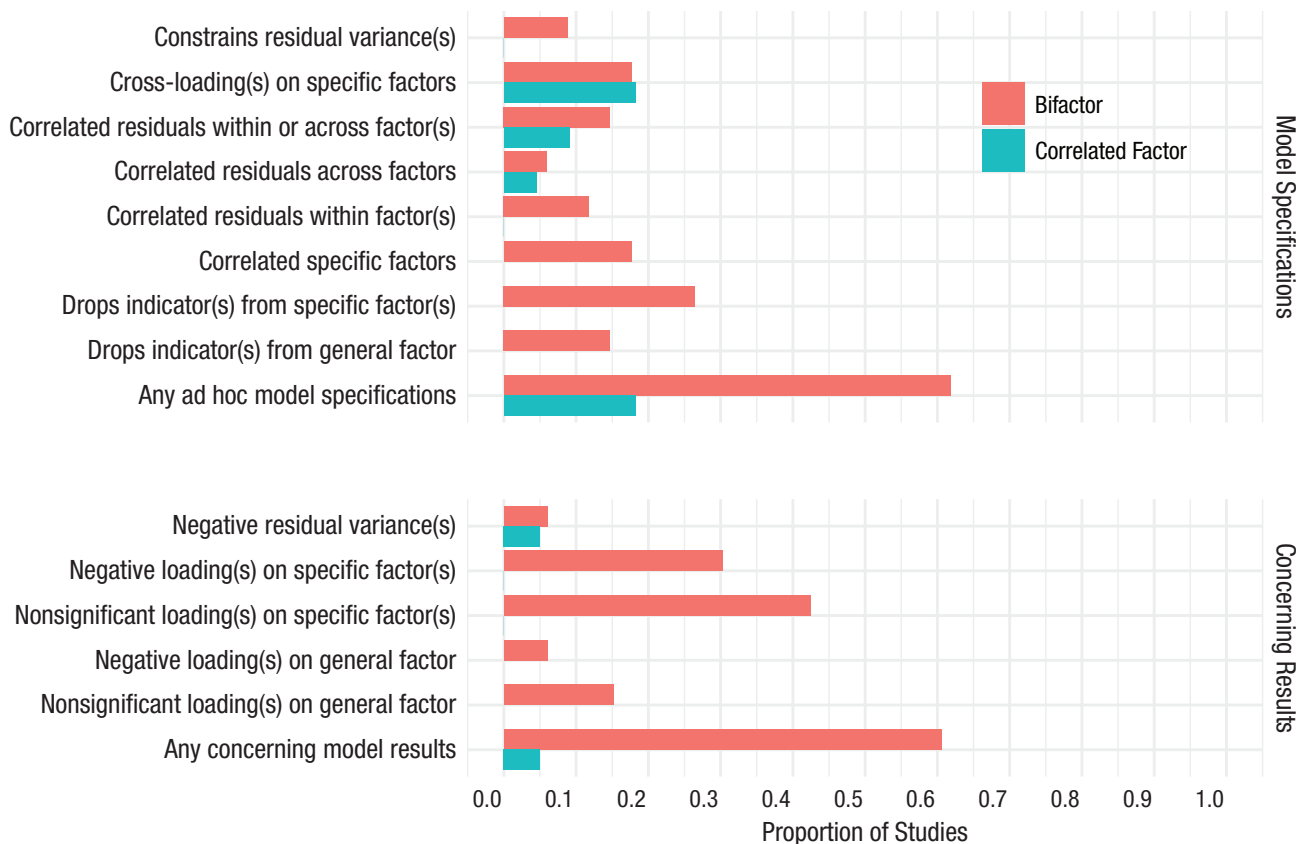


Fig. 2. Ad hoc model specifications and concerning results in best-fitting bifactor and correlated-factors models. Out of all studies that used confirmatory factor analysis, 34 modeled both a bifactor and correlated-factors model. Thirty-four studies provided sufficient information to assess ad hoc model specifications in bifactor models, and 33 studies reported sufficient results to assess concerning results in bifactor models. Twenty-two studies provided sufficient information to assess ad hoc model specifications in correlated-factor models, and 20 studies reported sufficient results to assess concerning results in correlated-factor models. Twelve studies provided requisite information to assess concerning model specifications in bifactor models but not correlated-factors models. Fourteen studies provided the requisite information to assess concerning results in bifactor models but not correlated-factors models.

researchers may often resort to such ad hoc model specifications, thus increasing the likelihood of obtaining chance findings that will not replicate. This highlights the critical importance of preregistration of one’s data analyses, in particular the details of a principled approach to model-fit improvement.

Advantages and shortcomings of model-based reliability and alternative indices for adjudicating among models

In growing awareness of problems with the overreliance on fit indices for adjudicating among structural models of psychopathology, such as overfitting (Bonifay & Cai, 2017; Bonifay et al., 2017; Preacher, 2006), researchers have begun to use (Bornovalova et al., 2020; Martel et al., 2017; Watts et al., 2019) and suggest (Forbes, Greene et al., 2021; Waldman, 2017) augmenting model fit with various model-based reliability indices (e.g., H , ω_H) first proposed in the psychometric

literature (McDonald, 1985, 1999; Reise, 2012; Rodriguez et al., 2016; Zinbarg et al., 2005). As shown in their formulas (Rodriguez et al., 2016; Zinbarg et al., 2005), these indices are driven not only by the magnitude of their indicators’ factor loadings but also by the number of factor indicators. For example, as the factor-loading magnitudes and the number of indicators increase, H approaches 1. These indices have begun to play a useful role in evaluating alternative structural models of psychopathology, as reflected by their increasing use in the literature, and have seen particular application in interpreting the results of bifactor models, especially in assessments of the reliability of the general and specific (i.e., group) factors in such models both within and across studies (Forbes, Greene et al., 2021; Martel et al., 2017; Watts et al., 2019). Use of these indices is meant to put the reliability of different factors in a model—either within a study or across studies—on an equal footing and to assess their usefulness in applied research. For example, an arbitrary threshold value of

$H \geq .7$ has been recommended for interpreting a factor as having adequate construct replicability (Rodriguez et al., 2016). In Figure 3, we show notched-box-and-whiskers plots of H and the median, standard deviation, and standard error of standardized factor loadings for the six most commonly characterized psychopathology factors (i.e., general, Externalizing, Internalizing, Distress, Fear, and Thought Disorder) from both bifactor and correlated-factors models in the 100 studies we reviewed. For all but the general factor, the values of these indices from the bifactor model are calculated for the specific (i.e., group) factors that accompany the general factor, in contrast to their unresidualized values from the correlated-factors model.

As Figure 3a shows, H was much higher for the factors in the correlated-factors model than for the specific factors in the bifactor model, $F(1, 265) = 216.23$, $p = 3.38 \times 10^{-36}$, partial $\eta^2 = .45$, suggesting that specific factors in bifactor models consistently explained less variance in their indicators than the factors in correlated-factors models. This is true in large part because the specific factors in the bifactor model are residuals in the sense that they explain the common variance in the indicators that is left over after the variance explained by the general factor. In addition, in the bifactor model, H was much higher for the general factor than for the specific factors, $F(5, 323) = 43.62$, $p = 2.62 \times 10^{-34}$, partial $\eta^2 = .40$. Across both models, H was highest for the general factor, followed by Externalizing and Internalizing, then by Fear, Thought Disorder, and Distress. This is likely due to the greater number of indicators used to specify common factors at higher than lower levels of generality. In addition to the values of H being much higher for factors in the correlated-factors model than for the corresponding specific factors in the bifactor model, differences in H across the factors in the correlated-factors model were nonsignificant and much smaller, $F(4, 125) = 1.40$, $p = .237$, partial $\eta^2 = .04$, than those in the bifactor model. These results suggest that the bifactor model provides substantial reliability in operationalizing a general factor but performs inferiorly at the level of specific factors.

Different pictures emerged for differences across factors for the median and standard deviation of standardized factor loadings, shown in Figures 3b and 3c, respectively. First, it is noteworthy that the magnitude of the median loadings was considerably lower than the magnitude of H . Similar to the findings for H , the median loadings were much higher for factors in the correlated-factors model than for the specific factors in the bifactor model, $F(1, 1230) = 190.16$, $p = 2.49 \times 10^{-40}$, partial $\eta^2 = .13$. Although the median loadings differed substantially and significantly across the six factors in the bifactor model, $F(5, 198) = 17.45$, $p = 2.67 \times 10^{-14}$,

partial $\eta^2 = .31$, differences in the median loadings among the factors in the correlated-factors model were much smaller and nonsignificant, $F(4, 125) = 2.34$, $p = .058$, partial $\eta^2 = .07$. In other words, correlated-factors models tended to result in consistently high loadings of their indicators across the dimensions that were modeled, whereas specific factors in the bifactor model tended to have weaker and less consistent loadings that were less interpretable. The pattern of these differences also was quite different from that for H , as the median loadings for factors in the bifactor model were highest for Externalizing, followed by the general factor, Fear, Internalizing, Distress, and Thought Disorder, and the pattern of these differences across factors in the correlated-factors model was quite different from that in the bifactor model. These findings likely reflect the different number of indicators per dimension and its influence on H but not on the median factor loadings.

The standard deviations of the factor loadings did not differ across the factors in either the bifactor model, $F(5, 198) = 1.04$, $p = .398$, partial $\eta^2 = .02$, or the correlated-factors model, $F(4, 125) = 2.29$, $p = .063$, partial $\eta^2 = .07$. Despite this, the loadings' standard deviations were significantly and substantially higher for factors in the bifactor than the correlated-factors model (.17, 95% CI = [.16, .18], and .11, 95% CI = [.09, .12], respectively), $F(1, 327) = 35.21$, $p = 7.53 \times 10^{-9}$, partial $\eta^2 = .10$. This indicates that loadings are more variable for factors in the bifactor than the correlated-factors model.

Another useful index for adjudicating among alternative factor models is the statistical property of efficiency, as instantiated using the standard errors of the factor loadings and factor correlations estimated within a given model. In addition to testing some hypotheses, a central goal of all statistical analyses is to estimate some quantities and to estimate them with greater than lesser precision. A model is useful to the extent that it facilitates this goal, and we can thus evaluate and adjudicate among alternative models of psychopathology partly on the basis of the extent to which their factor loadings and factor correlations are precisely estimated. In Figure 3d, the standard errors of the factor loadings are shown for the general and specific factors in the bifactor model and the factors in the correlated-factors models. There are several noteworthy features of this figure. First, factor loadings in the correlated-factors models are estimated quite precisely, as indicated by median standard errors that are quite low (.047) relative to their moderate to high factor loadings. Second, although loadings on the general factor are estimated almost as precisely (.056), they tended to show greater variability across studies. Third, loadings on the specific factors in the bifactor model are estimated much less precisely, as indicated by median standard errors for

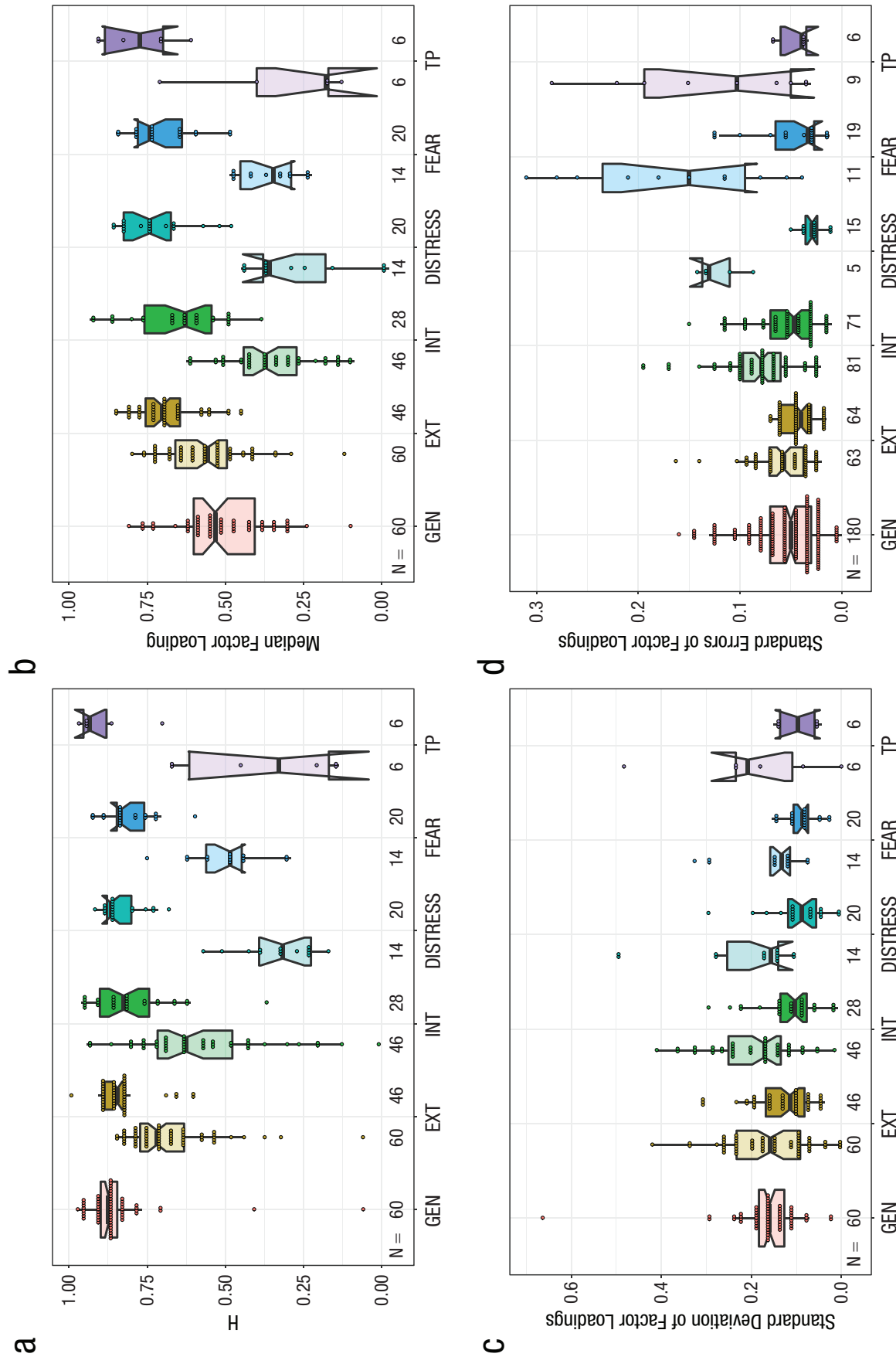


Fig. 3. Values of H (a) and the median (b), standard deviation (c), and standard errors (d) of standardized factor loadings from correlated-factors (darker hues) and bifactor (lighter hues) models, separately for each of the six most commonly characterized psychopathology factors. In each box-and-whisker plot, the horizontal line indicates the median, the upper and lower boundaries of the box indicate the interquartile range, and the whiskers mark values 1.5 times the interquartile range. Dots above or below the ends of the whiskers represent outliers. Lack of overlap in the notches in the boxes roughly corresponds to statistically significant differences among them (Tukey, 1977). N shows the number of studies that contributed to each plot. GEN = general, EXT = Externalizing, INT = Internalizing, TP = Thought Problems/Thought Disorders.

Table 1. *H* as a Function of Factor-Loading Magnitude (λ) and Number of Indicators

λ	Number of indicators																			
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
.33	.20	.27	.33	.38	.42	.46	.49	.52	.55	.57	.59	.61	.63	.65	.66	.68	.69	.70	.71	
.4	.28	.36	.43	.49	.53	.57	.60	.63	.66	.68	.70	.71	.73	.74	.75	.76	.77	.78	.79	
.6	.53	.63	.69	.74	.77	.80	.82	.84	.85	.86	.87	.88	.89	.89	.90	.91	.91	.91	.92	
.8	.78	.84	.88	.90	.91	.93	.93	.94	.95	.95	.96	.96	.96	.96	.97	.97	.97	.97	.97	

Note: Bolded numbers indicate the number of indicators necessary for equaling or surpassing the $H > .7$ threshold.

each specific factor that are almost twice as high as their counterparts in the correlated-factors models (.092, 95% CI = [.085, .100], and .047, 95% CI = [.040, .054], respectively), $F(1, 360) = 113.0$, $p = 3.81 \times 10^{-23}$, partial $\eta^2 = .24$. In addition, the precision with which factor loadings were estimated was more consistent across factors in the correlated-factors model, $F(4, 191) = 1.90$, $p = .111$, partial $\eta^2 = .04$, than across the specific factors in the bifactor model, $F(4, 169) = 18.9$, $p = 7.39 \times 10^{-13}$, partial $\eta^2 = .31$; similar to Bonifay & Cai, 2017; Bonifay et al., 2017).

Despite the potential utility of model-based reliability indices, there are some unforeseen shortcomings to their application for adjudicating among alternative structural models of psychopathology. For example, as mentioned above, H is dependent on both the magnitude of factor loadings within a factor and on the number of indicators used to represent a factor. Values of $H \geq .7$ for a factor can be achieved with factor loadings that range from .8 to .33 simply by increasing the number of factor indicators from 2 to 19 (see Table 1). This should come as no surprise, given that it has long been known that a test or scale can be made more reliable by increasing the number of items it contains (Nunnally & Bernstein, 1994). We find it problematic that two factors with such different properties (i.e., average factor loadings of .8 and .33) can be judged as having similar levels of construct replicability when one factor explains 64% of its indicators' variance on average and the other factor explains only 10% of its indicators' variance. Another view on this is shown in Figure S3 in the Supplemental Material, in which we compared values of H with the median factor loadings in the 100 studies we reviewed, binning by the number of indicators on each factor (i.e., < 5 , $5-10$, > 10). As the number of indicators increased, values of H increasingly exceeded the median loadings, and the correlation between H and the median loadings decreased. This demonstrates a decreased reliance of H on the magnitude of factor loadings versus the number of indicators as the latter increases. Given these findings, it might be better to rely on indices that are unaffected by the number of indicators, such as the median or mean,

standard deviation, and standard error of factor loadings within a factor, as suggested in Figures 3a to 3d, as the median loadings and their standard deviations show a clear superiority of factors with consistently high loadings regardless of their number of indicators.

To summarize the results shown in Figures 3a to 3d and Table 1, average factor loadings of indicators on the factors in correlated-factors models were higher, measured more precisely, and more consistent than for the factors in bifactor models. Although H for the general factor was as high as H for the factors in correlated-factors models, this was driven by the general factor's greater number of indicators, as the median factor loadings on the general factor was considerably lower than for the factors in the correlated-factors model. This suggests that researchers should use indices that are more sensitive to the percentage of variance that factors explain in their indicators, rather than the number of indicators on a factor. In addition, researchers should attend to the precision and consistency of factor loadings within the factors in a model in adjudicating among alternative structural models of psychopathology.

The consistency and sensitivity of factors to their constituent indicators

For a general factor of psychopathology to be considered truly general, the factor loadings of its indicators should be relatively consistent across the domains it covers. A general factor with large fluctuations in the magnitude of loadings across domains and studies is both quantitatively and practically meaningless in its interpretation. Although it is unrealistic to expect no variation in the average factor loadings across domains, such cross-domain variation in the factor loadings of indicators should be relatively small and ideally reflect only random fluctuations. In real-world applications, it is unrealistic to expect all indicators to be parallel (i.e., to have equal factor loadings and residual variances) or even tau equivalent (i.e., to have equal factor loadings) but rather for their factor loadings to be consistently moderate to high. In Figure 4a, using data from the 100 studies we reviewed, we show the distributions

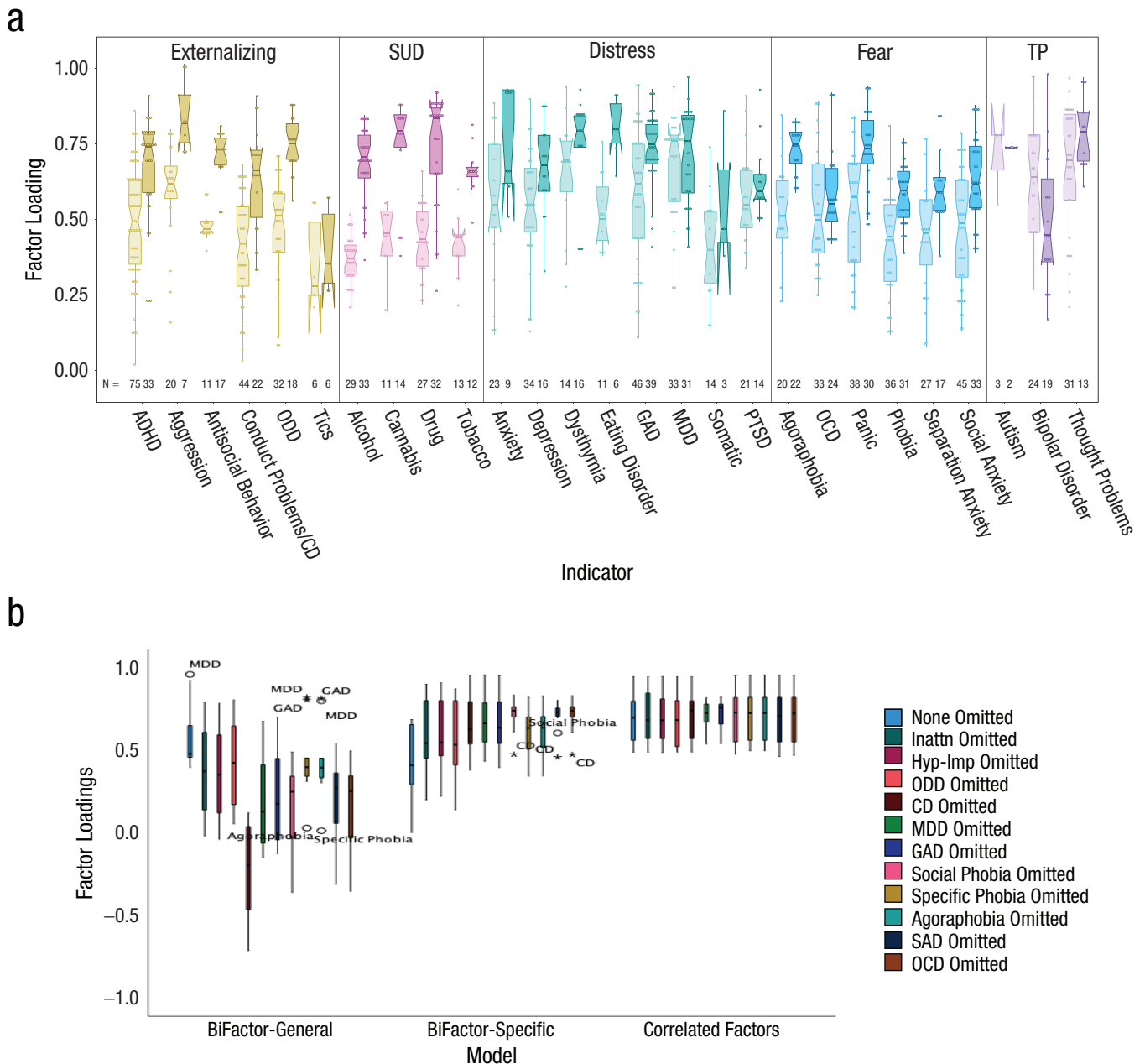


Fig. 4. Consistency and sensitivity of factor loadings for bifactor and correlated-factors models. In (a), standardized factor loadings are shown for symptom dimensions reflecting five commonly studied broad domains of psychopathology on the general factor from the bifactor model (the light boxes and whiskers) and on factors from the correlated-factors model (the dark boxes and whiskers). In (b), standardized factor loadings are shown for symptom dimensions for the general factor from the bifactor model, the Externalizing and Internalizing specific factors from the bifactor model, and the factors from the correlated factors model. In each box-and-whisker plot, the horizontal line indicates the median, the upper and lower boundaries of the box indicate the interquartile range, and the whiskers mark values 1.5 times the interquartile range. Dots or asterisks above or below the ends of the whiskers represent outliers. Lack of overlap in the notches in the boxes roughly corresponds to statistically significant differences among them (Tukey, 1977). *N* shows the number of studies that contributed to each plot. SUD = substance use disorder, TP = thought problems/thought disorders, ADHD = attention-deficit/hyperactivity disorder, CD = conduct disorder, ODD = oppositional defiant disorder, GAD = generalized anxiety disorder, MDD = major depressive disorder, Somatic = somatic complaints, PTSD = posttraumatic stress disorder, OCD = obsessive-compulsive disorder, Inattn = inattention, HYP-IMP = hyperactivity-impulsivity, SAD = separation anxiety disorder.

of standardized factor loadings of symptom dimensions reflecting five commonly studied broad domains of psychopathology on the general factor from the bifactor

model (the light boxes and whiskers) and on factors from the correlated-factors model (the dark boxes and whiskers). There is substantial cross-domain variation in

the general factor loadings, $F(4, 716) = 25.7, p = 6.32 \times 10^{-20}$, partial $\eta^2 = .13$, whereas this variation is much smaller for loadings in the correlated-factors model, $F(4, 514) = 5.5, p = .00025$, partial $\eta^2 = .04$. Factor loadings were also higher on the correlated factors than on the general factor in the bifactor model (.68, 95% CI = [.66, .70], and .51, 95% CI = [.50, .52], respectively), $F(1, 1238) = 267.1, p = 1.62 \times 10^{-54}$, partial $\eta^2 = .18$.

Another criterion that may be useful for adjudicating among structural models of psychopathology is how sensitive or robust a factor is to the inclusion or exclusion of its indicators (Reise, 2012). The optimal case for the validity of a factor is that the loadings of its indicators should be relatively consistent and moderate to high in magnitude (K. Bollen, 2011; K. A. Bollen, 2020; Fabrigar et al., 1999; Reise, 2012; Savalei & Reise, 2019; Yang & Green, 2010). In our reading of the literature, however, this is often not the case. We have shown one view of this issue in Figures 3a to 3d, namely, calculating the median and standard deviation of factor loadings. Another perspective on this, given a sufficient number of indicators, is to reexamine the median and variability of factor loadings on a factor when one removes each of the indicators in turn. We illustrate this below using data from the TTS described in the Method section (Waldman et al., 2016). We present factor loadings for each symptom dimension on the general and Externalizing and Internalizing specific factors from a bifactor model and from the Externalizing and Internalizing factors from a correlated-factors model when each symptom dimension is omitted in turn from the CFA. As shown in Figure 4b, variability in the magnitude and spread of the loadings was greatest for the general factor from the bifactor model, intermediate for the Externalizing and Internalizing specific factors from the bifactor model, and minimal for the Externalizing and Internalizing factors from the correlated-factors model. These results echo those presented in Figure 3a, in which H was higher and more consistent for factors in correlated-factors models than for the specific factors in bifactor models.

Adjuncts to CFA for adjudicating among alternative models—advantages and limitations

Above, we emphasized an approach to adjudicating among alternative structural models of psychopathology that relies heavily on CFA. Researchers have recently suggested two other approaches for adjudicating among alternative structural models of psychopathology that can be used as adjuncts to—or replacements for—the exclusive use of CFA. These include the reliance on associations of latent psychopathology dimensions with external criteria (Bonifay et al.,

2017) and the use of ESEMs or exploratory factor analysis (EFA) as a complement to CFA (Greene et al., 2022). In the following two sections, we will explore and discuss the advantages and limitations of these two approaches.

Patterns and magnitude of associations with relevant criterion variables across models

Some researchers have suggested that although model fit may not be particularly useful for adjudicating among alternative models of psychopathology, meaningful differences among alternative models will be evident in the variance explained in, and patterns of associations with, relevant criterion variables (Bonifay et al., 2017; Ferrando & Lorenzo-Seva, 2019; Forbes, Greene, et al., 2021; Watts et al., 2019). Such assertions are especially common in support of the general factor of psychopathology. We examined this assertion in the TTS data set described in the Method section (Waldman et al., 2016), contrasting the variance explained in and the patterns of association with relevant criterion variables (Figs. 5a and 5b; see also Figs. S4a and S4b in the Supplemental Material). As Figure 5a and Figure S4a show, the bifactor model containing general, Externalizing, and Internalizing factors explained a virtually identical amount of variance in outcomes as the correlated Externalizing and Internalizing factors alone. Also, as Figure 5b and Figure S4b show, the general factor had a nearly identical pattern of associations with the outcomes as the Externalizing and (to a lesser extent) Internalizing factors. A very similar pattern of findings emerged from another study (see Fig. 2 in Watts et al., 2019). These results fail to justify the incremental value of including a general factor over and above the correlated Externalizing and Internalizing factors alone. Given that these results are from just two studies, it is important for researchers to examine whether similar results will emerge in their studies and from the literature more generally. This will be difficult, however, because researchers reported external validity analyses from alternative models in only 29% of the studies we reviewed. It also is worth noting that associations with external criteria have often been misused in the bifactor literature to support the substantiveness of the p factor by contrasting the magnitude of relations with external correlates of the p factor versus the specific factors, which is not a fair comparison given the diminished model-based reliability of the specific factors, as shown in Figure 3a.

Furthermore, it is important to recognize that such comparisons in variance explained cannot be made using a higher-order general factor, as a model in which external variables are regressed on the higher-order

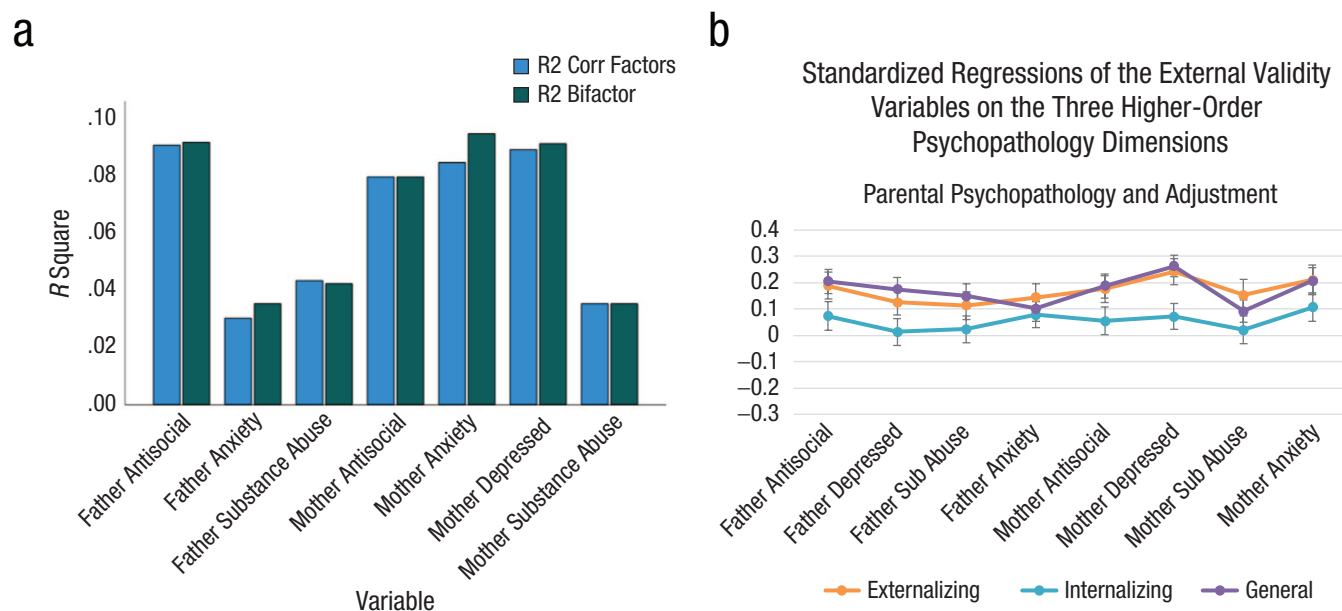


Fig. 5. Magnitude and patterns of associations with outcomes in bifactor and correlated-factors models, with (a) showing the percent of variance (R-square) explained in each of 7 criterion variables by the bifactor model (in green) and the correlated factors model (in blue) and (b) showing the standardized regression coefficient (Beta) and its 95% confidence interval for predicting 8 criterion variables from the Externalizing factor (in orange), Internalizing factor (in teal), and general factor (in gray).

general factor and lower-order factors simultaneously is unidentified. Given certain model constraints, only the bifactor structure allows one to separately and simultaneously examine the unique and shared variance associated with outcomes between the general and specific factors. Although this property is a desirable feature of bifactor models in principle, it does not guarantee that inclusion of a general factor will explain additional variance in, or show a different pattern of associations with, causes or outcomes over and above the factors in a correlated-factors model.

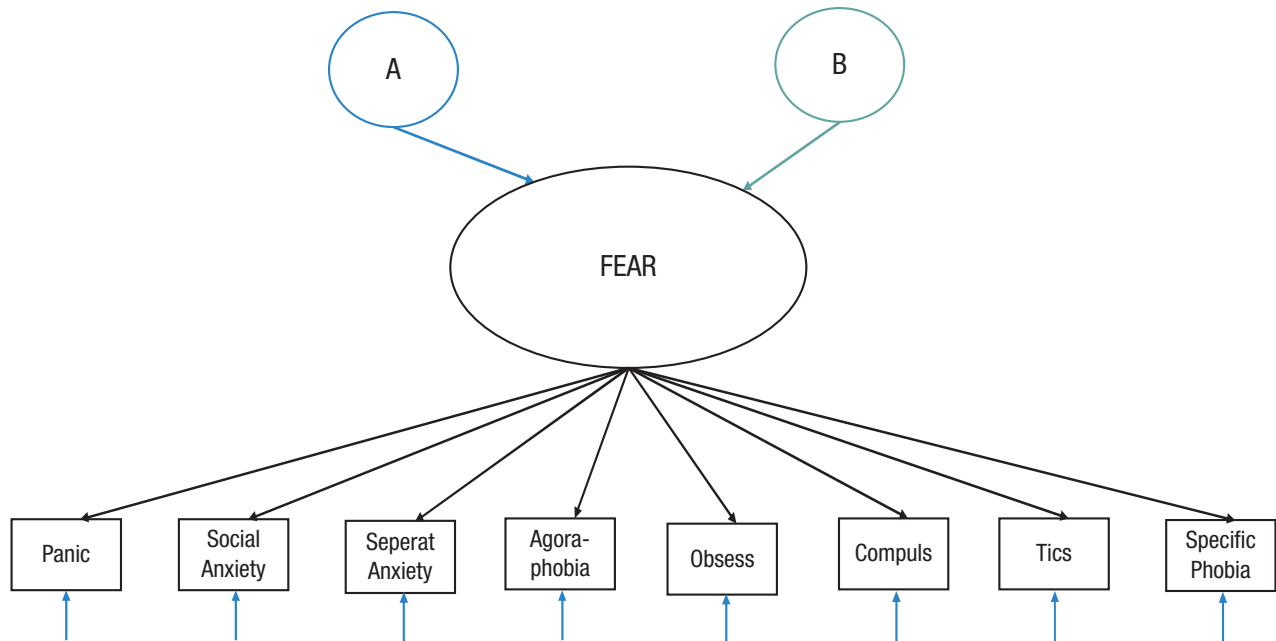
One way that tests of alternative structural models of psychopathology can be made more rigorous is by formally contrasting their associations with causes or outcomes. As an example, one can contrast the relations of the factors and their indicators with causes or outcomes under two alternative models that are commonly used in the multivariate behavior genetics literature (Neale & Cardon, 2013) but have rarely been used in the literature on the structure of psychopathology (but see Conway et al., 2022, and Forbes et al., 2020, for a somewhat similar model comparison). These models are the *common-* and *independent-pathway* models (Neale & Cardon, 2013), illustrated in Figures 6a and 6b, respectively. In the common-pathway model, associations of the variables A and B (here representing causes but which may also represent outcomes) with the symptom dimensions are mediated by the Fear factor, whereas in the independent-pathway model,

associations of the variables A and B with the symptom dimensions are direct and unmediated by the Fear factor. Comparison of these two models is tantamount to testing whether associations of the causes or outcomes with the symptom dimensions are reducible to associations of the variables A and B with the hypothesized latent factor or whether the symptom dimensions have meaningful associations with the causes or outcomes that are not captured by the hypothesized factor. A similar model comparison has been suggested in the context of genome-wide association studies (Grotzinger et al., 2022).

Greater awareness of statistically distinguishable versus indistinguishable models

Here, we relied on the bifactor model as a way of including a general psychopathology factor along with specific factors that parallel those in correlated-factors models. An alternative operationalization of a general psychopathology factor is via a higher-order model in which the general factor accounts for the shared variance among the second-order dimensions (e.g., Internalizing, Externalizing, Thought Disorder). Although there are important substantive distinctions between the interpretation and parameterization of bifactor and higher-order models, there are several challenges to distinguishing them on the basis of fit indices. First, a

a



b

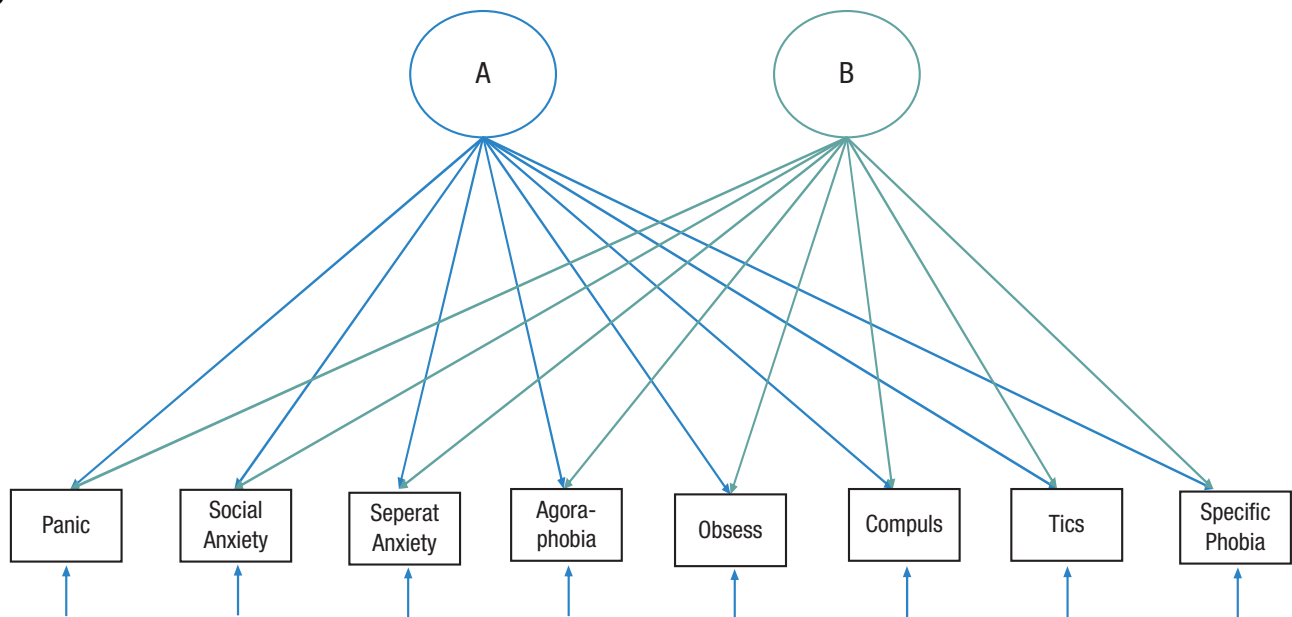


Fig. 6. Common-pathway (a) and independent-pathway (b) models for the structure of psychopathology.

higher-order model requires more than three indicators (i.e., lower-order dimensions) in order to be overidentified and thus testable against the correlated-factors model that is its logical alternative (Loehlin & Beaujean, 2016). Second, even under seemingly favorable conditions in which there are four or more indicators, the fit of the bifactor and higher-order models is often identical or nearly so (Gignac, 2008; Markon, 2019; Mulaik

& Quartetti, 1997; Yung et al., 1999). Given these issues, it has recently been suggested that researchers use other criteria for adjudicating between these alternative models containing a general as well as specific factors (Forbes, Greene, et al., 2021; Markon, 2019).

Given increasing concerns with the ability of CFA to definitively adjudicate among competing models, there has been a resurgence of interest in more exploratory

Table 2. Distinguishable and Indistinguishable Models: Contrasting Fits of Confirmatory Factor Analysis (CFA) Models and Exploratory Structural Equation Models (ESEMs)

Model	χ^2	<i>df</i>	TLI	RMSEA	SRMR	BIC	$r_{\text{EXT-INT}}$ OR $r_{\text{EXT-Distress}}$
CFA							
Full model (GAD and MDD on general factor only)	424	34	.96	.06	.03	29,228	-.13 [-.19, -.07]
Three oblique factors (fear, distress, externalizing)	1,032	41	.91	.09	.06	29,959	.58 [.54, .62]
Three orthogonal factors (fear, distress, externalizing)	2,849	43	.78	.14	.22	31,874	
ESEM							
Bifactor with two specific factors	333	25	.95	.06	.02	29,169	-.16 [-.42, .11]
Three oblique factors	333	25	.95	.06	.02	29,169	.57 [.51, .63]
Three orthogonal factors	333	25	.95	.06	.02	29,169	

Note: In the correlation column, values in brackets indicate 95% confidence intervals. TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; BIC = Bayesian information criterion; EXT = Externalizing; INT = Internalizing; GAD = generalized anxiety disorder; MDD = major depressive disorder.

approaches for investigating the structure of psychopathology. These have included EFA (Greene et al., 2022; Murray et al., 2019; Ringwald et al., 2023) and its variants, such as exploratory bifactor analysis (Greene et al., 2022; Jennrich & Bentler, 2011, 2012; Lorenzo-Seva & Ferrando, 2019; Mansolf & Reise, 2016; Markon, 2019; Pezzoli et al., 2017; Ringwald et al., 2019; Sellbom et al., 2015; Sharp et al., 2015), the “bass-ackwards” approach (Goldberg, 2006; Kim & Eaton, 2015; Levin-Aspensson et al., 2019), and ESEMs (Asparouhov & Muthén, 2009; Marsh et al., 2014; Wright & Simms, 2015). Although this shift may end up paying dividends over the undue reliance on CFA, this is as yet an open question. One relevant issue that has received insufficient attention, however, is that many of these exploratory models are statistically indistinguishable from each other despite the fact that they are substantively very different (Ringwald et al., 2019), similar to the distinction between the bifactor and higher-order models above. This is illustrated in Table 2, in which we present reanalyses of previously published data from the TTS. We show the fit of three CFA models and three ESEMs, all of which are conceptually quite different. Despite the substantive differences among the models, the three alternative CFA models are distinguishable by their fit statistics, whereas the three ESEMs are completely indistinguishable, notwithstanding the dramatic differences in their substantive interpretations. Although this issue of indistinguishable fit in EFA has long been known in the technical statistical literature, it is often ignored in applied studies of the structure of psychopathology. Thus, although augmenting CFAs with more exploratory methods—especially in a sequential fashion in which EFA methods are used as a sensitivity check to investigate sources of covariance missed by CFAs (Greene

et al., 2022)—is an exciting direction for further exploration, authors conducting applied research need to be more cognizant of distinguishable versus indistinguishable models and thus more cautious in their application.

Discussion

Conclusions and future directions

There are several conclusions that may be drawn from the analyses and results presented here. In addition, on the basis of these results, we have several suggestions for changes that can lead to more consistent, replicable, and comprehensive models of the underlying structure of psychopathology. First, given increased recognition that fit indices are useful in discriminating among some types of models but not others, researchers need to be mindful of these contextual differences in their adjudication of alternative models. Specifically, although conventional fit indices appear to perform well at discriminating among various correlated-factors models and models containing only a single general factor—even in cases in which those alternative models are quite similar in their fit to the data—conventional fit indices are susceptible to and cannot detect overfitting in bifactor models, as we and others have shown (Bonifay & Cai, 2017; Bonifay et al., 2017; Greene et al., 2019; Watts et al., 2019).

Second, as a corollary to the previous point, researchers need to be wary of using ad hoc or post hoc model modifications to improve model fit, especially the fit of their hypothesized best-fitting model, as chasing model fit is most likely to result in models of the structure of psychopathology that do not

replicate across studies or factor analytic methods (i.e., exploratory vs. confirmatory).

Third, researchers should pay greater attention to various model characteristics—such as the magnitude, precision, and consistency of factor loadings and factor correlations—in evaluating alternative structural models of psychopathology. Here, we showed that factor loadings and factor correlations were estimated more consistently, more precisely, and with less bias in correlated-factors than in bifactor models. In addition, the factors in correlated-factors models demonstrated greater parameter invariance, as they were less sensitive to the inclusion or exclusion of any particular indicator than the specific or general factors in bifactor models. Along these lines, the magnitude of factor loadings on the general factor in bifactor models showed considerable variability across the major psychopathology domains and their constituent factors. This relatively weak level of indicator invariance for factors in the bifactor model translates into weaker support for the construct validity and reliability of a general psychopathology construct (Reise, 2012) and supports the notion that specific factors in bifactor models may be untrustworthy as measures of narrow constructs (Kelley & Pornprasertmanit, 2016).

Fourth, although there are good reasons to augment CFAs with exploratory modeling methods (such as ESEMs), given the overreliance on the former (Greene et al., 2022), it is important to recognize the fact that alternative EFA models that are quite different substantively will show identical fit to the data so long as they include the same number of factors. In addition, although replication across samples is always important, this is true to an even greater extent for the findings from EFAs and ESEMs, given their exploratory nature.

Fifth, researchers need to conduct more rigorous tests of the associations of their hypothesized best models and alternative models with external criteria than are currently practiced before declaring victory for the superiority of their hypothesized model. We have illustrated this here by borrowing the concept of common- versus independent-pathway models from the quantitative genetics literature (Neale & Cardon, 2013; see also Forbes et al., 2020; Grotzinger et al., 2022).

Sixth, an extension of the previous point is that researchers need to more systematically contrast the external validity of alternative models to test for differences in the explanatory power of their hypothesized best-fitting model over that of alternative models. We and others (Watts et al., 2019) illustrated this by demonstrating that a bifactor model with two specific Externalizing and Internalizing factors explained no more variance in a set of relevant outcomes than a model

with only the two correlated factors. This is akin to the well-known situation in multiple regression in which the variance explained in an outcome by several predictors is decomposed into the variance that is shared among the predictors and the variance that is unique to each predictor. Although inclusion of a general factor in a bifactor model can be useful pragmatically by capturing this common variance, it may often give the illusion that one is gaining something incremental over the correlated factors, both statistically and substantively, which would be misleading (Fried et al., 2021).

Seventh, although we did not have space to examine this issue here, more attention needs to be paid to the appropriate and optimal levels of granularity in the selection of factor indicators in structural models of psychopathology. Diagnoses, symptom dimensions, and individual symptoms have all been used as indicators of higher-order psychopathology dimensions, and each has its advantages and disadvantages. For example, diagnoses are available for very large samples (e.g., $\geq 35,000$ in the National Epidemiologic Survey on Alcohol and Related Conditions [NESARC]; Forbes, Greene, et al., 2021; Lahey et al., 2012), but despite the increased statistical power given such a large sample, the factor loadings from some models fitted to NESARC data are quite imprecise, as reflected by large standard errors, and lead to some nonsignificant factor loadings (e.g., Lahey et al., 2012). Many studies have used symptom dimensions as indicators, which can be advantageous because they provide greater information than diagnoses (Faure & Forbes, 2021; Markon, 2010; Markon et al., 2011; van der Sluis et al., 2013; Waszczuk et al., 2020; Wright et al., 2013; Wright & Simms, 2015) but have the disadvantage of often being severely nonnormally distributed and highly skewed and kurtotic. Finally, individual symptoms are the most granular indicators in relatively common use and have the advantages that they better allow one to build structural models “from the ground up” (Forbes, Sunderland, et al., 2021), and building latent-variable models using them can better account for measurement error. A recent study has also shown that in the context of alcohol use disorder, even individual symptoms may be insufficiently granular and lead to spurious evidence for unidimensionality if too few symptoms are used (Watts et al., 2021). In addition, results of a recent study (Forbes, Sunderland, et al., 2021) suggest that symptom-level homogeneity likely inflates the similarity and consequent covariation of some *DSM-5* disorders and thus represents a potential source of bias in studies analyzing their patterns of covariation.

Eighth, authors conducting applied research should strongly consider integrating simulations with their analyses of real data to gain a better understanding of

which models can be successfully discriminated from each other and which cannot and what model features (e.g., correlated residuals; Greene et al., 2019) might lead to spurious evidence in favor of their proposed model (McNeish & Wolf, 2021). It is fair to say that, despite their utility, simulations are considerably underused in the study of the structure of psychopathology and that the field would benefit from their increased use. This extends to assessments not only of model fit, as used here and elsewhere (Bonifay & Cai, 2017; Preacher, 2006), but also of parameter bias and imprecision.

Ninth, although it may seem rather prosaic, researchers should both examine and provide readers much more detailed results from their studies of the structure of psychopathology than is currently the norm. In our search through 100 articles for this review, we were dismayed at the low rates of reporting of details crucial for adjudicating among alternative structural models of psychopathology. These included factor loadings and standard errors from the best-fitting model (91% and 18%, respectively), factor loadings and standard errors from multiple alternative models (52% and 9%, respectively), and relations of external criteria with factors in alternative models as well as in the best-fitting model (29% of studies that tested multiple models). Given the wide availability of Supplemental Material for most journals, researchers are no longer limited in their reporting of such information as they were in the past.

Tenth, and finally, researchers should test a greater number of alternative models, broadening their evaluation to models that supplement their hypothesized best-fitting model (or models) to avoid confirmatory biases (Fudge, 2014; Platt, 1964). As shown in Figures S5a and S5b in the Supplemental Material, in our review of 100 studies, we found that researchers tested relatively few alternative models ($M = 4$, $SD = 3$) and that the number of alternative models tested declined somewhat from 1999 to 2021 (estimates from 4.3 to just over 3.5). In addition to increasing the likelihood of confirmatory bias, testing few models ignores the fact that there may be a set of fungible models with indistinguishable fit (MacCallum et al., 1993; Raykov & Penev, 1999), some of which may end up being better contenders given replication and criteria other than model fit (e.g., relations with criterion variables, utility). Thus, a better analytic strategy might be to test a fuller set of models and select models that the data can more definitively rule out than to “pick a winner.” Rather than trying to decide on the best model, it might be more realistic and useful to say that several models are consistent with the data and await adjudication by further research, whereas other models can be more reliably eliminated (e.g., Kim & Eaton, 2015). Increasing the number of models tested can also aid in examining

replicability across studies. Relatedly, researchers should rely less on fit indices based on null-hypothesis significance tests (e.g., χ^2 difference tests of exact fit, RMSEA, BIC, comparative fit index, Tucker-Lewis index), given that these scale with sample size and often devolve to context-dependent rules of thumb (Greene et al., 2022; Marsh et al., 2004; McNeish & Wolf, 2021) and endeavor to represent the magnitude of differences in fit among alternative models.

Like all recommendations, the ones proposed here have important caveats. As one example, large and consistent factor loadings may result from selecting items that are highly similar to one another, a psychometrically undesirable strategy. Fortunately, one can guard against this using item-response-theory methods to ensure that factor indicators provide information and reliable measurement across the intended range of the latent psychopathology dimension. This illustrates the fact that although the proposed indices are useful, they are not the only considerations in evaluating the reliability and validity of models of psychopathology.

Limitations

There are several limitations to the current study. First, we did not consider the use of multi-informant data, which may be problematic, especially given its importance in studies of youth psychopathology, nor did we consider how various problems with alternative operationalizations of psychopathology indicators (e.g., the use of diagnoses, symptom dimensions, or individual symptoms) might vary systematically by sample characteristics (e.g., age, sex, ancestry).

Second, in our attempt to provide general guidelines for methods and indices for adjudicating among alternative structural models of psychopathology, we inevitably faced problems with incomplete and inconsistent coverage of psychopathology across studies, which was exacerbated by the differential developmental relevance of psychopathological conditions and constructs across studies.

Third, the simulations we conducted had certain characteristics that may limit their generalizability. These include the use of symptom dimensions as indicators and the modeling of these as normally distributed; the use of only a single large sample size; the use of only two indicators on the Distress factor, which limits the possible models that are identifiable (Loehlin & Beaujean, 2016); and the use of only two sets of true parameter values that were drawn from a single published study (Watts et al., 2019). More extensive simulation studies using alternative true parameter values from alternative best-fitting models using symptoms, symptom dimensions, and diagnoses as indicators at a

variety of plausible sample sizes are needed. These will yield a better understanding of the role of each of these factors in model nonconvergence, model fit, bias in the percentage of variance explained in the indicators, and bias and imprecision in estimating factor loadings and factor correlations. These simulations should also yield a clearer picture of the variables involved in adjudicating among alternative structural models of psychopathology and, in distinction to the results presented here, might reveal scenarios in which bifactor models are disadvantaged despite being the true generating model.

Fourth, we focused here on alternative models for the *structure* of psychopathology. Thus, we did not consider alternative models of psychopathology, such as network approaches (Borsboom & Cramer, 2013; Borsboom et al., 2018; McNally, 2016, 2021; Robinaugh et al., 2020), despite their popularity and increased use (but see Forbes et al., 2017; Forbes, Wright, et al., 2021).

Fifth, similar to many studies in the field, those we reviewed used participants of predominantly European ancestry. Although we believe that the proposed methods and indices for adjudicating among alternative structural models of psychopathology are equally applicable to individuals from all demographic backgrounds, this is a hypothesis that should be evaluated in subsequent research. As a specific example, formal tests of measurement invariance can be leveraged to elucidate similarities and differences in the structure of psychopathology and the validity of the measures thereof across various groups, including sex and ancestry.

Implications for modeling the structure of psychopathology

Deciding among rival models of psychopathology is integral to many areas of psychopathology research. How can we hope to find the underlying genetic and environmental risk factors, neurobiological underpinnings, course and outcome, and most effective treatments for dimensions of psychopathology if we do not know how best to classify those dimensions? There are numerous unresolved issues in the structure of psychopathology that bear on this question. Here, we focused on CFA, but there are many other analytic methods (e.g., various forms of EFA, hierarchical clustering) that are potentially useful for elucidating the structure of psychopathology. Another issue is illuminating the “dark matter” of psychopathology, namely better understanding the placement in a hierarchical taxonomy of psychopathology of conditions the classification of which is unclear (e.g., ADHD, mania, neurodevelopmental disorders, obsessive-compulsive disorder, and dissociation). One reason for the uncertainty surrounding the classification of these conditions is that they are likely

multidimensional, a hypothesis that should be tested in future research. These conditions also reflect the balance between well-established and more provisional aspects of a hierarchical taxonomy such as HiTOP, and analytic methods designed for explicitly investigating this balance (e.g., Procrustes or target rotations; Browne et al., 2002; Zhang et al., 2019) may be particularly useful in clarifying the placement of such conditions. Finally, many contemporary models of psychopathology, such as HiTOP, are hierarchical, with lower-level dimensions of psychopathology nested within higher-order dimensions of greater generality. This highlights the importance of determining the relevance of different levels of the hierarchy for different purposes (e.g., etiology, utility), as well as for research on psychopathology more generally.

Another application of the methods and indices proposed here is to investigate the genetic (and environmental) etiology of psychopathology. Valid classification is integral to finding genes and biological pathways that underlie both higher- and lower-order psychopathology dimensions. Although several studies have used novel analytic methods to examine the structure of psychopathology at the genomic level (Grotzinger et al., 2022; Lee et al., 2021; Waldman et al., 2020), each of these studies has found a different higher-order dimensional structure of psychopathology using largely the same data sets. In addition, although none of these studies found evidence for a general psychopathology factor at the genomic level, multiple studies have reported single-nucleotide polymorphism (SNP)-based heritabilities for such a general factor (Neumann et al., 2016; Riglin et al., 2020). It is imperative to better establish the higher-order phenotypic and genetic dimensional structure of psychopathology in order to find the genes and biological pathways underlying these dimensions, as well as their SNP-based heritabilities and genetic correlations with relevant variables.

An overarching theme of this article is that alternative structural models of psychopathology are testable and subject to revision rather than set in stone. For example, an important and often unappreciated feature of the HiTOP framework (Haeffel et al., 2022) is that it is a dynamic entity, one subject to revision in light of new evidence relevant to the classification of psychopathology (DeYoung et al., 2022). The methods and indices described here should facilitate this effort by helping improve studies of the structure of psychopathology that will form the basis of such proposals for revision.

In sum, we recommend that when adjudicating among alternative structural models, psychopathology researchers should supplement the use of fit indices by examining the median or mean, standard deviation, and

standard errors of factor loadings for each of the factors within each of the models fit. As a practical matter, it may make sense to also average these indices across the factors in the model and then contrast the averages across the various models examined. It also would be advantageous to examine the sensitivity of the factor loadings on each factor to the inclusion or exclusion of each of its indicators and to report associations with relevant causes or outcomes not only for the best-fitting model but also for alternative models. Using these criteria to augment conventional fit indices and having greater awareness of the fit propensity of alternative models should help increase the validity and replicability of such models and advance progress toward a consensus model of the structure of psychopathology. To return to where we started this article, although Scott Lilienfeld's research interests and publications branched out far and wide beyond his early work on classification and comorbidity, we like to think that he would approve of our suggestions here and view them as steps toward constructing more valid and replicable models of psychopathology.

Transparency

Action Editor: William O'Donohue

Editor: Jennifer L. Tackett

Author Contributions

Irwin D. Waldman: Conceptualization; Data curation; Formal analysis; Methodology; Project administration; Resources; Software; Supervision; Visualization; Writing – original draft; Writing – review & editing.

Christopher D. King: Conceptualization; Data curation; Formal analysis; Resources; Supervision; Visualization; Writing – original draft; Writing – review & editing.

Holly E. Poore: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Supervision; Visualization; Writing – original draft; Writing – review & editing.

Justin M. Lunningham: Conceptualization; Methodology; Writing – review & editing.

Richard M. Zinbarg: Conceptualization; Methodology; Writing – review & editing.

Robert F. Krueger: Conceptualization; Writing – review & editing.

Kristian E. Markon: Conceptualization; Writing – review & editing.

Marina Bornovalova: Conceptualization; Writing – review & editing.

Michael Chmielewski: Conceptualization; Writing – review & editing.

Christopher Conway: Conceptualization; Writing – review & editing.

Michael Dretsch: Conceptualization; Writing – review & editing.

Nicholas R. Eaton: Conceptualization; Writing – review & editing.

Miriam K. Forbes: Conceptualization; Writing – review & editing.

Kelsie Forbush: Conceptualization; Writing – review & editing.

Kristin Naragon-Gainey: Conceptualization; Writing – review & editing.

Ashley Lauren Greene: Conceptualization; Writing – review & editing.

J. D. Haltigan: Conceptualization; Writing – review & editing.

Masha Ivanova: Conceptualization; Writing – review & editing.

Keanan Joyner: Conceptualization; Writing – review & editing.

Katherine M. Keyes: Conceptualization; Writing – review & editing.

Kevin M. King: Conceptualization; Writing – review & editing.

Roman Kotov: Conceptualization; Writing – review & editing.

Holly Levin-Aspenson: Conceptualization; Writing – review & editing.

Thomas Olinio: Conceptualization; Writing – review & editing.

Jason A. Oliver: Conceptualization; Writing – review & editing.

Christopher J. Patrick: Conceptualization; Writing – review & editing.

David Preece: Conceptualization; Writing – review & editing.

Lauren A. Rutter: Conceptualization; Writing – review & editing.

Martin Sellbom: Conceptualization; Writing – review & editing.

Susan South: Conceptualization; Writing – review & editing.

Nicholas J. Wagner: Conceptualization; Writing – review & editing.

Ashley L. Watts: Conceptualization; visualization; Writing – review & editing.

Sylia Wilson: Conceptualization; Writing – review & editing.

Aidan G. C. Wright: Conceptualization; Writing – review & editing.


David Zald: Conceptualization; Methodology; Writing – review & editing.


Declaration of Conflicting Interests


The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.


ORCID iDs

Irwin D. Waldman  <https://orcid.org/0000-0002-6862-1837>











Richard M. Zinbarg  <https://orcid.org/0000-0002-2670-6330>

Kristian E. Markon  <https://orcid.org/0000-0002-7230-9694>

Christopher Conway  <https://orcid.org/0000-0002-6848-2638>

Miriam K. Forbes  <https://orcid.org/0000-0002-6954-3818>

Kelsie Forbush  <https://orcid.org/0000-0002-5900-4204>

Kristin Naragon-Gainey  <https://orcid.org/0000-0002-2245-6880>
 Keanan Joyner  <https://orcid.org/0000-0002-5759-0879>
 Kevin M. King  <https://orcid.org/0000-0001-8358-9946>
 Holly Levin-Aspenson  <https://orcid.org/0000-0002-3980-8299>
 Thomas Olino  <https://orcid.org/0000-0001-5139-8571>
 Christopher J. Patrick  <https://orcid.org/0000-0003-1794-1969>
 Susan South  <https://orcid.org/0000-0003-0341-7074>
 Ashley L. Watts  <https://orcid.org/0000-0002-8210-4748>
 Syla Wilson  <https://orcid.org/0000-0003-0144-524X>
 Aidan G. C. Wright  <https://orcid.org/0000-0002-2369-0601>

Acknowledgments

We thank Niels Waller for his very helpful comments on an earlier draft of this article. This article is dedicated to the memory of Scott Lilienfeld, whose 40-year friendship and 30-year collaboration enriched the first author's personal and professional life to an immeasurable extent. Scott's contributions to research, teaching, mentoring, and service in psychology are legendary. We miss the opportunity to have had Scott discuss, disagree, and kibitz with us over aspects of this article.

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/21677026221144256>

References

- Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs: General and Applied*, 80(7), 1–37. <https://doi.org/10.1037/h0093906>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16(3), 397–438. <https://doi.org/10.1080/10705510903008204>
- Barlow, D. H., Farchione, T. J., Bullis, J. R., Gallagher, M. W., Murray-Latin, H., Sauer-Zavala, S., Bentley, K. H., Thompson-Hollands, J., Conklin, L. R., Boswell, J. F., Ametaj, A., Carl, J. R., Boettcher, H. T., & Cassiello-Robbins, C. (2017). The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders. *JAMA Psychiatry*, 74(9), 875–884. <https://doi.org/10.1001/jama.psychiatry.2017.2164>
- Barlow, D. H., Farchione, T. J., Sauer-Zavala, S., Latin, H. M., Ellard, K. K., Bullis, J. R., Bentley, K. H., Boettcher, H. T., & Cassiello-Robbins, C. (2017). *Unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide* (2nd ed.). Oxford University Press.
- Biederman, J., Newcorn, J., & Sprich, S. (1991). Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders. *The American Journal of Psychiatry*, 148(5), 564–577. <https://doi.org/10.1176/ajp.148.5.564>
- Bollen, K. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly*, 35, 359–372. <https://doi.org/10.2307/23044047>
- Bollen, K. A. (2020). When good loadings go bad: Robustness in factor analysis. *Structural Equation Modeling*, 27(4), 515–524. <https://doi.org/10.1080/10705511.2019.1691005>
- Bonifay, W., & Cai, L. (2017). On the complexity of item response theory models. *Multivariate Behavioral Research*, 52(4), 465–484. <https://doi.org/10.1080/00273171.2017.1309262>
- 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>
- Bornoalova, 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., & Cramer, A. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91–121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>
- Borsboom, D., Robinaugh, D. J., Rhemtulla, M., & Cramer, A. O. J. (2018). Robustness and replicability of psychopathology networks. *World Psychiatry*, 17(2), 143–144. <https://doi.org/10.1002/wps.20515>
- Browne, M. W., MacCallum, R. C., Kim, C.-T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, 7(4), 403–421. <https://doi.org/10.1037//1082-989X.7.4.403>
- Burt, S. A. (2009). Are there meaningful etiological differences within antisocial behavior? Results of a meta-analysis. *Clinical Psychology Review*, 29(2), 163–178. <https://doi.org/10.1016/j.cpr.2008.12.004>
- Burt, S. A. (2012). How do we optimally conceptualize the heterogeneity within antisocial behavior? An argument for aggressive versus non-aggressive behavioral dimensions. *Clinical Psychology Review*, 32, 263–279. <https://doi.org/10.1016/j.cpr.2012.02.006>
- 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>
- Caspi, A., & Moffitt, T. E. (2018). All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry*, 175(9), 831–844. <https://doi.org/10.1176/appi.ajp.2018.17121383>
- Conway, C. C., Forbes, M. K., & South, S. C. (2022). A Hierarchical Taxonomy of Psychopathology (HiTOP) primer for mental health researchers. *Clinical Psychological Science*, 10(2), 236–258. <https://doi.org/10.1177/21677026211017834>
- DeYoung, C. G., Kotov, R., Krueger, R. F., Cicero, D. C., Conway, C. C., Eaton, N. R., Forbes, M. K., Hallquist,

- M. N., Jonas, K. G., Litzman, R. D., Rodriguez-Seijas, C., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., Widiger, T. A., & Wright, A. G. C. (2022). Answering questions about the Hierarchical Taxonomy of Psychopathology (HiTOP): Analogies to whales and sharks miss the boat. *Clinical Psychological Science*, *10*(2), 279–284. <https://doi.org/10.1177/21677026211049390>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Faure, K., & Forbes, M. K. (2021). Clarifying the placement of obsessive-compulsive disorder in the empirical structure of psychopathology. *Journal of Psychopathology and Behavioral Assessment*, *43*(3), 671–685. <https://doi.org/10.1007/s10862-021-09868-1>
- Fava, M., Rankin, M. A., Wright, E. C., Alpert, J. E., Nierenberg, A. A., Pava, J., & Rosenbaum, J. F. (2000). Anxiety disorders in major depression. *Comprehensive Psychiatry*, *41*(2), 97–102. [https://doi.org/10.1016/S0010-440X\(00\)90140-8](https://doi.org/10.1016/S0010-440X(00)90140-8)
- Feinstein, A. R. (1970). The pre-therapeutic classification of co-morbidity in chronic disease. *Journal of Chronic Diseases*, *23*(7), 455–468. [https://doi.org/10.1016/0021-9681\(70\)90054-8](https://doi.org/10.1016/0021-9681(70)90054-8)
- Ferrando, P. J., & Lorenzo-Seva, U. (2019). An external validity approach for assessing essential unidimensionality in correlated-factor models. *Educational and Psychological Measurement*, *79*(3), 437–461. <https://doi.org/10.1177/0013164418824755>
- 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. (2021). Three recommendations based on a comparison of the reliability and validity of the predominant models used in research on the empirical structure of psychopathology. *Journal of Abnormal Psychology*, *130*(3), 297–317. <https://doi.org/10.1037/abn0000533>
- Forbes, M. K., Magson, N. R., & Rapee, R. M. (2020). Evidence that different types of peer victimization have equivalent associations with transdiagnostic psychopathology in adolescence. *Journal of Youth and Adolescence*, *49*(3), 590–604. <https://doi.org/10.1007/s10964-020-01202-4>
- Forbes, M. K., Sunderland, M., Rapee, R. M., Batterham, P. J., Calear, A. L., Carragher, N., Ruggero, C., Zimmerman, M., Baillie, A. J., Lynch, S. J., Mewton, L., Slade, T., & Krueger, R. F. (2021). A detailed hierarchical model of psychopathology: From individual symptoms up to the general factor of psychopathology. *Clinical Psychological Science*, *9*(2), 139–168. <https://doi.org/10.1177/2167702620954799>
- Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*, *126*(7), 969–988. <https://doi.org/10.1037/abn0000276>
- Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2021). On unreplicable inferences in psychopathology symptom networks and the importance of unreliable parameter estimates. *Multivariate Behavioral Research*, *56*(2), 368–376. <https://doi.org/10.1080/00273171.2021.1886897>
- Fried, E. I., Greene, A. L., & Eaton, N. R. (2021). The p factor is the sum of its parts, for now. *World Psychiatry*, *20*(1), 69–70. <https://doi.org/10.1002/wps.20814>
- Fudge, D. S. (2014). Fifty years of J. R. Platt's strong inference. *Journal of Experimental Biology*, *217*(8), 1202–1204. <https://doi.org/10.1242/jeb.104976>
- Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: g as superordinate or breadth factor? *Psychology Science*, *50*(1), 21–43.
- Goldberg, L. R. (2006). Doing it all bass-ackwards: The development of hierarchical factor structures from the top down. *Journal of Research in Personality*, *40*(4), 347–358. <https://doi.org/10.1016/j.jrp.2006.01.001>
- 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., Litzman, R. D., Patrick, C. J., Reininghaus, 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>
- Greene, A. L., Watts, A. L., Forbes, M. K., Kotov, R., Krueger, R. F., & Eaton, N. R. (2022). Misbegotten methodologies and forgotten lessons from Tom Swift's electric factor analysis machine: A demonstration with competing structural models of psychopathology. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000465>
- Grotzinger, A. D., Mallard, T. T., Akingbuwa, W. A., Ip, H. F., Adams, M. J., Lewis, C. M., McIntosh, A. M., Grove, J., Dalsgaard, S., Lesch, K.-P., Strom, N., Meier, S. M., Mattheisen, M., Børglum, A. D., Mors, O., Breen, G., iPSYCH, Tourette Syndrome and Obsessive Compulsive Disorder Working Group of the Psychiatric Genetics Consortium, Bipolar Disorder Working Group of the Psychiatric Genetics Consortium, . . . Nivard, M. G. (2022). Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic, and molecular genetic levels of analysis. *Nature Genetics*, *54*, 548–599. <https://doi.org/10.1038/s41588-022-01057-4>
- Haefel, G. J., Jeronimus, B. F., Kaiser, B. N., Weaver, L. J., Soyster, P. D., Fisher, A. J., Vargas, I., Goodson, J. T., & Lu, W. (2022). Folk classification and factor rotations: Whales, sharks, and the problems with the Hierarchical Taxonomy of Psychopathology (HiTOP). *Clinical Psychological Science*, *10*(2), 259–278. <https://doi.org/10.1177/21677026211002500>
- Harzing, A.-W. (2016, February 6). *Publish or perish*. Harzing.com. <https://harzing.com/resources/publish-or-perish>
- Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika*, *76*(4), 537–549. <https://doi.org/10.1007/s11336-011-9218-4>
- Jennrich, R. I., & Bentler, P. M. (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika*, *77*(3), 442–454. <https://doi.org/10.1007/s11336-012-9269-1>
- Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of

- methods, recommendations, and software for composite measures. *Psychological Methods*, 21(1), 69–92. <https://doi.org/10.1037/a0040086>
- Kessler, R. C., Gruber, M., Hettema, J. M., Hwang, I., Sampson, N., & Yonkers, K. A. (2008). Comorbid major depression and generalized anxiety disorders in the national comorbidity survey follow-up. *Psychological Medicine*, 38(3), 365–374. <https://doi.org/10.1017/S0033291707002012>
- Kim, H., & Eaton, N. R. (2015). The hierarchical structure of common mental disorders: Connecting multiple levels of comorbidity, bifactor models, and predictive validity. *Journal of Abnormal Psychology*, 124(4), 1064–1078. <https://doi.org/10.1037/abn0000113>
- 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., Krueger, R. F., Watson, D., Cicero, D. C., Conway, C. C., DeYoung, C. G., Eaton, N. R., Forbes, M. K., Hallquist, M. N., Latzman, R. D., Mullins-Sweatt, S. N., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., & Wright, A. G. C. (2021). The Hierarchical Taxonomy of Psychopathology (HiTOP): A quantitative nosology based on consensus of evidence. *Annual Review of Clinical Psychology*, 17(1), 83–108. <https://doi.org/10.1146/annurev-clinpsy-081219-093304>
- Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry*, 56(10), 921–926. <https://doi.org/10.1001/archpsyc.56.10.921>
- Krueger, R. F., Kotov, R., Watson, D., Forbes, M. K., Eaton, N. R., Ruggero, C. J., Simms, L. J., Widiger, T. A., Achenbach, T. M., Bach, B., Bagby, R. M., Bornovalova, M. A., Carpenter, W. T., Chmielewski, M., Cicero, D. C., Clark, L. A., Conway, C., DeClercq, B., DeYoung, C. G., . . . Zimmermann, J. (2018). Progress in achieving quantitative classification of psychopathology. *World Psychiatry*, 17(3), 282–293. <https://doi.org/10.1002/wps.20566>
- Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is there a general factor of prevalent psychopathology during adulthood? *Journal of Abnormal Psychology*, 121(4), 971–977. <https://doi.org/10.1037/a0028355>
- Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D., & Zald, D. H. (2017a). A hierarchical causal taxonomy of psychopathology across the life span. *Psychological Bulletin*, 143(2), 142–186. <https://doi.org/10.1037/bul0000069>
- Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D., & Zald, D. H. (2017b). Validity and utility of the general factor of psychopathology. *World Psychiatry*, 16(2), 142–144. <https://doi.org/10.1002/wps.20410>
- 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>
- Lee, P. H., Feng, Y.-C. A., & Smoller, J. W. (2021). Pleiotropy and cross-disorder genetics among psychiatric disorders. *Biological Psychiatry*, 89(1), 20–31. <https://doi.org/10.1016/j.biopsych.2020.09.026>
- Levin-Aspenson, H. F., Khoo, S., & Kotelnikova, Y. (2019). Hierarchical taxonomy of psychopathology across development: Associations with personality. *Journal of Research in Personality*, 81, 72–78. <https://doi.org/10.1016/j.jrp.2019.05.006>
- Levin-Aspenson, H. F., Watson, D., Clark, L. A., & Zimmerman, M. (2021). What is the general factor of psychopathology? Consistency of the p factor across samples. *Assessment*, 28(4), 1035–1049. <https://doi.org/10.1177/1073191120954921>
- Lilienfeld, S. O. (1992). The association between antisocial personality and somatization disorders: A review and integration of theoretical models. *Clinical Psychology Review*, 12(6), 641–662. [https://doi.org/10.1016/0272-7358\(92\)90136-V](https://doi.org/10.1016/0272-7358(92)90136-V)
- Lilienfeld, S. O., Van Valkenburg, C., Larntz, K., & Akiskal, H. S. (1986). The relationship of histrionic personality disorder to antisocial personality and somatization disorders. *The American Journal of Psychiatry*, 143(6), 718–722. <https://doi.org/10.1176/ajp.143.6.718>
- Lilienfeld, S. O., & Waldman, I. D. (1990). The relation between childhood attention-deficit hyperactivity disorder and adult antisocial behavior reexamined: The problem of heterogeneity. *Clinical Psychology Review*, 10(6), 699–725. [https://doi.org/10.1016/0272-7358\(90\)90076-M](https://doi.org/10.1016/0272-7358(90)90076-M)
- Lilienfeld, S. O., & Waldman, I. D. (2004). Comorbidity and chairman Mao. *World Psychiatry*, 3(1), 26–27.
- Lilienfeld, S. O., Waldman, I. D., & Israel, A. C. (1994). A critical examination of the use of the term and concept of comorbidity in psychopathology research. *Clinical Psychology: Science and Practice*, 1(1), 71–83. <https://doi.org/10.1111/j.1468-2850.1994.tb00007.x>
- Loehlin, J. C., & Beaujean, A. A. (2016). *Latent variable models: An introduction to factor, path, and structural equation analysis* (5th ed.). Taylor & Francis.
- Lorenzo-Seva, U., & Ferrando, P. J. (2019). A general approach for fitting pure exploratory bifactor models. *Multivariate Behavioral Research*, 54(1), 15–30. <https://doi.org/10.1080/00273171.2018.1484339>
- MacCallum, R., Wegener, D., Uchino, B., & Fabrigar, L. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, 114, 185–199. <https://doi.org/10.1037//0033-2909.114.1.185>
- Mansolf, M., & Reise, S. P. (2016). Exploratory bifactor analysis: The Schmid-Leiman orthogonalization and Jennrich-Bentler analytic rotations. *Multivariate Behavioral Research*, 51(5), 698–717. <https://doi.org/10.1080/00273171.2016.1215898>
- Markon, K. E. (2010). Modeling psychopathology structure: A symptom-level analysis of Axis I and II disorders. *Psychological Medicine*, 40(2), 273–288. <https://doi.org/10.1017/S0033291709990183>

- Markon, K. E. (2019). Bifactor and hierarchical models: Specification, inference, and interpretation. *Annual Review of Clinical Psychology, 15*, 51–69. <https://doi.org/10.1146/annurev-clinpsy-050718-095522>
- Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin, 137*(5), 856–879. <https://doi.org/10.1037/a0023678>
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2
- Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*(1), 85–110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- 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. (1985). *Factor analysis and related methods*. Psychology Press.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum.
- McNally, R. J. (2016). Can network analysis transform psychopathology? *Behaviour Research and Therapy, 86*, 95–104. <https://doi.org/10.1016/j.brat.2016.06.006>
- McNally, R. J. (2021). Network analysis of psychopathology: Controversies and challenges. *Annual Review of Clinical Psychology, 17*, 31–53. <https://doi.org/10.1146/annurev-clinpsy-081219-092850>
- McNeish, D., & Wolf, M. G. (2021). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000425>
- Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research? A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence, 3*(1), 2–20. <https://doi.org/10.3390/jintelligence3010002>
- Mulaik, S. A., & Quartetti, D. A. (1997). First order or higher order general factor? *Structural Equation Modeling, 4*(3), 193–211. <https://doi.org/10.1080/10705519709540071>
- Murray, A. L., Booth, T., Eisner, M., Obsuth, I., & Ribeaud, D. (2019). Quantifying the strength of general factors in psychopathology: A comparison of CFA with maximum likelihood estimation, BSEM, and ESEM/EFA bifactor approaches. *Journal of Personality Assessment, 101*(6), 631–643. <https://doi.org/10.1080/00223891.2018.1468338>
- Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41*(5), 407–422. <https://doi.org/10.1016/j.intell.2013.06.004>
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.).
- Neale, M., & Cardon, L. R. (2013). *Methodology for genetic studies of twins and families*. Springer.
- Neumann, A., Pappa, I., Lahey, B. B., Verhulst, F. C., Medina-Gomez, C., Jaddoe, V. W., Bakermans-Kranenburg, M. J., Moffitt, T. E., van IJzendoorn, M. H., & Tiemeier, H. (2016). Single nucleotide polymorphism heritability of a general psychopathology factor in children. *Journal of the American Academy of Child & Adolescent Psychiatry, 55*(12), 1038–1045. <https://doi.org/10.1016/j.jaac.2016.09.498>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Pezzoli, P., Antfolk, J., & Santtila, P. (2017). Phenotypic factor analysis of psychopathology reveals a new body-related transdiagnostic factor. *PLOS ONE, 12*(5), Article e0177674. <https://doi.org/10.1371/journal.pone.0177674>
- Platt, J. R. (1964). Strong inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science, 146*(3642), 347–353. <https://doi.org/10.1126/science.146.3642.347>
- Preacher, K. J. (2006). Quantifying parsimony in structural equation modeling. *Multivariate Behavioral Research, 41*(3), 227–259. https://doi.org/10.1207/s15327906mbr4103_1
- Raykov, T., & Penev, S. (1999). On structural equation model equivalence. *Multivariate Behavioral Research, 34*(2), 199–244. <https://doi.org/10.1207/S15327906Mbr340204>
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>
- Riglin, L., Thapar, A. K., Leppert, B., Martin, J., Richards, A., Anney, R., Davey Smith, G., Tilling, K., Stergiakouli, E., Lahey, B. B., O'Donovan, M. C., Collishaw, S., & Thapar, A. (2020). Using genetics to examine a general liability to childhood psychopathology. *Behavior Genetics, 50*(4), 213–220. <https://doi.org/10.1007/s10519-019-09985-4>
- Ringwald, W. R., Beeney, J. E., Pilkonis, P. A., & Wright, A. G. C. (2019). Comparing hierarchical models of personality pathology. *Journal of Research in Personality, 81*, 98–107. <https://doi.org/10.1016/j.jrp.2019.05.011>
- Ringwald, W. R., Forbes, M. K., & Wright, A. G. C. (2023). Meta-analysis of structural evidence for the Hierarchical Taxonomy of Psychopathology (HiTOP) model. *Psychological Medicine, 53*(2), 533–546. <https://doi.org/10.1017/S0033291721001902>
- Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine, 50*(3), 353–366. <https://doi.org/10.1017/S0033291719003404>
- 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>

- Savalei, V., & Reise, S. P. (2019). Don't forget the model in your model-based reliability coefficients: A reply to McNeish (2018). *Collabra: Psychology*, 5(1), Article 36. <https://doi.org/10.1525/collabra.247>
- Sellbom, M., Cooke, D. J., & Hart, S. D. (2015). Construct validity of the Comprehensive Assessment of Psychopathic Personality (CAPP) concept map: Getting closer to the core of psychopathy. *International Journal of Forensic Mental Health*, 14(3), 172–180. <https://doi.org/10.1080/14999013.2015.1085112>
- Sharp, C., Wright, A. G. C., Fowler, J. C., Frueh, B. C., Allen, J. G., Oldham, J., & Clark, L. A. (2015). The structure of personality pathology: Both general ('g') and specific ('s') factors? *Journal of Abnormal Psychology*, 124(2), 387–398. <https://doi.org/10.1037/abn0000033>
- Singh, A. L., & Waldman, I. D. (2010). The etiology of associations between negative emotionality and childhood externalizing disorders. *Journal of Abnormal Psychology*, 119(2), 376–388. <https://doi.org/10.1037/a0019342>
- Smith, G. T., Atkinson, E. A., Davis, H. A., Riley, E. N., & Oltmanns, J. R. (2020). *The general factor of psychopathology*. *Annual Review of Clinical Psychology*, 16, 75–98. <https://doi.org/10.1146/annurev-clinpsy-071119-115848>
- Tukey, J. (1977). *Exploratory data analysis*. Pearson.
- van der Sluis, S., Posthuma, D., Nivard, M. G., Verhage, M., & Dolan, C. V. (2013). Power in GWAS: Lifting the curse of the clinical cut-off. *Molecular Psychiatry*, 18(1), 2–3. <https://doi.org/10.1038/mp.2012.65>
- Waldman, I. D. (2017, July 28). *Issues in the validation of the general factor of psychopathology* [Conference session]. Behavior Genetics Association Annual Meeting, Oslo, Norway.
- Waldman, I. D., & Lilienfeld, S. O. (1991). Diagnostic efficiency of symptoms for oppositional defiant disorder and attention-deficit hyperactivity disorder. *Journal of Consulting and Clinical Psychology*, 59(5), 732–738. <https://doi.org/10.1037/0022-006X.59.5.732>
- Waldman, I. D., & Lilienfeld, S. O. (2001). Applications of taxometric methods to problems of comorbidity: Perspectives and challenges. *Clinical Psychology: Science and Practice*, 8(4), 520–527. <https://doi.org/10.1093/clipsy.8.4.520>
- Waldman, I. D., Lilienfeld, S. O., & Lahey, B. B. (1995). Toward construct validity in the childhood disruptive behavior disorders. In T. H. Ollendick & R. J. Prinz (Eds.), *Advances in clinical child psychology* (pp. 323–363). Springer. https://doi.org/10.1007/978-1-4757-9044-3_8
- Waldman, I. D., Poore, H. E., Luningham, J. M., & Yang, J. (2020). Testing structural models of psychopathology at the genomic level. *World Psychiatry*, 19(3), 350–359. <https://doi.org/10.1002/wps.20772>
- 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>
- Waldman, I. D., Rowe, D. C., Abramowitz, A., Kozel, S. T., Mohr, J. H., Sherman, S. L., Cleveland, H. H., Sanders, M. L., Gard, J. M. C., & Stever, C. (1998). Association and linkage of the dopamine transporter gene and attention-deficit hyperactivity disorder in children: Heterogeneity owing to diagnostic subtype and severity. *The American Journal of Human Genetics*, 63(6), 1767–1776. <https://doi.org/10.1086/302132>
- Waszczuk, M. A., Eaton, N. R., Krueger, R. F., Shackman, A. J., Waldman, I. D., Zald, D. H., Lahey, B. B., Patrick, C. J., Conway, C. C., Ormel, J., Hyman, S. E., Fried, E. I., Forbes, M. K., Docherty, A. R., Althoff, R. R., Bach, B., Chmielewski, M., DeYoung, C. G., Forbush, K. T., . . . Kotov, R. (2020). Redefining phenotypes to advance psychiatric genetics: Implications from hierarchical taxonomy of psychopathology. *Journal of Abnormal Psychology*, 129(2), 143–161. <https://doi.org/10.1037/abn0000486>
- Watts, A. L., Boness, C. L., Loeffelman, J. E., Steinley, D., & Sher, K. J. (2021). Does crude measurement contribute to observed unidimensionality of psychological constructs? A demonstration with DSM–5 alcohol use disorder. *Journal of Abnormal Psychology*, 130(5), 512–524. <https://doi.org/10.1037/abn0000678>
- 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>
- Wright, A. G. C., Krueger, R. F., Hobbs, M. J., Markon, K. E., Eaton, N. R., & Slade, T. (2013). The structure of psychopathology: Toward an expanded quantitative empirical model. *Journal of Abnormal Psychology*, 122(1), 281–294. <https://doi.org/10.1037/a0030133>
- Wright, A. G. C., & Simms, L. J. (2015). A metastructural model of mental disorders and pathological personality traits. *Psychological Medicine*, 45(11), 2309–2319. <https://doi.org/10.1017/S0033291715000252>
- Yang, Y., & Green, S. B. (2010). A note on structural equation modeling estimates of reliability. *Structural Equation Modeling*, 17(1), 66–81. <https://doi.org/10.1080/10705510903438963>
- Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, 64(2), 113–128. <https://doi.org/10.1007/BF02294531>
- Zhang, G., Hattori, M., Trichtinger, L. A., & Wang, X. (2019). Target rotation with both factor loadings and factor correlations. *Psychological Methods*, 24(3), 390–402. <https://doi.org/10.1037/met0000198>
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω_H : Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123–133. <https://doi.org/10.1007/s11336-003-0974-7>