



Published in final edited form as:

Psychol Inq. 2020 ; 31(4): 310–320. doi:10.1080/1047840x.2020.1853476.

Building theories on top of, and not independent of, statistical models: The case of the p -factor

Ashley L. Watts¹, Sean P. Lane², Wes Bonifay³, Douglas Steinley¹, Francisco A. C. Meyer⁴

¹Department of Psychological Sciences, University of Missouri, Columbia, MO, USA

²Department of Psychological Sciences, Purdue University, West Lafayette, IN, USA

³Department of Education, School, and Counseling Psychology, University of Missouri, Columbia, MO, USA

⁴Department of Psychology, Vanderbilt University, Nashville, TN, USA

Eiko Fried (this issue) outlines several ways in which factor and network models do not bear on the theories they purport to test. He argues, in turn, that there is a major inferential gap between our theories and our models in the psychological literature. Fried inspired us to muse on the inferential gap between theory and model in the p -factor literature specifically, because it represents a particularly salient intersection of psychological theory and statistical modeling. Herein, we argue that the p -factor literature is a microcosm of the broader psychological literature Fried portrays: theory and models are confounded; our models are not risky enough tests of our theories, if they bear on them at all; and our theories are weak, and thus difficult to adjudicate or falsify.

Theory, models, and data are necessarily intertwined but, as we noted, unfortunately confused in the p -factor literature, so much so that it is often difficult to glean whether one is referring to a theory or a model when describing the p -factor. For instance, the terms “ p -factor” and “general factor of psychopathology” are used to refer to both the theoretical and statistical constructs (Caspi et al., 2014; Caspi & Moffitt, 2018; Lahey et al., 2012; Smith et al., 2020). Like Fried (this issue), we think it is wise to use terms that distinguish theory and model. In this commentary, we use “ p -factor” when referring to the theoretical construct, and “general factor of psychopathology” when referring to a statistical instantiation of the p -factor (Watts et al., 2019). The term “general factor” came about because early models of the p -factor tested it with the bifactor model, which contains a global, “general” factor that accounts for the shared variance of all indicators in the model (Caspi et al., 2014; Lahey et al., 2012).

During the exercise of distinguishing theory and model as they pertain to the p -factor, we identified four major issues with the existing literature, each of which arguably arise from the theory-model inference gap that Fried (this issue) explicates:

1. Model fit is an unreliable marker of model validity.

2. Most existing explanations of the p -factor are weak theories.
3. The p -factor is ill-defined and even protean.
4. The p -factor is described as a causal mechanism, but latent variable models do not reveal causality.

In what follows, after we introduce the concept of the p -factor and how it came about, we expand on each of these four issues in more detail and situate them within the p -factor and broader quantitative psychology literatures. Along the way, we offer suggestions for moving forward, specifically, ways to place the p -factor at stronger theoretical risk. We hope such suggestions refine and strengthen tests and theories of the p -factor, should it exist.

The “Discovery” of the p -factor

Categorical classification systems, such as those outlined in our diagnostic manuals, are plagued by comorbidity, or diagnostic co-occurrence (Krueger & Markon, 2006). Most people that meet diagnostic criteria for one mental disorder meet for two (61–88%); those that meet criteria for two disorders have increased odds of meeting for three, and so on (Couwenbergh et al., 2006; see also Kessler et al., 2005). Observing this strong overlap among putatively distinct forms of psychopathology, the quantitatively inclined have enlisted factor analytic techniques to empirically reorganize psychopathology.

Much work has emphasized two superordinate dimensions: externalizing and internalizing (Achenbach, 1966; Krueger & Markon, 2006). Externalizing comprises psychiatric conditions characterized by poor emotional and behavioral control and is exemplified by the likes of antisociality, conduct problems, oppositionality, substance use, and attention problems. Internalizing comprises conditions characterized by negative emotionality, and is exemplified by the likes of anxiety, mood, fears, and eating pathologies. Other work has bifurcated internalizing into distress and fears (Krueger, 1999), and externalizing into disinhibition and antagonism (Kotov et al., 2011). Still other work has identified additional superordinate dimensions, such as thought problems and detachment (Keyes et al., 2013; Kotov et al., 2011). These efforts have resulted in a reasonably fine-grained, hierarchically-organized structure of psychopathology (Kotov et al., 2017). At the base of the hierarchy are signs and symptoms, which are organized in terms of increasingly broad, transdiagnostic dimensions as one scales it.

Although externalizing and internalizing were once thought to lie at the pinnacle of the psychopathology hierarchy, the number and nature of highest-order dimensions atop it has been contested fairly recently. For instance, it is well-demonstrated that externalizing and internalizing dimensions are at least moderately correlated ($r \approx .50 - .60$ in most samples; Krueger & Markon, 2006), which has spawned inquiries into potential sources of that covariation. Enter the p -factor, which is thought to reflect an overarching, unitary dimension that gives rise to psychopathology broadly construed (Caspi et al., 2014; Lahey et al., 2012).

At present, researchers have claimed to detect a general factor of psychopathology in virtually any sample type, age group, and psychopathology assessment. This work has led to the conclusion that the p -factor is robust and replicable across a variety of important

moderators (Caspi & Moffitt, 2018; Smith et al., 2020). Many studies have emphasized the real-world correlates of a general factor of psychopathology, and most studies tend to conclude that the p -factor is associated with numerous constructs reflecting decreased functioning. Reasonably-established correlates of the p -factor include lower levels of intelligence and executive functioning, income, conscientiousness, and agreeableness, as well as higher levels of family history of psychopathology, impairment, childhood maltreatment, harsh parenting, peer delinquency, self-harm, neuroticism, impulsivity, sensation seeking, aggression, and hopelessness (e.g., Albott et al., 2018; Caspi et al., 2014; Harden et al., 2020; Lahey et al., 2012; Tackett et al., 2013; Watts et al., 2019). Other work has shown that a general factor of psychopathology is moderately heritable (Tackett et al., 2013; Waldman et al., 2016), with a smattering of neurobiological substrates (e.g., altered connectivity in the anterior cingulate cortex and cerebellum, reduced grey matter volume; Kaczkurkin et al., 2018; Romer, Elliott, et al., 2019; Romer, Knodt, et al., 2019).

Conceptual and Statistical Issues in the p -factor Literature

Although much research – by our count, over 100 studies since 2012 – has proceeded by presuming that the p -factor is an entity, recent research has challenged the psychological and statistical evidence that forms the basis of p -factor theory. In recent years, we have argued that the p -factor literature has, in one way or another, developed incautiously. The p -factor literature tends to confuse a well-fitting model with a valid and structurally replicable one, and a well-fitting (generally bifactor) model has been used to reify the p -factor as a causal entity. Nevertheless, the conclusion that the p -factor exists occurred only after finding that a bifactor model fit better than the widely-used correlated factors model (Caspi et al., 2014; Lahey et al., 2012; for discussion, see: Bonifay et al., 2017; Littlefield et al., in press; Watts et al., 2019).¹ Quite explicitly, the interpretation of the p -factor has been iteratively retconned (i.e., retrospectively reconfigured to accommodate pre-existing notions) so as to empirically sustain itself. As the p -factor literature has blossomed, so too have our concerns.

Here, we describe evidence that challenges the very notion of a p -factor and the model by which it is typically examined (the bifactor model), which we structure around the four observations we outlined earlier. We divide these observations into their own sections for the sake of organization, but much of the research described later informs more than one of these observations.

To characterize the extent of various practices (e.g., selecting a bifactor model based on model fit alone) and cross-study consistency in the general factors of psychopathology reported in the literature, we conducted a condensed review that randomly sampled 25 percent of studies that extracted a general factor of psychopathology using a bifactor model. We limited our search to studies that used diagnoses or symptom counts, as opposed to individual items, for the sake of simplicity and to facilitate comparison across studies. Fifteen studies' loadings on a general factor are reviewed here (see Table 1). The purpose of

¹The higher-order model is less widely used in tests of alternative models (cf., Forbes et al., under review) because one needs at least four second-order dimensions (e.g., externalizing, distress, fears, thought disorder) to differentiate the higher-order model from a correlated factors model in terms of model fit (Markon, 2019).

this review is not to be comprehensive but illustrative. From time to time, we will summarize trends in the literature based on these studies.

Issue #1: Model fit is an unreliable marker of model validity.

The most popular way to model a general factor of psychopathology is using the bifactor model. The bifactor model decomposes covariation among all indicators in the model, say psychopathology symptoms or disorders, into a general factor and any number of specific factors; all factors in the model are assumed to be orthogonal (Reise, 2012), although this last requirement is ignored routinely (e.g., Caspi et al., 2014; Tackett et al., 2013; Waldman et al., 2016). The general factor directly influences all of its indicators, which, from the reflective, latent variable perspective, can be (and has been) interpreted to mean that the general factor underlies all variables in the data (Reise, 2012). Specific factors are typically viewed as influencing and explaining the covariance among subsets of indicators that are presumably highly similar in content (e.g., externalizing, internalizing). As such, bifactor models appear at face value to offer something appealing to psychopathology researchers: the ability to model general and specific sources of variance in psychopathology (Brandes et al., 2019).

Bifactor models have also been widely criticized. Chief among these criticisms is that bifactor model fit is an unreliable indicator of model validity, that is, the degree to which the model represents some natural or “true” structure. Importantly, the bifactor model tends to show superior goodness of fit over other models (i.e., correlated factors models, higher-order models) even when the true model does not follow a bifactor structure (e.g., Maydeu-Olivares & Coffman, 2006; Morgan et al., 2015; Murray & Johnson, 2013). This bias in favor of the bifactor model in terms of model fit has been established through the use of simulation, where one can specify the true (“population”) structure and test any series of competing models and compare their fit. Bias is present when model fit indicates a preference for a model other than the true structure. Bifactor model fit bias has been well-established in the intelligence literature (e.g., Murray & Johnson, 2013) and was more recently extended to the psychopathology literature (Greene et al., 2019). Greene and colleagues (2019) showed that model fit more often than not could not identify the true structure.

Perhaps the most serious shortcoming of the bifactor model is that it tends to fit better than other models, not because it is a valid representation of the data, but due to its tendency to overfit data. A number of studies have shown that the bifactor model has high fitting propensity (Bonifay & Cai, 2017), meaning that it accommodates any possible data, including random and nonsensical response patterns (Bonifay & Cai, 2017; Reise et al., 2016). In fact, the extensive simulations conducted by Bonifay and Cai showed that a (putatively confirmatory) bifactor model and an exploratory factor analytic (EFA) model have similar propensities to fit well, and on occasion, the confirmatory bifactor model may actually fit *better* than the EFA model. In short, the bifactor model is inherently highly flexible; it can model important trends and noise, but the trends and noise are indistinguishable.

Considering the bifactor model's high fit propensity, model fit should not be used as evidence that some theory or hypothesis has been supported. Of the 15 studies we reviewed, 80 percent either selected the bifactor model on the basis of preferential model fit (60%), or simply did not consider an alternative structure (20%; e.g., correlated factors model). Put another way, only 20 percent of studies considered criteria other than model fit in adjudicating structural models. Notably, these few studies did not necessarily conclude that the bifactor model of psychopathology was the best model of those tested, and instead presented more than one model, usually a bifactor model and correlated factors model (Conway et al., 2019; Shields et al., 2019; Watts et al., 2019). Our condensed review suggests that the practice of relying on model fit to justify the use of the bifactor model is widespread, despite model fit's inability to bear on validity.

Additionally, the bifactor model tends to be associated with other less than ideal properties. Given the bifactor model's inclination to overfit data, it tends also to produce specific factors with irregular loading patterns (Eid et al., 2017). Indeed, many bifactor models of psychopathology have yielded specific factors with low and even negative loadings (e.g., Caspi et al., 2014; Castellanos-Ryan et al., 2016; Martel et al., 2017; Watts et al., 2019). For instance, thought disorder specific factors have been, ironically, negatively indicated by psychosis (Caspi et al., 2014; Martel et al., 2017), and distress specific factors have been weakly or negatively indicated by generalized anxiety disorder and major depression; in both of these cases, these indicators are thought to typify their constructs of interest (e.g., Forbes et al., under review; Tackett et al., 2013; Watts et al., 2019).

Notwithstanding concerns over interpretability of these and other specific factors (Bagby et al., 2007; Vanheule et al., 2008), a natural consequence of poorly defined specific factors is that they are inconsistent across studies and sometimes unreliable. We have raised concerns about poor factor reliability in bifactor models of psychopathology in other work (Bonifay et al., 2017; Watts et al., 2019; see also Forbes et al., under review; Kim, Greene, et al., 2019), and Rodriguez and colleagues (2016a, 2016b) noted this tendency of the bifactor model more generally. Of the 15 studies included in our review, 80 percent reported one or more specific factors with standardized loadings of less than .20, and many of these loadings were near 0 or even negative. The median reliabilities (per omega hierarchical subscale, Rodriguez et al., 2016) were .49 for externalizing specific factors and .30 for internalizing specific factors.

Considering the limitations of the bifactor model, we have argued that model selection should not rely exclusively on model fit (Bonifay et al., 2017; Bornovalova et al., 2020; Murray & Johnson, 2013; Roberts & Pashler, 2000; Sellbom & Tellegen, 2019; Watts et al., 2019). We might take this a step further: Model fit does not necessarily indicate model validity or superiority. Potentially innumerable models that are conceptually non-equivalent but equally well-fitting can be observed in the same data (Raykov & Marcoulides, 2001; Tomarken & Waller, 2003). In this way, concluding that the bifactor model of psychopathology is a better representation of nature because it fits better than another model is indefensible.

Still, researchers defend the use of the bifactor model for two reasons. The first reason is that there is generally a positive manifold among forms of psychopathology. Under a positive manifold, the century old Perron-Frobenius theorem (Horn & Johnson, 2012) guarantees that all loadings on the so-called general factor will be positive.² Nevertheless, a positive manifold should not be taken as evidence for a *p*-factor. General factors may also arise when the data are not simple structured (i.e., with no cross-loadings; Revelle & Wilt, 2012), which is an assumption rarely met in modeling assessments of psychological constructs (Sellbom & Tellegen, 2019). Additionally, Hoffman and colleagues (2019) showed that specific characteristics of item design can induce a positive manifold among forms of psychopathology. Thus, there are a number of methodological, as opposed to substantive, reasons a general factor of psychopathology may arise in data (e.g., skip-out assessment structure, evaluative consistency bias; see also Bäckström et al., 2009; Chang et al., 2012; Davies et al., 2015; DeYoung, 2006).

Moreover, general factors may simply mimic a specific factor. For instance, although all indicators are allowed to load onto a general factor in a bifactor model, there are a number of bifactor models of psychopathology with general factors that are overrepresented by some forms of psychopathology, and underrepresented (or even unrepresented or negatively indicated) by others. Revelle and Wilt (2013) highlight that a general factor with large variance in terms of its factor loadings is an indication of a specific factor rather than a general factor. Table 1 displays the range in factor loadings on the general factor for each study included in our condensed review. Many studies' general factors were characterized by considerable variability, with an extreme example being that a general factor was composed of loadings that ranged from .00 to .87 (Olino et al., 2018). Typically, studies that include thought disorder indicators tend to find that the general factor is defined by them (e.g., Caspi et al., 2014; Romer, Knodt, et al., 2019), whereas studies that do not include thought disorder indicators tend to find that distress defines the general factor (e.g., Lahey et al., 2012; Tackett et al., 2013; Watts, Poore, & Waldman, 2019). Externalizing and fears (and fears, particular) rarely define general factors.

The second reason researchers justify the use of a bifactor model is that it often yields factors that relate differentially with external criteria. Relatedly, researchers tend to conclude that the *p*-factor is substantive because the general factor correlates with real-world behaviors. This leads us to three important considerations regarding validity. First, the detection of differential correlates is insufficient justification for the use of any model. Differential external validity is only interesting when we are reasonably certain that the model closely approximates the natural structure of the phenomena under investigation. Based on model fit, there is no clear way to determine whether a bifactor model is the valid structural representation of psychopathology, and so whether its factors bear distinguishable relations with external criteria is neither here nor there. In other words, the external validity of the model is only relevant once its internal validity is established (Campbell & Fiske, 1959).

²Also, others support the presence of a general factor of psychopathology on the basis of the first factor or principal component due to the (large) size of the first Eigenvalue, but it is well-known that the estimate of the first Eigenvalue is positively biased (Lawley, 1956); this bias was the primary motivation for the development of parallel analysis (Horn, 1965).

Second, thus far, the *p*-factor literature emphasizes convergent validity and entirely neglects discriminant validity. Few studies consider what should *not* correlate with a general factor, and thus they fail to provide one avenue towards its potential falsification. Are there correlates the *p*-factor should *not* have? If not, justifying the *p*-factor's existence because the general factor correlates with other constructs is insufficient (Caspi & Moffitt, 2018). Plus, many correlates of general factors of psychopathology (e.g., income, childhood abuse or neglect, cerebellar dysfunction) are so nonspecific that they cannot inform the substantive meaning of the *p*-factor, or adjudicate among competing theories. Are there correlates of the *p*-factor that might bear on the validity of one theory over another?

Third, even still, the traditional (“tripartite”) effort to establish validity – characterized by criterion, content, and construct validation methods – may be a lost cause. Borsboom and colleagues (2004) cast doubt on correlational evidence, in particular, arguing that “the primary objective of validation research is not to establish that the correlations go in the right directions but to offer a theoretical explanation of the processes that lead up to the measurement outcomes” (p. 1067; cf., Westen & Rosenthal, 2005). Kane (2013) completely overhauled the traditional view of validity through his 8-step interpretation/use argument (IUA) approach. According the IUA approach, at the outset of a study, researchers must form a careful argument about precisely *how* the measurements from a particular instrument should be used and interpreted. Validation is then based on this argument rather than simple correlations with external criteria. In the present context, adoption of the IUA perspective would require psychopathology researchers to carefully consider the indicators that they include in their models, the exact meaning of the factors and their scores, and the expected generalizability of the results obtained by fitting the model.

Issue #2: Most existing explanations of the *p*-factor are weak theories.

Although early work on the *p*-factor was relatively agnostic with regards to mechanisms thought to underlie the *p*-factor (Lahey et al., 2012), later work has advanced numerous mechanisms, including: negative emotionality (Brandes et al., 2019; Lahey et al., 2017; Tackett et al., 2013), poor emotional and behavioral control (Carver et al., 2017), disordered thought (Caspi & Moffitt, 2018), severity of psychopathology (Caspi et al., 2014; Caspi & Moffitt, 2018), and functional impairment (Smith et al., 2020; Widiger & Oltmanns, 2017). Each of these theories are sensible but positing that the *p*-factor reflects a relatively generic mechanism makes it difficult to adjudicate. Based on Fried's definition of weak theories, we conclude that most *p*-factor theories are weak for two reasons: because they 1) “are narrative and imprecise accounts of hypotheses” (p. *TBD*), and 2) “predict observations predicted by other theories” (p. *TBD*). We use impairment as an exemplar of weak *p*-factor theories, but what we argue here could be said of negative emotionality, thought disorder, and severity, among others.

Impairment is the most plausible substantive interpretation of the *p*-factor (Smith et al., 2020). Arguably, that the *p*-factor reflects impairment might not even be a theory *per se*, as much as a conciliatory interpretation of the literature given the cross-study inconsistencies in general factors. We note that the impairment theory is most plausible, only because it is among the most flexible of theories: Disorders include impairment by definition, and

extremes of latent symptom dimensions are invariably associated with impairment (Nuzum et al., 2019). As such, to support the impairment theory of the p -factor, the general factor simply has to be indicated by some form of psychopathology, no matter what indicators are included, and no matter their magnitude of contribution to the factor. As we highlighted earlier, general factors of psychopathology will always contain some psychopathology-relevant variance out of mathematical necessity. So, any weak p -factor theory can be “supported” simply because a general factor is modeled. This renders the impairment theory virtually non-falsifiable, and any conclusion that the p -factor reflects some substantive mechanism is essentially tautological.

As a brief aside, others have concluded that the p -factor reflects disordered thought (Caspi & Moffitt, 2018). In early attempts to model the p -factor, Caspi and colleagues (2014) found that thought problems essentially defined the general factor in a bifactor model, and there was a negative residual variance for psychosis when a thought problems specific factor was included in the model; essentially, there was no remaining variance in thought problems after modeling the general factor. In turn, Caspi and colleagues dropped the thought problems specific factor from the model, testing what Eid and colleagues (2017) termed the bifactor S-1 model. In the bifactor S-1 model, the specific factor for the subset of indicators thought to define the general factor is dropped, in turn converting the general factor a reference domain reflecting whichever specific factor is dropped (for other examples, see Lahey et al., 2018; Tackett et al., 2013; Waldman et al., 2016). So, the conclusion that the general factor reflects those indicators is circular. Also, due to pragmatic challenges with assessment and data collection, psychosis (among other dimensions) is rarely included in structural models of psychopathology. Should the p -factor reflect thought problems, models that do not include features that exemplify thought problems are likely not well-suited to test that theory. At a minimum, the exclusion of thought problems and related phenomena (e.g., obsessive-compulsive disorder, mania) raises questions about the compatibility between general factors that do and do not include those indicators.

To avoid acceptance of weak theories, Fried (this issue) called for greater emphasis on falsifiability, which is the possibility of an assertion, hypothesis, or theory to be disproven by contradictory evidence. A strong theory should predict observations that *must occur* to support it, and if those observations do not occur, the theory is false (see also Meehl, 1978). As Popper (1962), the progenitor of scientific falsification, argued, “A theory which is not refutable by any conceivable event is non-scientific.” In accord with riskier prediction (Meehl, 1978), researchers should prespecify the indicators that they expect to exemplify a general factor. Better yet, one could prespecify a reasonably narrow range of loadings for each indicator based on the existing literature or explicit hypotheses, and further use Bayesian approaches to compare such “constrained” theoretical accounts with the predominant “agnostic” approaches toward fitting a general factor of psychopathology.

For instance, if one believes that disordered thought gives rise to all forms of psychopathology (Caspi & Moffitt, 2018), one might prespecify that their general factor should be most highly defined by indicators that exemplify thought problems, such as schizophrenia, mania, and obsessive-compulsive disorder. Similarly, if one believes that a tendency toward negative mood states gives rise to all forms of psychopathology, they might

specify that a general factor should be most defined by indicators that exemplify negative mood, including depression, anxiety, and fears. Nevertheless, how these predictions can be disentangled from the possibility that the *p*-factor reflects severity or impairment remains unclear. We are not aware of a single *p*-factor paper that has explicitly outlined which indicators are expected to define a general factor. Such predictions would place the *p*-factor at stronger theoretical risk.

Issue #3: The *p*-factor is ill-defined and even protean.

Implicit in reviews of the *p*-factor is that it replicates across studies (Caspi & Moffitt, 2018; Smith et al., 2020). Even Fried (this issue), in critiquing the *p*-factor literature, noted that the “statistical *p*-factor is indeed well-replicated” (p. *TBD*). In the context of the *p*-factor literature, “replicates” may simply mean that the same form of model was tested, with little consideration of whether, for instance, its patterning of loadings or correlates replicated. Extracting a factor does not mean that one is extracting the same factor across studies (Lane et al., 2016). To date, no review has explicitly characterized the contents of general factors of psychopathology across studies.

We thought it might be useful to do so briefly. Figure 1 presents the standardized factor loadings on general factors of psychopathology for each of the 15 studies included in our condensed review. This figure shows that an indicator’s loading on a general factor of psychopathology can vary quite dramatically from study to study. The most extreme example is panic disorder/agoraphobia, whose loadings on general factors ranged from $-.02$ to $.81$. Depression and generalized anxiety similarly varied in terms of their loadings on general factors (ranges: depression = $.26 - .94$, generalized anxiety = $.30 - .95$). Further supporting this idea is that the structure of *p*-factor was highly inconsistent across the studies in our condensed review ($ICC=.24$), reflecting questionable replicability.

We suspect that part of the reason why general factors may not replicate across studies include is because they include different indicators of psychopathology (Bornoalova et al., 2020). By this, we do not mean that psychopathology is assessed with different instruments but that the same forms of psychopathology are not consistently included in each general factor. Table 1 displays the representation of each indicator across the random selection of general factors we considered in our condensed review. Certain indicators are invariably included in the modeling of a general factor of psychopathology, such as major depression, generalized anxiety, and some form of antisociality, be it antisocial personality or conduct disorder. Other indicators appear much more sporadically, some of which appear in few studies. Examples of lesser represented indicators include autism spectrum disorder, bipolar disorder, schizophrenia or thought disorder, and obsessive-compulsive disorder.

In earlier work, Watts and colleagues (2019) showed that general factors of psychopathology can vary considerably as a function of what is included when it is modeled. Using a reasonably comprehensive set of 14 child psychopathology indicators (symptom dimensions), Watts and colleagues (2019) fit a bifactor model of psychopathology and re-estimated general factors by dropping each specific (symptom) dimension from it in turn, resulting in 15 models tested, with 14 loadings on the general factor freely estimated per model (see Figure 2). Some indicators’ loadings on the general factor were robust to the

exclusion of other indicators, such as tics (sudden and repetitive vocalizations or motor movements), whereas others' loadings were highly variable, such as conduct disorder, depression, and generalized anxiety. The most variable indicators also happened to be those that were the strongest markers of the general factor in the full model (with all indicators included). One dramatic example is conduct disorder, whose loadings on the general factor ranged from $-.45$ when obsessions were excluded from the general factor to $.63$ when tics were excluded. In sum, excluding some indicators caused the sign of other indicators' loadings on the general factor to flip (see Watts et al., 2019, for a lengthier description). When Watts and colleagues (2019) correlated general factor scores from the full model with those from models where one of the 14 indicators were dropped, correlations were essentially 0 for models that dropped conduct disorder and major depression ($r_s = .01$ and $-.09$, respectively), and strongly negative for models that dropped obsessions and oppositional defiance ($r_s = -.89$ and $-.65$). These findings suggest that when general factors are constructed using even slightly different sets of indicators, they can bear little resemblance to one another.

Consider one final example. Mallard and colleagues (2019) recently claimed to find not one but two polygenic p -factors among the same set of individuals, one defined by self-reported disorders and the other defined by clinician-reported disorders. These two general factors were only modestly correlated ($r = .34$), perhaps because they included different forms of psychopathology. The self-reported general factor included major depressive disorder, other depression symptoms, irritability, mania, and psychosis, whereas the clinician-reported general factor included schizophrenia, bipolar disorder, psychosis, and major depressive disorder. Where there was overlap in terms of disorders onto both self- and clinician-defined general factors, those disorders' loadings on each factor varied considerably. Psychosis' loadings on the clinician- and self-defined general factors were low ($\lambda = .21$) and moderately high ($\lambda = .66$), respectively. Similarly, major depression's loadings were low ($\lambda = .18$) and high ($\lambda = .77$), respectively. Although the authors conclude that self- and clinician-defined general factors arise from relatively distinct genetic liabilities, it is possible that the low correlation between these two general factors arises because of differences in what was included in each general factor. That is, *even when the same individual is being rated*, differences in measurement across raters can produce remarkably different general factors of psychopathology.

So, what is the p -factor? Even referring to it as an entity, as we just did, presupposes that it is relatively well-replicated across studies. The data we reviewed call into question that assumption. Based on the evidence we report here, the commonly used phrase "general factor of psychopathology" could have hundreds of meanings. As such, instead of the " p -factor replicates" or "is supported", it is perhaps more accurate to state that a general factor was "extracted" or "modeled." We prefer "extracted" or "modeled" because a general factor from one study might not resemble a general factor from another.

Issue #4: The *p*-factor is described as a causal mechanism, but latent variable models do not reveal causality.

Many people adopt common cause explanations of the *p*-factor. Latent variables, however, are not tests of the common cause model. Instead, latent variables are statistical summaries of data that take into account the intercorrelations among indicators. Thus, like the study of any correlation, latent variables are completely agnostic with respect to causes of the interrelatedness among indicators. Data can be modeled with a latent variable when a researcher believes the data covary due to some common, indirectly measured cause, but their being modeled in such a way does not mean that there *is* a common cause of those indicators. At best, a latent variable is potentially consistent with, but is not a rigorous test of, a common cause model (Bollen & Lennox, 1991; Boorsbom et al., 2003; Wright, 2020; see VanderWeele & Vansteelandt, 2020, for a more rigorous test).

There are also apparent incompatibilities between certain *p*-factor theories and the use of latent variables. Latent variables assume indicators are fungible, or sampled from a universe of exchangeable indicators (e.g., Eid et al., 2017). This assumption means, then, that the general factor should be robust to the inclusion or exclusion of other indicators. Thus, its composition should not change appreciably. This includes studies that use the same indicators to model a general factor *and* studies that use different indicators. As we demonstrated earlier, general factors of psychopathology can be quite capricious inasmuch as they appear highly sensitive to their contents (Watts et al., 2019).

Moreover, certain theories of the *p*-factor are incompatible with a common cause model. If general factors of psychopathology are simply capturing impairment or severity associated with their contents, the *p*-factor does not reflect a latent entity. In this case, the *p*-factor is not a causal mechanism but an epiphenomenon (or downstream consequence) of psychopathology. Should the *p*-factor capture impairment or severity, it is likely a formative as opposed to reflective construct (e.g., Bainter & Bollen, 2014; Edwards & Bagozzi, 2000; McCoach & Kenny, 2014). As Fried (this issue) argues, it makes little theoretical sense to extract a latent variable for impairment, much as it is nonsensical to construct a latent variable of socioeconomic status.

Conclusion

Early examinations of a general factor of psychopathology marked a watershed moment in quantitative psychopathology and clinical psychology more broadly (Caspi et al., 2014; Lahey et al., 2012). Amid the neo-Kraepelinian era of psychiatric classification (Compton & Guze, 1995), which is epitomized by delineating among putatively discrete forms of psychopathology, the notion that one's liability towards psychopathology can be described by a single dimension is quite radical. At the same time, there are numerous conceptual and statistical reasons to be wary of the *p*-factor literature: it conflates theory and model, it uses a mathematical representation that generally does not bear precisely on its theories, and it promotes theories that are difficult to adjudicate or falsify because they can be all be confirmed under the same model.

We venture that much of the issues we describe stem from an underappreciation of the very theory-model inference gap that Fried (this issue) outlines. In the p -factor literature, theory and model are intertwined (for good reason) but are generally conflated, likely because the lion's share of research on the p -factor uses a single type of model to examine it (i.e., the bifactor model). Although much research has extolled the virtues of the p -factor in terms of its ability to enhance research on the sources, correlates, and even treatment implications of psychopathology (Caspi & Moffit, 2018; Conway et al., 2019; Lahey et al., 2017; Meier & Meier, 2018), we caution against the adoption of a theoretical model that is built on a methodological house of cards. For the field to move towards a more precise understanding of the p -factor, should it exist, it needs to grapple with the issues we raise in this commentary.

Funding:

Douglas Steinley is funded through R01AA024133 (Principal Investigator: Kenneth J. Sher) and R01AA027264 (Principal Investigator: Sean P. Lane/Erin P. Hennes); Sean P. Lane is funded through R01AA027264 (Principal Investigator: Sean P. Lane/Erin P. Hennes).

The authors wish to thank Cassandra E. Boness, Thomas H. Costello, and Kenneth J. Sher for their comments and the interesting conversation.

References

- *An asterisk designates studies that were included in the condensed review.
- Achenbach TM (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs: General and Applied*, 80(7), 1–37. 10.1037/h0093906
- Albott CS, Forbes MK, & Anker JJ (2018). Association of childhood adversity with differential susceptibility of transdiagnostic psychopathology to environmental stress in adulthood. *JAMA Network Open*, 1(7), e185354–e185354. 10.1001/jamanetworkopen.2018.5354 [PubMed: 30646399]
- Bäckström M, Björklund F, & Larsson MR (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality*, 43(3), 335–344. 10.1016/j.jrp.2008.12.013
- Bagby RM, Taylor GJ, Quilty LC, & Parker JDA (2007). Reexamining the factor structure of the 20-Item Toronto Alexithymia Scale: Commentary on Gignac, Palmer, and Stough. *Journal of Personality Assessment*, 89(3), 258–264. 10.1080/00223890701629771 [PubMed: 18001226]
- Bollen K, & Lennox R (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305–314.
- Bonifay W, & Cai L (2017). On the complexity of item response theory models. *Multivariate Behavioral Research*, 52(4), 465–484. 10.1080/00273171.2017.1309262 [PubMed: 28426237]
- Bonifay W, Lane SP, & Reise SP (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5(1), 184–186. 10.1177/2167702616657069
- Borsboom D, Mellenbergh GJ, & Van Heerden J (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203–219. 10.1037/0033-295X.110.2.203 [PubMed: 12747522]
- Borsboom D, Mellenbergh GJ, & Van Heerden J (2004). The concept of validity. *Psychological Review*, 111(4), 1061–1071. 10.1037/0033-295X.111.4.1061 [PubMed: 15482073]
- Bornovalova MA, Choate AM, Fatimah H, Petersen KJ, & Wiernik BM (2020). Appropriate use of bifactor analysis in psychopathology research: Appreciating benefits and limitations. *Biological Psychiatry*, 88(1), 18–27. 10.1016/j.biopsych.2020.01.013 [PubMed: 32199605]
- Brandes CM, Herzhoff K, Smack AJ, & Tackett JL (2019). The p factor and the n factor: Associations between the general factors of psychopathology and neuroticism in children. *Clinical Psychological Science*, 7(6), 1266–1284. 10.1177/2167702619859332

- Campbell DT, & Fiske DW (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. 10.1037/h0046016 [PubMed: 13634291]
- Carver CS, Johnson SL, & Timpano KR (2017). Toward a functional view of the p factor in psychopathology. *Clinical Psychological Science*, 5(5), 880–889. 10.1177/2167702617710037 [PubMed: 29057170]
- *Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, ... & Moffitt TE (2014). The p Factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, 2(2), 119–137. 10.1177/2167702613497473 [PubMed: 25360393]
- Caspi A, & Moffitt TE (2018). All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry*, 175(9), 831–844. 10.1176/appi.ajp.2018.17121383
- Castellanos-Ryan N, Brière FN, O’Leary-Barrett M, Banaschewski T, Bokde A, ... & Consortium TI (2016). The structure of psychopathology in adolescence and its common personality and cognitive correlates. *Journal of Abnormal Psychology*, 125(8), 1039–1052. 10.1037/abn0000193 [PubMed: 27819466]
- Chang L, Connelly BS, & Geeza AA (2012). Separating method factors and higher order traits of the Big Five: A meta-analytic multitrait–multimethod approach. *Journal of Personality and Social Psychology*, 102(2), 408–426. 10.1037/a0025559 [PubMed: 21967007]
- Compton WM, & Guze SB (1995). The neo-Kraepelinian revolution in psychiatric diagnosis. *European Archives of Psychiatry and Clinical Neuroscience*, 245(4–5), 196–201. 10.1007/BF02191797 [PubMed: 7578281]
- Conway CC, Forbes MK, Forbush KT, Fried EI, Hallquist MN, ... & Krueger RF, ... Eaton NR (2019). A hierarchical taxonomy of psychopathology can transform mental health research. *Perspectives in Psychological Science*, 14(3), 419–436. 10.1177/1745691618810696
- *Conway CC, Mansolf M, & Reise SP (2019). Ecological validity of a quantitative classification system for mental illness in treatment-seeking adults. *Psychological Assessment*, 31(6), 730–740. 10.1037/pas0000695 [PubMed: 30667268]
- Couwenbergh C, van den Brink W, Zwart K, Vreugdenhil C, van Wijngaarden-Cremers P, & van der Gaag RJ (2006). Comorbid psychopathology in adolescents and young adults treated for substance use disorders. *European Child & Adolescent Psychiatry*, 15(6), 319–328. 10.1007/s00787-006-0535-6 [PubMed: 16648966]
- Davies SE, Connelly BS, Ones DS, & Birkland AS (2015). The general factor of personality: The “Big One,” a self-evaluative trait, or a methodological gnat that won’t go away? *Personality and Individual Differences*, 81, 13–22. 10.1016/j.paid.2015.01.006
- DeYoung CG (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, 91(6), 1138–1151. 10.1037/0022-3514.91.6.1138 [PubMed: 17144770]
- Eid M, Geiser C, Koch T, & Heene M (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, 22(3), 541–562. 10.1037/met0000083 [PubMed: 27732052]
- Forbes MK, Greene AL, Levin-Aspenson HF, Watts AL, Hallquist M, ... & Krueger RF (under review). A comparison of the reliability and validity of the predominant models used in research on the empirical structure of psychopathology. 10.31219/osf.io/fhp2r
- *Gluschkoff K, Jokela M, & Rosenström T (2019). The general psychopathology factor: structural stability and generalizability to within-individual changes. *Frontiers in Psychiatry*, 10, 594. 10.3389/fpsy.2019.00594 [PubMed: 31543833]
- Greene AL, Eaton NR, Li K, Forbes MK, Krueger RF, ... & Kotov R (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *Journal of Abnormal Psychology*, 128(7), 740–764. 10.1037/abn0000434 [PubMed: 31318246]
- *Harden KP, Engelhardt LE, Mann FD, Patterson MW, Grotzinger AD, Savicki SL, Thibodeaux ML, Freis SM, Tackett JL, Church JA, & Tucker-Drob EM (2020). Genetic associations between executive functions and a general factor of psychopathology. *Journal of the American Academy of Child & Adolescent Psychiatry*, 59(6), 749–758. 10.1016/j.jaac.2019.05.006 [PubMed: 31102652]
- *He Q, & Li JJ (2019, 5 22). Factorial invariance in hierarchical factor models of mental disorders in African American and European American youths. 10.31234/osf.io/bw2gy

- Hoffman M, Steinley D, Trull TJ, & Sher KJ (2019). Estimating transdiagnostic symptom networks: The problem of “skip outs” in diagnostic interviews. *Psychological Assessment*, 31(1), 73–81. 10.1037/pas0000644 [PubMed: 30160496]
- Horn RA, & Johnson CR (2012). *Matrix analysis* (2nd ed). Cambridge University Press.
- Kaczurkin AN, Moore TM, Calkins ME, Ciric R, Detre JA, Elliott MA, Foa EB, de La Garza AG, Roalf DR, Rosen A, Ruparel K, Shinohara RT, Xia CH, Wolf DH, Gur RE, Gur RC, & Satterthwaite TD (2018). Common and dissociable regional cerebral blood flow differences associate with dimensions of psychopathology across categorical diagnoses. *Molecular Psychiatry*, 23(10), 1981–1989. 10.1038/mp.2017.174 [PubMed: 28924181]
- Kane M (2013). Articulating a validity argument In *The Routledge handbook of language testing* (pp. 48–61). Routledge.
- Kessler RC, Chiu WT, Demler O, & Walters EE (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 617–627. 10.1001/archpsyc.62.6.617 [PubMed: 15939839]
- Keyes KM, Eaton NR, Krueger RF, Skodol AE, Wall MM, Grant B, Siever LJ, & Hasin DS (2013). Thought disorder in the meta-structure of psychopathology. *Psychological Medicine*, 43(8), 1673–1683. 10.1017/S0033291712002292 [PubMed: 23171498]
- Kim H, Greene A, Eaton NR, Lerner M, & Gadow K (2019). Mr. Kim et al. reply *Journal of the American Academy of Child and Adolescent Psychiatry*, 58(10), 1021–1025. 10.1016/j.jaac.2019.04.024
- Kotov R, Ruggero CJ, Krueger RF, Watson D, Yuan Q, & Zimmerman M (2011). New dimensions in the quantitative classification of mental illness. *Archives of General Psychiatry*, 68(10), 1003–1011. 10.1001/archgenpsychiatry.2011.107 [PubMed: 21969458]
- Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, ... & Zimmerman M (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454–477. 10.1037/abn0000258 [PubMed: 28333488]
- Krueger RF (1999). The structure of common mental disorders. *Archives of General Psychiatry*, 56(10), 921–926. 10.1001/archpsyc.56.10.921 [PubMed: 10530634]
- Krueger RF, & Markon KE (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology*, 2(1), 111–133. 10.1146/annurev.clinpsy.2.022305.095213
- *Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, & Rathouz PJ (2012). Is there a general factor of prevalent psychopathology during adulthood? *Journal of Abnormal Psychology*, 121(4), 971–977. 10.1037/a0028355 [PubMed: 22845652]
- Lahey BB, Krueger RF, Rathouz PJ, Waldman ID, & Zald DH (2017). A hierarchical causal taxonomy of psychopathology across the life span. *Psychological Bulletin*, 143(2), 142–186. 10.1037/bul0000069 [PubMed: 28004947]
- *Lahey BB, Zald DH, Perkins SF, Villalta-Gil V, Werts KB, Van Hulle CA, ... & Watts AL (2018). Measuring the hierarchical general factor model of psychopathology in young adults. *International Journal of Methods in Psychiatric Research*, 27(1), e1593 10.1002/mpr.1593
- Lane SP, Steinley D, & Sher KJ (2016). Meta-analysis of DSM alcohol use disorder criteria severities: Structural consistency is only ‘skin deep.’ *Psychological Medicine*, 46(8), 1769–1784. 10.1017/S0033291716000404 [PubMed: 27019218]
- Littlefield AK, Lane SP, Gette JA, Watts AL, & Sher KJ (in press). The “Big Everything”: Integrating and investigating dimensional models of psychopathology, personality, and intelligence. *Personality Disorders: Theory, Research, & Treatment*.
- Mallard TT, Linnér RK, Okbay A, Grotzinger AD, de Vlaming R, Meddens SFW, Tucker-Drob EM, Kendler KS, Keller MC, Koellinger PD, & Harden KP (2019). Not just one p: Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *BioRxiv*, 603134 10.1101/603134
- Martel MM, Pan PM, Hoffmann MS, Gadelha A, Rosário M. C. do, Mari JJ, Manfro GG, Miguel EC, Paus T, Bressan RA, Rohde LA, & Salum GA (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. 10.1037/abn0000205 [PubMed: 27748619]
- Maydeu-Olivares A, & Coffman DL (2006). Random intercept item factor analysis. *Psychological Methods*, 11(4), 344–362. 10.1037/1082-989X.11.4.344 [PubMed: 17154751]
- Meier MA, & Meier MH (2018). Clinical implications of a general psychopathology factor: A cognitive-behavioral transdiagnostic group treatment for community mental health. *Journal of Psychotherapy Integration*, 28, 253–268. 10.1037/int0000095
- Morgan GB, Hodge KJ, Wells KE, & Watkins MW (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. 10.3390/jintelligence3010002
- Murray AL, & 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. 10.1016/j.intell.2013.06.004
- *Noordhof A, Krueger RF, Ormel J, Oldehinkel AJ, & Hartman CA (2015). Integrating autism-related symptoms into the dimensional internalizing and externalizing model of psychopathology. The TRAILS Study. *Journal of Abnormal Child Psychology*, 43(3), 577–587. 10.1007/s10802-014-9923-4 [PubMed: 25099360]
- Nuzum H, Shapiro JL, & Clark LA (2019). Affect, behavior, and cognition in personality and functioning: An item-content approach to clarifying empirical overlap. *Psychological Assessment*, 31(7), 905–912. 10.1037/pas0000712 [PubMed: 30869967]
- *Olino TM, Bufferd SJ, Dougherty LR, Dyson MW, Carlson GA, & Klein DN (2018). The development of latent dimensions of psychopathology across early childhood: Stability of dimensions and moderators of change. *Journal of Abnormal Child Psychology*, 46(7), 1373–1383. 10.1007/s10802-018-0398-6 [PubMed: 29359267]
- *Pezzoli P, Antfolk J, & Santtila P (2017). Phenotypic factor analysis of psychopathology reveals a new body-related transdiagnostic factor. *PLoS One*, 12, e0177674 10.1371/journal.pone.0177674 [PubMed: 28542328]
- Popper K (1963). *Conjectures and refutations: The growth of scientific knowledge*. London, UK: Routledge.
- Raykov T, & Marcoulides GA (2001). Can there be infinitely many models equivalent to a given covariance structure model? *Structural Equation Modeling: A Multidisciplinary Journal*, 8(1), 142–149. 10.1207/S15328007SEM0801_8
- Reise SP (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667–696. 10.1080/00273171.2012.715555 [PubMed: 24049214]
- Reise SP, Kim DS, Mansolf M, & Widaman KF (2016). Is the bifactor model a better model or is it just better at modeling implausible responses? Application of iteratively reweighted least squares to the Rosenberg Self-Esteem Scale. *Multivariate Behavioral Research*, 0–0. 10.1080/00273171.2016.1243461
- Revelle W, & Wilt J (2013). The general factor of personality: A general critique. *Journal of Research in Personality*, 47(5), 493–504. 10.1016/j.jrp.2013.04.012 [PubMed: 23956474]
- Roberts S, & Pashler H (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107(2), 358–367. 10.1037/110033-295X.107i358 [PubMed: 10789200]
- Rodriguez A, Reise SP, & Haviland MG (2016a). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137–150. 10.1037/met0000045 [PubMed: 26523435]
- Rodriguez A, Reise SP, & Haviland MG (2016b). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98(3), 223–237. 10.1080/00223891.2015.1089249 [PubMed: 26514921]
- Romer AL, Elliott ML, Knodt AR, Sison ML, Ireland D, Houts R, ... & Hariri AR (2019). A pervasively thinner neocortex is a transdiagnostic feature of general psychopathology. *BioRxiv*, 788232 10.1101/788232
- *Romer AL, Knodt AR, Sison ML, Ireland D, Houts R, Ramrakha S, ... & Hariri AR (2019). Replicability of structural brain alterations associated with general psychopathology: Evidence

from a population-representative birth cohort. *Molecular Psychiatry*, 1–8. 10.1038/s41380-019-0621-z

- Sellbom M, & Tellegen A (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, 31(12), 1428–1441. 10.1037/pas0000623 [PubMed: 31120298]
- *Shields AN, Reardon KW, Brandes CM, & Tackett JL (2019). The p factor in children: Relationships with executive functions and effortful control. *Journal of Research in Personality*, 82, 103853. 10.1016/j.jrp.2019.103853
- Smith GT, Atkinson EA, Davis HA, Riley EN, & Oltmanns JR (2020). The general factor of psychopathology. *Annual Review of Clinical Psychology*, 16(1), 75–98. 10.1146/annurev-clinpsy-071119-115848
- *Snyder HR, Friedman NP, & Hankin BL (2019). Transdiagnostic mechanisms of psychopathology in youth: Executive functions, dependent stress, and rumination. *Cognitive Therapy and Research*, 43(5), 834–851. 10.1007/s10608-019-10016-z [PubMed: 31551642]
- *Tackett JL, Lahey BB, Hulle CV, Waldman I, Krueger RF, & Rathouz PJ (2013). Common genetic influences on negative emotionality and a general psychopathology factor in childhood and adolescence. *Journal of Abnormal Psychology*, 122(4), 1142–1153. 10.1037/a0034151 [PubMed: 24364617]
- Tomarken AJ, & Waller NG (2003). Potential problems with “well fitting” models. *Journal of Abnormal Psychology*, 112(4), 578–598. 10.1037/0021-843X.112.4.578 [PubMed: 14674870]
- VanderWeele TJ, & Vansteelandt S (2020). A statistical test to reject the structural interpretation of a latent factor model. arXiv preprint arXiv:2006.15899.
- Vanheule S, Desmet M, Groenvynck H, Rosseel Y, & Fontaine J (2008). The factor structure of the Beck Depression Inventory–II: An evaluation. *Assessment*, 15(2), 177–187. 10.1177/1073191107311261 [PubMed: 18182478]
- Waldman ID, Poore H, van Hulle C, Rathouz P, & Lahey BB (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. 10.1037/abn0000183 [PubMed: 27819467]
- *Watts AL, Poore HE, & Waldman ID (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science*, 7(6), 1285–1303. 10.1177/2167702619855035
- Westen D, & Rosenthal R (2004). Improving construct validity: Cronbach, Meehl, and Neurath’s ship: Comment. *Psychological Assessment*, 17(4), 409–412. 10.1037/1040-3590.17.4.409
- Widiger TA, & Oltmanns JR (2017). The general factor of psychopathology and personality. *Clinical Psychological Science: A Journal of the Association for Psychological Science*, 5(1), 182–183. 10.1177/2167702616657042 [PubMed: 30416877]
- Wright AGC (in press). Latent variable models in clinical psychology In Wright AGC, & Hallquist MN (Eds.). *Cambridge handbook of research methods in clinical psychology*. New York, NY: Cambridge University Press.

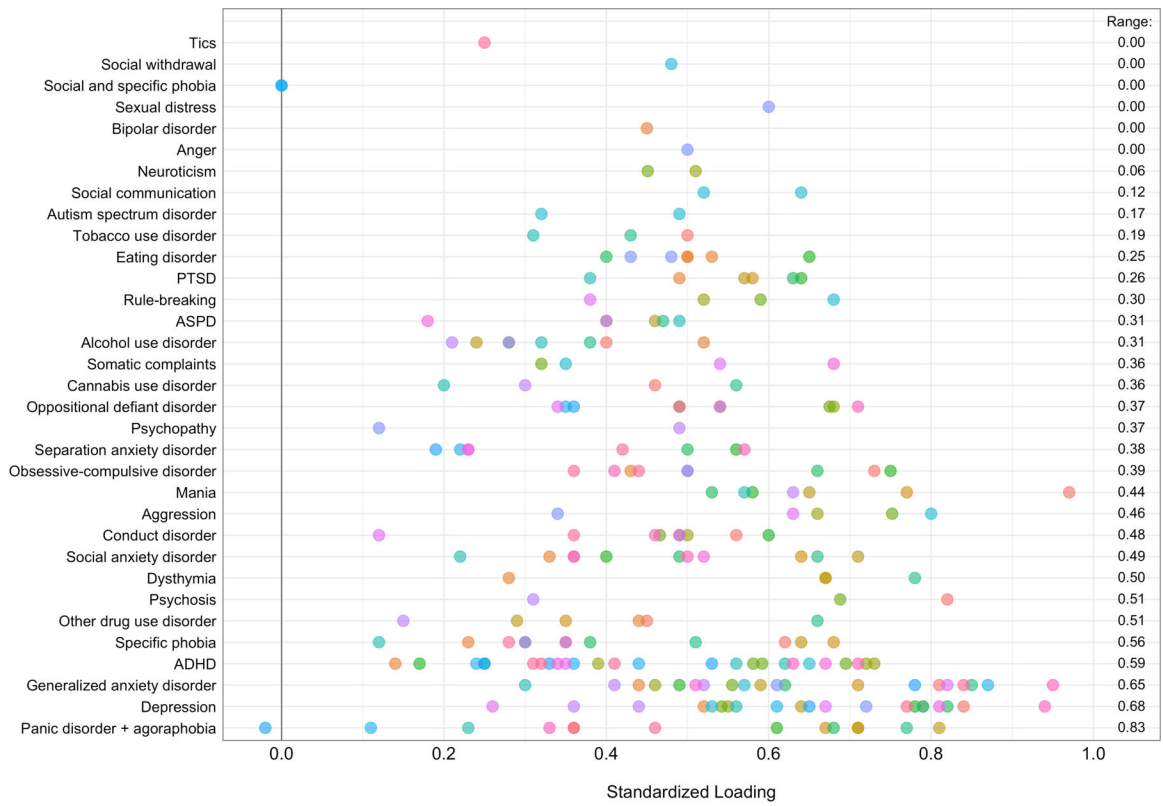


Figure 1. Standardized factor loadings on general factors of psychopathology across studies included in the condensed review.

Note. Each color pertains to a factor loading from a study. Some psychopathology indicators have multiple indicators from the same study if the study included multiple indicators of the same construct, or if the study included multiple samples. The rightmost column displays the range in factor loadings across studies. ADHD=attention-deficit/hyperactivity disorder; ASPD=antisocial personality disorder; PTSD=posttraumatic stress disorder.



Figure 2. Variability in symptom dimensions’ loadings on the general factor when other symptom dimensions are excluded from the model.
Note. Each color pertains to different model in which one symptom was dropped from the general factor. The asterisks depict each symptom dimension’s loading in the full model. The data presented along the bottom of the plot depict the range in factor loadings onto the general factor DEP=major depression; CND=conduct disorder; GAD=generalized anxiety disorder; OBS=obsessions; AGO=agoraphobia; PAN=panic disorder; SAD=separation anxiety disorder; COM=compulsions; SOC=social phobia; INA=inattention; SPH=specific phobia; ODD=oppositional defiant disorder; HYP=hyperactivity; IMP=impulsivity; TIC=tics.

Table 1.

Studies included in their condensed review.

Symptom dimension	Caspi et al. (2014)	Conway et al. (2019)	Gluschkoff et al. (2019)	Harden et al. (2020)	He & Li (2020)	Lahey et al. (2012)	Lahey et al. (2018)	Noordhof et al. (2015)	Olino et al. (2018)	Pezzoli et al. (2017)	Romer et al. (2019)	Shields et al. (2019)	Snyder et al. (2019)	Tackett et al. (2013)	Watts et al. (2019)	% models included
Attention-deficit/hyperactivity disorder		X		X	X		X	X	X			X	X	X	X	67
Aggression			X					X		X		X				27
Alcohol use disorder	X	X	X			X	X			X	X					47
Anger										X						7
Antisocial personality disorder			X			X	X				X					33
Autism spectrum disorder								X								7
Bipolar disorder		X														7
Cannabis use disorder	X					X	X				X					27
Conduct disorder	X			X	X		X	X			X	X	X	X	X	47
Depression	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	100
Dysthymia	X	X	X			X										20
Eating disorder		X			X					X						20
Generalized anxiety disorder	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	100
Mania	X		X		X		X				X					33
Neuroticism				X												7
Obsessive-compulsive disorder	X	X			X		X				X			X		40
Oppositional defiant disorder				X	X				X			X	X	X	X	47
Other drug use disorder	X	X	X			X					X					33

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Caspi et al. (2014)	Conway et al. (2019)	Gluschkoff et al. (2019)	Harden et al. (2020)	He & Li (2020)	Lahey et al. (2012)	Lahey et al. (2018)	Noordhof et al. (2015)	Olino et al. (2018)	Pezzoli et al. (2017)	Romer et al. (2019)	Shields et al. (2019)	Snyder et al. (2019)	Tackett et al. (2013)	Watts et al. (2019)	
Panic disorder + agoraphobia		X	X	X	X	X	X		X					X	X	53
Psychopathy										X	X					13
Psychosis/ thought problems	X		X	X							X					20
Post-traumatic stress disorder		X	X		X	X										27
Rule-breaking				X				X				X				20
Separation anxiety disorder				X	X				X			X	X	X	X	40
Sexual distress										X						7
Social and specific phobia								X								7
Social anxiety disorder		X	X	X	X	X	X					X	X	X	X	60
Social communication								X								7
Social withdrawal								X								7
Somatic complaints				X				X				X	X			27
Specific phobia	X	X	X		X	X	X	X			X			X	X	67
Tics															X	7
Tobacco use disorder	X					X	X									20
Range in factor loadings	.57	.39	.51, .57	.34, .43	.61, .57	.47	.50	.48	.87, .80	.60	.48	.70	.63	.62	.59	

Note. This figure displays the studies included in the condensed review. The leftmost column includes a list of each psychopathology indicator used to test a general factor of psychopathology, with its use in each study denoted with an X. The rightmost column displays the percent of studies included in the review that included each psychopathology indicator. The bottom row lists the range in factor loadings on the general factor for each study. A study might include two ranges in factor loadings if it reported more than one sample. Due to low representation of each subcategory across studies, attention-deficit/hyperactivity disorder combined attention problems/inattention, hyperactivity, and impulsivity assessed within the context of attention problems, Panic + agoraphobia combined panic disorder, panic disorder with agoraphobia, and agoraphobia. Psychosis combined schizophrenia and thought problems.