A critical evaluation of the p-factor literature

Watts, A.L.; Greene, A.L.; Bonifay, W.; Fried, E.I.

nature reviews psychology

https://doi.org/10.1038/s44159-023-00260-2

Perspective

Check for updates

A critical evaluation of the *p*-factor literature

Ashley L. Watts[1,2] ✉, Ashley L. Greene[3,4], Wes Bonifay[5] & Eiko I. Fried[6]

Abstract

The *p*-factor is a construct that is thought to explain and maybe even cause variation in all forms of psychopathology. Since its 'discovery' in 2012, hundreds of studies have been dedicated to the extraction and validation of statistical instantiations of the *p*-factor, called general factors of psychopathology. In this Perspective, we outline five major challenges in the *p*-factor literature, namely that it: mistakenly equates good model fit with validity; corroborates weak *p*-factor theories through underspecified construct validation efforts; produces poorly replicated general factors of psychopathology; violates assumptions of latent variable models; and reifies general factors of psychopathology as latent, causal entities. In turn, the *p*-factor literature neglects alternative models that are incompatible with the notion that a single dimension adequately summarizes variation in all forms of psychopathology. These challenges raise questions about substantive interpretations of the *p*-factor, undermining confidence that the *p*-factor is a real, latent entity, or that general factors of psychopathology are useful summaries of psychopathology variation. We conclude with ways to move forward, in the spirit of strengthening the *p*-factor literature and improving psychopathology classification, treatment and prevention across the lifespan.

Sections

[1]Department of Psychological Sciences, University of Missouri, Columbia, MO, USA. [2]Department of Psychological Sciences, Vanderbilt University, Nashville, TN, USA. [3]VISN 2 Mental Illness Research, Education and Clinical Center, James J. Peters VA Medical Center, Bronx, NY, USA. [4]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [5]Department of Educational, School, and Counseling Psychology, University of Missouri, Columbia, MO, USA. [6]Unit of Clinical Psychology, Faculty of Social and Behavioral Sciences, Leiden University, Leiden, NL, USA. ✉e-mail: ashleylwatts@gmail.com

Introduction

The *p*-factor is a construct that is thought to describe and maybe even cause variation in all forms of psychopathology[1–3]. Over the past decade, hundreds of studies have been dedicated to the extraction and validation of statistical instances of the so-called *p*-factor, which we refer to as general factors of psychopathology[4,5] to avoid conflating theoretical and statistical constructs, respectively. Despite the *p*-factor's relatively young age, it has achieved near-omnipresence in the psychopathology field, by way of its proponents as well as its skeptics, whose respective research has largely proceeded within two silos.

On the one hand, *p*-factor proponents have gathered empirical support for its structural and construct validity[3,6,7]. To date, general factors of psychopathology have been extracted across most major age groups and numerous cultures. Proponents conclude that general factors of psychopathology are relatively stable across important developmental periods[8–10]; moderately heritable[11–14]; associated with various neural substrates, such as increased blood flow to the anterior cingulate cortex and alterations in cerebellar circuitry[15,16]; and concurrently and prospectively associated with important life outcomes, including suicidality[1,17,18], criminal behaviour and convictions[1,18], employment termination[19], executive dysfunction[13,20], poorer academic performance and intelligence[1,18,21], and reduced foetal growth[22]. In these studies, *p*-factor researchers typically conclude that scores on a single dimension can quantify and even explain a person's liability towards mental disorder broadly construed — including the extent of comorbidity, severity and chronicity of psychiatric phenomena across their lifespan[1–3,6].

On the other hand, *p*-factor skeptics have raised substantial concerns regarding the statistical nature of general factors of psychopathology and the theoretical nature of the *p*-factor[4,5,23–28]. Quite explicitly, general factors of psychopathology were dubbed the *p*-factor as an explicit homage to theoretical constructs in other domains[1,6], most directly Spearman's general factor of intelligence (also known as *g*)[29]. As skeptics see it, *p*-factor researchers' allusions to *g* divert attention away from *g*'s long history of well demonstrated limitations and criticisms, including of the methods used to model *g*. In fact, although the dust is far from settled on the validity and utility of *g*, historical and contemporary intelligence scholars suggest that it is little more than an unfalsifiable product of factor analysis that has been erroneously reified as a latent, causal dimension[30–32]. Such criticisms of *g* (among others) also apply to the *p*-factor.

In this Perspective, we aim to go beyond existing reviews of the *p*-factor that have largely adjudicated its potential substantive interpretations and overlooked its limitations and criticisms[3,6,7] by bridging gaps between the abovementioned *p*-factor research silos (for another review of the limitations of substantive interpretations of general factors of psychopathology, see ref. 27). Specifically, we argue that the only way the *p*-factor literature can move forward is to integrate substantive interest in the *p*-factor with a balanced understanding of the theoretical and statistical issues that plague the literature. To that end, we synthesize competing perspectives on the *p*-factor and provide researchers with additional context for justified skepticism. After introducing the *p*-factor and its purported theoretical meanings, we outline concerns with the statistical nature of general factors of psychopathology and the theoretical nature of the *p*-factor. Finally, we offer ways to move forward while highlighting important limitations of current practices that hinder progress on the classification, treatment and prevention of psychopathology across the lifespan.

Overview of the *p*-factor

In this section, we summarize the statistical practices in the psychopathology literature that gave rise to the *p*-factor, as well as the *p*-factor theories that followed the adoption and widespread use of statistical models that directly incorporate a general factor of psychopathology.

Statistical development

Some sixty years ago, Menninger and colleagues[33] claimed that there is only one mental illness: all mental disorders, although superficially dissimilar in terms of their expressions, are fundamentally related and differ only in terms of severity. Although Menninger's bold supposition was largely overlooked for 50 years, the sentiment behind this assertion arguably presaged the notion of a *p*-factor. Indeed, amid neo-Kraepelinian efforts to demarcate putatively distinct forms of psychopathology (such as dementia praecox and manic depression), a practice termed splitting, Menninger[33] offered a radical alternative that has appealed to contemporary psychopathologists:[6] lumping[34]. A natural but unfortunate ramification of splitting was the observation of rampant co-occurrence (or comorbidity) among presumably non-overlapping disorders. That is, splitting resulted in drawing arbitrary borders around psychiatry's favoured disorders, so much so that comorbidity is now increasingly regarded as the rule rather than the exception in psychiatric classification[6].

Much contemporary research has focused on the causes of psychiatric comorbidity, with Achenbach introducing a major, substantive cause: transdiagnostic dimensions of psychopathology. Achenbach observed[35] that conditions characterized by high negative emotionality (for example, major depression, generalized anxiety and phobias) covaried strongly, as did conditions characterized by low behavioural and emotional control (for example, oppositional defiant disorder, antisocial personality disorder and attention-deficit hyperactivity disorder). He termed the shared variation among these sets of conditions internalizing and externalizing, respectively[36–38]. The study of internalizing and externalizing marked a sea change in psychopathology research, in part because Achenbach's work offered olive branches to both lumpers and splitters[34]. In conceptualizations of psychopathology as hierarchically arranged[35,39], conditions housed under broader dimensions can (and do) share features, and they can also have unique or dissociable features.

Achenbach's early models paved the way for groundbreaking research on the causes and consequences of transdiagnostic dimensions of psychopathology, but those models are incomplete representations of psychopathology's structure[40]. First, other dimensions are necessary to describe individual differences in psychopathology, such as thought disorder, detachment and potentially somatoform[39]. Second, the correlation between internalizing and externalizing dimensions is sufficiently large (about 0.5 to 0.6)[36,41] to be potentially caused by a superordinate dimension.

Seminal works on the *p*-factor explicitly sought to leverage the covariation among psychopathology dimensions[1,2] observed in correlated factors models (Fig. 1a) by decomposing it into multiple levels of breadth. To do so, researchers typically use a bifactor model, which decomposes the variance of items into two sets of factors: a general factor of psychopathology that represents the shared variance among all forms of psychopathology included in the model, and several specific factors that capture additional covariance shared by subsets of psychopathology (Fig. 1b). When modelling a general factor of psychopathology with a bifactor model, researchers tend to find that general factors of psychopathology explain anywhere from around 50%[42] to 90%[22,43,44]

of the variance in psychopathology indicators (for an exception, see ref. 23). Such findings have been taken as evidence of the *p*-factor. Consequently, many researchers argued[1–3,6] that the dimensions that were once thought to perch at the summit of the psychopathology hierarchy (for example, externalizing, internalizing) might need to be taken down a peg.

Substantive interpretations

The rapid extraction of general factors of psychopathology and the *p*-factor literature was not kicked off by theory, but rather by the observation that a bifactor model of psychopathology provided a better fit to the data, on average, than a model without a general factor of psychopathology. Over time, after concluding that a bifactor model of psychopathology is a valid if not optimal representation of psychopathology's structure, researchers developed post-hoc explanations of the *p*-factor that emphasize the substantive cause or causes of statistical general factors of psychopathology. Usually, researchers assume that the *p*-factor reflects some latent cause, although there is no strong consensus on the specific mechanisms involved.

The meaning of the *p*-factor is sometimes based on an interpretation of the specific magnitudes of model parameters in general factors of psychopathology (structural validity) or on a general factor of psychopathology's associations with external criteria (construct validity). Either way, researchers have proposed numerous substantive explanations of the *p*-factor, including: unspecified causal mechanisms[2,7], deficits in intellectual functioning[6], disordered thought[1,6], negative emotionality[11,45] and emotion dysregulation[46]. We consider each of these explanations interesting and expand on them below, but we acknowledge that they are, at best, weak theories[47,48]. Each of the following explanations is fledgling, arguably underspecified, is difficult to falsify and has not yet been subjected to risky or otherwise scrutinous tests[4,5,49].

Unspecified causal mechanisms. The most general theory of the *p*-factor is that it captures causal, albeit unspecified (or nonspecific), mechanisms[2,3,50]. That is, general factor of psychopathology scores are interpreted as quantifying risk or liability for every form of psychopathology by capturing their shared causal or etiologic processes. Some *p*-factor studies further claim support for a "generalist genes, specialist environments"[51] model, gesturing to the observation that general factors of psychopathology explain a relatively high degree of genetic variation (about 50%) and a low degree of environmental variation in all of their indicators. These findings have inspired claims that most genetic variation in individual mental disorders is shared with all others[11,12,52]. A critical limitation of this explanation of the *p*-factor is that it is loosely defined and inherently flexible, so much so that it is virtually unfalsifiable; this theory can be corroborated by essentially modelling any form of psychopathology. Indeed, it is unclear what empirical evidence could not be interpreted as support for a causal *p*-factor.

Intellectual functioning. Another theory proposes that the *p*-factor reflects low intellectual functioning[6], in part because some general factors of psychopathology are associated with low cognitive functioning broadly construed[1,20,53]. Nevertheless, this theory has not gained considerable traction, in part because correlations between general factors of psychopathology and performance on intelligence and other cognitive tasks are small (*r* values of 0.1 to 0.2)[1]. Explaining less than 4% of the variance in a general factor of psychopathology is hardly likely to qualify as robust support for a particular causal mechanism[54].

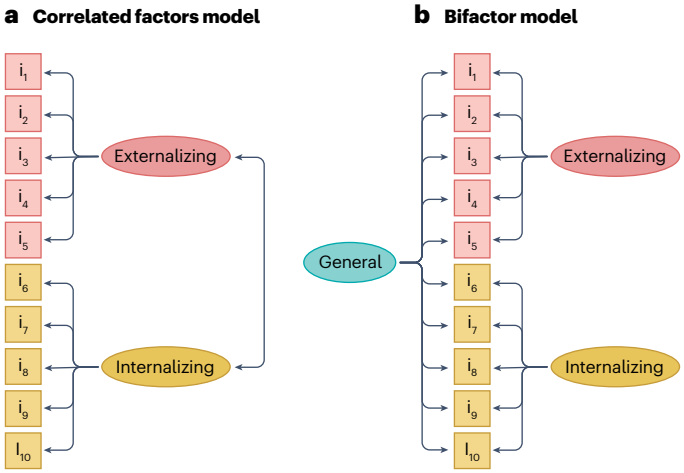


Fig. 1 | Frequntly used structural models of psychopathology. a, A correlated factors model includes separate but correlated dimensions. This example includes externalizing and internalizing dimensions, but the model could include other dimensions, such as thought disorder[1] or somatoform[38], and other dimensions can be split apart. For instance, popular models split internalizing into distress and fears[2,4], and others split externalizing into disinhibited and antagonistic forms. The extent of covariation among dimensions in the correlated factors model has been taken as indirect evidence of a general factor of psychopathology[1,2]. **b**, The bifactor model directly models a general factor of psychopathology. Bifactor models include a general factor onto which all indicators are allowed to load, and a number of specific factors that include subsets of indicators. The general factor is set to be uncorrelated with specific factors in order to fully disentangle general from specific sources of variance. Traditionally, the specific factors are further constrained to be uncorrelated with one another. This modelling assumption is not a mathematical necessity, but an a priori decision based on theory. That is, specific factors are often constrained to be uncorrelated with each other because researchers presume that the general factor accounts for the covariation between factors in the correlated factors model. Thus, although specific factors in a bifactor model are typically labelled using the same terms as dimensions in a correlated factors model, they differ conceptually and empirically. For instance, externalizing in a bifactor model reflects the variance attributable to externalizing once the variance it shares with the general factor has been removed. It is possible to allow specific factors in the bifactor model to correlate with one another, although allowing for all possible correlations among specific factors tends to weaken the general factor, and such correlations should be modelled according to theory[4].

Disordered thought. Indicators of psychosis tend to load most strongly onto general factors of psychopathology when they are included in them, so much so that indicators of disordered thought defined early general factors of psychopathology[1]. This observation led to the theory that the *p*-factor reflects disordered thought[1,6]. This theory's champions argue that cognitive distortion permeates all forms of psychopathology[1,6]. Delusions and hallucinations constitute hallmarks of psychosis, but potentially subtler manifestations of cognitive distortion arise in numerous other forms of psychopathology, including irrational fears in generalized anxiety and specific phobias, automatic negative thoughts in depression, traumatic reexperiencing in post-traumatic stress disorder and intrusive thoughts in obsessive–compulsive disorder. Nevertheless, such an idea rests on two major and, so far, unsubstantiated assumptions: that delusions and hallucinations constitute the extreme end of a dimension upon which subtler forms

of cognitive distortion lie, and that the same mechanisms that cause severe psychotic delusions and hallucinations also cause other forms of cognitive distortion.

Negative emotionality. When indicators of thought disorder are not included in the model, major depression and generalized anxiety tend to load most strongly (even nearly perfectly) onto some general factors of psychopathology[2,4,19]. Further, some general factors of psychopathology are highly correlated with a latent negative emotionality factor[45,54], and much of the phenotypic variation shared between a general factor of psychopathology and negative emotionality has been attributed to shared additive genetic variation[11]. Supporting the possibility that the *p*-factor reflects negative emotionality, negative emotionality is associated with a broad array of psychopathology[55] and is one of the most studied correlates of general factors of psychopathology[1,45,54,56–61]. Even so, it is unsurprising that general factors of psychopathology are highly correlated with negative emotionality[45,54] because internalizing psychopathology captures emotions (such as sadness, anxiety, dysphoria and irritability) that are also assessed in negative emotionality measures. Thus, existing tests of this theory are essentially guaranteed to support it because they include negative emotionality on both sides of the regression equation (that is, in both the independent and dependent variables).

Emotion dysregulation. One final hypothesis is that poor emotional control defines the *p*-factor[46]. Impulsive responses to negative (as opposed to positive) affect are implicated in a wide array of psychopathology[62]. To test this theory, others have relied on proxies of emotion dysregulation (such as impulsivity, low conscientiousness and poor response inhibition on laboratory tasks), which are sometimes correlated with general factors of psychopathology[1,4,20,53]. Although numerous forms of psychopathology are indeed associated with impulsive responses to negative affect, it is unclear how such a construct is relevant to behaviours marked by emotional overcontrol as opposed to undercontrol, such as food restriction in anorexia nervosa[63] or delayed discounting in obsessive–compulsive personality disorder[64].

Challenges in the p-factor literature

General factors of psychopathology are products of factor analysis (or related measurement models) that might or might not adequately represent the theoretical *p*-factor. Although theoretical models are routinely tested with statistical ones, the closely intertwined relationship between the two has created some important challenges for the statistical meaning of general factors of psychopathology and interpretations of the substantive *p*-factor.

Equating good fit and validity

Researchers have used a variety of statistical techniques and models to adjudicate competing accounts of the structure of psychopathology. The *p*-factor literature has largely fixated on bifactor models — Table 1 shows alternative modelling strategies — in large part because such models, which include a general factor of psychopathology, tend to fit observed data better than models without one[1,2]. But any such 'fit contest'[65] that includes a bifactor model is akin to rolling a pair of loaded dice: numerous simulation studies have demonstrated that the bifactor model has qualities that make it more likely than its competitors — including other factor models — to fit any data, including randomly generated and implausible data[65–68]. Consequently, the bifactor model has demonstrated superior fit even when the true data-generating mechanism corresponds to an alternative structure[66,67].

Key to interpreting these simulation studies is recognizing that the bifactor model has a relatively high fitting propensity, or an inherent tendency to accommodate a wide range of possible data patterns. Fitting propensity includes parametric complexity (the number of parameters in the model) and configural complexity (the particular configuration of those parameters)[69]. On balance, models with more complexity will be more adept at fitting data[70] (Box 1).

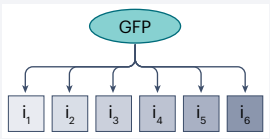
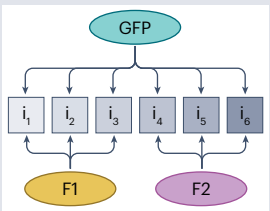
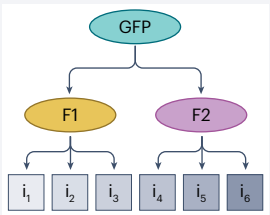
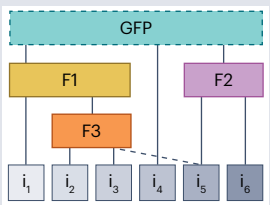
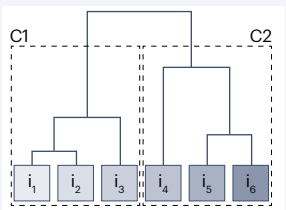
Researchers routinely mitigate parametric complexity by relying on relative fit statistics (such as the Bayesian Information Criteria (BIC))[71,72] that penalize good fit if it comes at the cost of more parameters. Nevertheless, parametric complexity alone does not account for all sources of bifactor model complexity. Yet, *p*-factor researchers have argued that the use of BIC avoids selecting the most complex model (for example, "The general bifactor model fit [sic] also fit somewhat better according to BIC, even though BIC penalizes less parsimonious models to avoid the most complex model always being the best fitting")[2], which falsely equates fitting propensity with parametric complexity and overlooks configural complexity. Fit indices that penalize parametric complexity (such as BIC) are consistently biased towards more complex models[66], rendering this common reasoning misguided.

In contrast with parametric complexity, configural complexity cannot be detected by simply counting model parameters. Models that differ in their configural complexity can yield different fits to the data even when the contain the same number of parameters[68,69,73]. One study directly tested this assertion by fitting a confirmatory bifactor model and an exploratory factor analysis model with the same number of parameters to 1,000 randomly generated data sets[68]. The confirmatory bifactor model achieved good fit for almost as many data patterns as the exploratory factor model and the confirmatory bifactor model achieved better fit than the exploratory factor model for a small proportion of data patterns. In other words, although confirmatory models are supposed to impose constraints on the data, the purportedly confirmatory bifactor model is so adept at fitting data that it verges on atheoretical (indeed some have suggested that the 'confirmatory' in 'confirmatory factor analysis' is a misnomer!)[74]. All told, the proliferation of *p*-factor research is primarily attributable to the good fit of bifactor models of psychopathology, but the bifactor model's inherent tendency to fit well shows that goodness-of-fit is not a strong test of any *p*-factor theory[75,76].

Underspecified construct validation

The *p*-factor literature tends to classify virtually any statistically significant correlation between a general factor of psychopathology and a criterion variable as evidence of its validity, with little attention paid to the magnitude of the observed association, discriminant validity or theory falsification. Such a focus on statistical significance impedes theory development. For instance, the hypothesis that the *p*-factor reflects low intellectual functioning would generate greater skepticism if researchers attended closely to the magnitude of the observed associations, which show that performance on various intelligence tasks explains essentially none of the variance in general factors of psychopathology across existing studies[1,20,53]. Similarly, associations between general factors of psychopathology and negative emotionality vary widely, with correlations ranging from 0.13 (ref. 59) to 0.88 (ref. 54) (the median is 0.42) (Fig. 2). Values on the lower end of this range are similar to those identified for intelligence, presenting a correlation

Table 1 | Considerations for modelling general factors of psychopathology

Model	Specifications	Considerations
One-factor 	Summarizes the variance shared among all indicators but does not model residual variance among subsets of its indicators	One dimension may not adequately summarize the data Fewer parameters than the other models, so good fit is not likely. If a good fit is achieved with a simple model, it is more likely to have captured a generalizable trend[86,126] The factor generally correlates highly with other general factors of psychopathology
Bifactor 	Allows for residual variance among subsets of indicators that take form as specific factors	Specific factors reflect the variance shared among their indicators after taking into account what they share with the general factor of psychopathology[127] Typically associated with relatively good model fit compared with other models, but strong fit does not indicate that the model is an adequate or valid description of the data[66,68,107]
Higher-order 	Includes a set of lower-order dimensions whose shared variance is described by a superordinate higher-order dimension	With two lower-order factors, general factor of psychopathology loadings simply recapitulate the correlation between the lower-order factors With three lower-order factors, the model fit is equivalent to a model with three correlated factors With four or more lower-order factors, the model is differentiable from the correlated factors model Specific factors are residuals when their external validity is examined Less parametrically complex than the bifactor model, so its model fit facilitates a stronger test of whether a general factor of psychopathology is valid when there are four or more lower-order factors
Bass-ackwards 	Sequentially extracts factors (or principal components) from one all the way up to the number of indicators included in the model	Capable of modelling a more detailed hierarchical structure than the bifactor and higher-order models[128] Begins by extracting a general factor of psychopathology, regardless of whether a general factor of psychopathology is supported in the data. The general factor of psychopathology should not be reified without further validation
Hierarchical clustering 	Sorts indicators into a detailed tree-like structure from either the top-down (divisive) or the bottom-up (agglomerative)	Capable of modelling a more detailed hierarchical structure than the bifactor and higher-order models Begins or ends with a single cluster (that is, a general factor of psychopathology) but that cluster should not be reified without further validation[129] Stopping rules indicate whether a general factor of psychopathology (that is, a single-cluster solution) is supported[130]; here, two clusters (depicted in grey boxes), not one, are supported

C, cluster; F, factor; GFP, general factor of psychopathology; i, item.

too small to corroborate claims about substantial causal processes or mechanisms.

There are several possible explanations for such varied associations between general factors of psychopathology and theoretically relevant external criteria. First, such variability is further evidence of bifactor model complexity. A model with high complexity will minimally constrain possible outcomes and flexibly accommodate data. Thus, it is unsurprising that an overly complex model produces a wide range of associations with external criteria across data sets[69,77]. That is, a parsimonious model will only fit well if the data patterns fall within a narrow range of possibilities, but a complex model can fit well and produce widely varied parameter estimates. Consequently, a complex model has more flexibility regarding associations with external criteria.

Second, general factors of psychopathology often correlate more highly with criterion variables when both are measured using the same method (for example, self-report), owing to shared (or mono-) method covariance, than when either psychopathology or the criterion variables are measured using different methods (such as caregiver

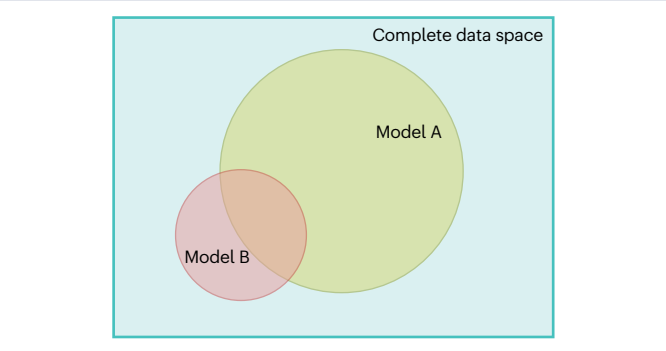
Box 1

Model complexity

In statistical modelling, complexity refers to "the property of a model that enables it to fit diverse patterns of data"[133]. By implication, a good fit to the observed data — if obtained with an overly complex model — might be meaningless[77]. Thus, researchers should quantify complexity when evaluating a statistical model.

To inform model complexity, researchers have simulated random data and fitted various competing models to the data. In the figure, the turquoise outer box represents the complete space of all possible data that could be fitted by two hypothetical models. The proportion of that space that will achieve a good fit (according to some arbitrary threshold, such as a Comparative Fit Index value of >0.95) for a given model is indicated by the size of the model's circle. The green circle depicts the proportion of data for which model A fits well and the pink circle depicts the proportion of data for which model B fits well. The overlap between models A and B represents the data patterns that both models will fit equally well, whereas the non-overlapping areas reflect patterns that are unique to each model.

In the figure, model A occupies a larger region than does model B. Thus, model A is better equipped to fit diverse patterns of data (that is, it has a higher fitting propensity)[25,69]. Moreover, the limited overlap between models A and B indicates that they reflect somewhat dissimilar data patterns. If models A and B have equivalent numbers of parameters, these results inform configural complexity and suggest that model A is, on average, more likely to fit random data than model B. If model A has more parameters than model B, these results cannot directly isolate and inform configural complexity because they are confounded by the models' differing degrees of parametric complexity.



This approach to model evaluation is crucial for rigorous theory testing for at least two reasons. First, an understanding of relative complexity directly corresponds to the philosophy of science notion of risky theory testing[76]. The figure shows that model B is the riskier choice: it will be less likely than model A to fit well, so its good fit, if achieved, will provide much stronger evidence in support of the theory[77,134]. Second, model complexity is inversely associated with generalizability. A well-fitting parsimonious model is more likely to capture the generalizable trend in the data, whereas a well-fitting complex model is likely to also capture sample-dependent noise[86,87]. Thus, if researchers want to conduct risky tests of a theory and draw generalizable inferences from their findings, they must ensure that the statistical model they use for testing is not excessively complex. For real applications to models used in the *p*-factor literature, see refs. 66,68,73.

or teacher reports, or laboratory tasks)[78]. Perhaps this virtual truism is why support for the *p*-factor is stronger in the context of shared method variance[23].

Third, predictor–criterion overlap occurs when a general factor of psychopathology contains the same (or closely related) content as the outcome variable. For instance, a high correlation between a general factor of psychopathology and a latent negative emotionality factor[45,54] is weak evidence of the theory that the *p*-factor reflects negative emotionality because the same feelings and behaviours (for example, sadness, anxiety, dysphoria and irritability) are captured in both psychopathology and negative emotionality measures. Indeed, several published general factors of psychopathology correlate so strongly with indices of negative emotionality that they border on being statistically isomorphic[45,54]. Similarly, the finding that general factors of psychopathology are associated with experiencing relationship problems[19] can be attributed at least in part to the fact that externalizing includes numerous conditions that explicitly incorporate interpersonal problems (for example, substance-use disorder and oppositional defiance). Further, impairment in social and other areas of functioning is required for a clinical diagnosis of virtually any condition outlined in The Diagnostic and Statistical Manual of Mental Disorders (DSM). Thus, impairment is baked into models that include diagnoses.

Ultimately, we suspect that weak theories are at least partially to blame for underspecified construct validation efforts. As with many areas of psychology, the *p*-factor literature has long relied on verbal theories that are rarely precisely articulated or quantified, making them difficult to falsify and extremely easy to corroborate[53]. As an example, disordered thought was promoted as the cause of the *p*-factor because it occurs "in the context of affective disorders, anxiety disorders, eating disorders, posttraumatic stress disorder, somatoform disorders, dissociative disorders, substance use disorders, and antisocial disorders"[6] (see also ref. 1). So defined, this theory is inherently tautological and self-sustaining: the *p*-factor is valid and reflects disordered thought if it is modelled using any form of psychopathology, because disordered thought permeates all psychopathology. Considering such weak *p*-factor theories, we are not convinced that establishing any statistically significant correlation between a general factor of psychopathology and another variable is informative about the validity and nature of the *p*-factor.

Poor replication across studies

Many narrative reviews of the *p*-factor claim that general factors of psychopathology replicate well across studies[1,2,21,79,80]. But the *p*-factor literature uses the term 'replicate' in its weakest possible meaning: that

of replication of good fit. General factors of psychopathology 'replicated' because researchers successfully fitted bifactor models – which are inherently successful at fitting data – to different data sets[81]. Unfortunately, pointing to two well-fitting models in different data without considering replication of specific model parameters (for example, factor loadings)[3,9,79] is entirely meaningless (see ref. 49 for a discussion of this issue). Such a practice is no different from fitting two linear regressions to two data sets that measure smoking and lung cancer and claiming that the findings replicate simply because linear regressions were estimated, and not because the associations between those variables were statistically similar. Within the *p*-factor literature, some studies show that psychosis loads most strongly onto a general factor of psychopathology[1], whereas others show that major depression and generalized anxiety load most strongly (even nearly perfectly) onto general factors of psychopathology[2,4,19]. Yet, in both circumstances, researchers interpreted their general factors of psychopathology as 'the *p*-factor', with the former concluding that it reflects disordered thought[1,6], and the latter concluding that it reflects negative emotionality[11,45].

Supporting the argument that general factors of psychopathology vary in their interpretive meaning across studies, one study found that general factors of psychopathology from three different epidemiologic samples had dramatically variable parameter estimates, producing varied interpretations of the general factors of psychopathology[24]. Some were defined by thought disorder, others by major depression and generalized anxiety, and others by post-traumatic stress disorder. A larger investigation found remarkably poor convergence among 15 published general factors of psychopathology (intraclass correlation coefficient of 0.23)[4] with loadings on these general factors of psychopathology once again varying substantially (Fig. 3): the loadings for panic disorder with agoraphobia ranged from −0.02 (ref. 42) to 0.81 (ref. 82) and loadings for attention deficit hyperactivity disorder ranged from 0.14 (ref. 83) to 0.73 (ref. 13).

Such poor replicability of general factors of psychopathology across studies – all of which are taken to reflect or tap into the same *p*-factor – renders narrative reviews[3,6,27] difficult to interpret. The *p*-factor is so ill-defined that one cannot know what is meant by any individual finding that mentions it. Alice might use 'the *p*-factor' to refer to a general factors of psychopathology that is largely represented by depression and anxiety[84], and Bob might use 'the *p*-factor' to refer to a general factor of psychopathology that is dominated by thought disorder[1]. But there is no strong rationale to claim that Bob has replicated Alice's findings. Rather, it is plausible that these two general factors of psychopathology reflect distinct sources of covariation or mechanisms that are study- and/or sample-specific.

Multiple interacting forces contribute to the *p*-factor's replication crisis[5,24]. We focus on two key forces here. First, model complexity is inversely associated with generalizability[85,86]. A well-fitting parsimonious model is more likely to capture the replicable signal in the data, whereas a well-fitting complex model is likely to also capture sample-dependent noise[86,87]. Thus, general factors of psychopathology are naturally less likely to replicate important model parameters and more likely to vary in their interpretive meaning across studies. If researchers want to detect the meaningful trend underlying psychopathology and thereby draw replicable inferences from their findings, they must ensure that the statistical model they use is not excessively complex. Complex phenomena might require complicated models, but complicated models require more rigorous evaluation.

Second, some of the poor replicability of general factors of psychopathology might arise from differences in psychopathology measurement across studies. To explore that possibility, we examined 12 published general factors of psychopathology[9,13,45,56,57,79,81,88–92] that were constructed using identical measurement (the Achenbach School-based Empirically Based Assessment scales for psychopathology)[93]. An advantage of constraining our study selection to these scales is that they are predominantly used in studies of youth, which further constrains the age distribution of the samples under consideration. We focused on three subscales (anxious-depressed, withdrawn-depressed and thought problems) that bear directly on two popular *p*-factor theories: that it reflects negative emotionality or thought disorder[1,6]. Even when general factors of psychopathology were modelled using identical measurement in samples with relatively constrained age variability, they were associated with substantial variability in their factor loadings (Fig. 4). The range of factor loadings on the general factors of psychopathology was 0.18 to 0.86 for anxious-depressed, 0.17 to 0.74 for withdrawn-depressed and 0.21 to 0.87 for thought problems. Thus, although measurement and sample characteristics

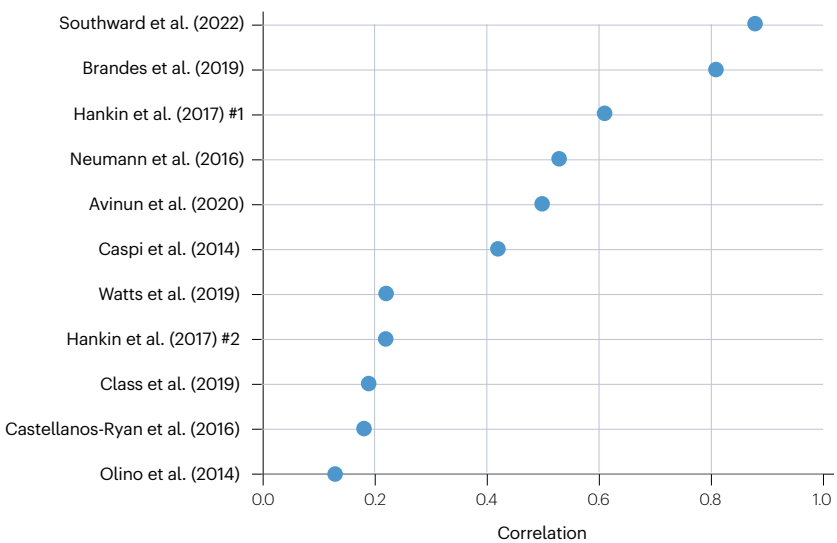


Fig. 2 | Correlations between general factors of psychopathology and negative emotionality. Associations between general factors of psychopathology and negative emotionality from ten studies[1,4,45,53,54,56,57,59–61], with each association depicted as a dot. The effect sizes range from 0.13 to 0.88.

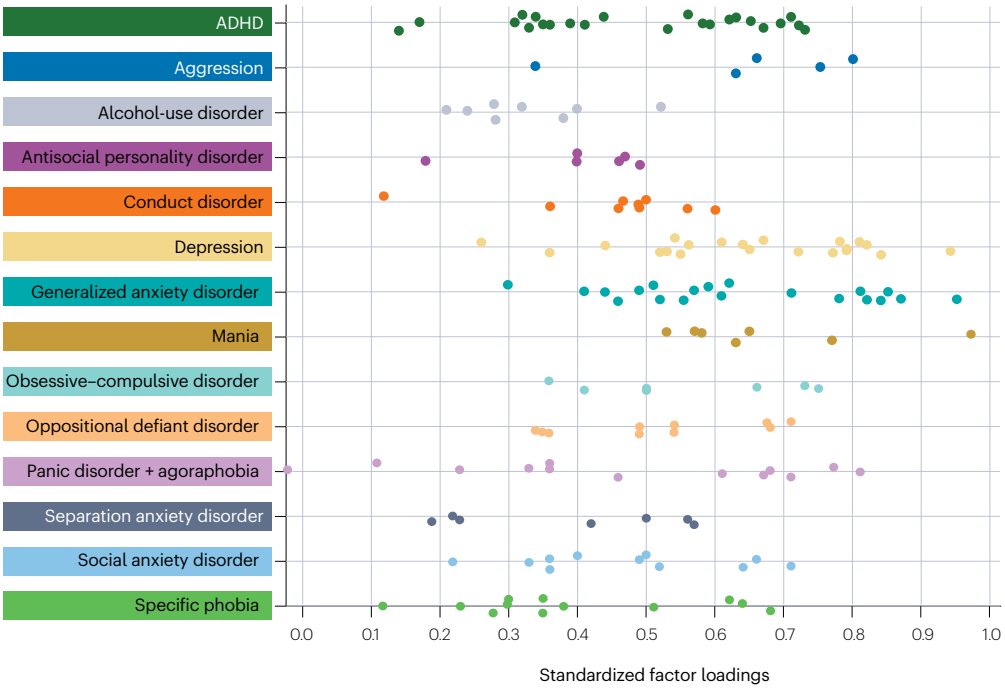


Fig. 3 | Standardized factor loadings on general factors of psychopathology. Standardized factor loadings on 15 published general factors of psychopathology[1,2,4,11,13,16,42,79,80,82,83,91,131,132], with a focus on indicators that were included in at least five general factors of psychopathology. Each loading is depicted as a dot. ADHD, attention deficit hyperactivity disorder.

are both relevant to replicating general factors of psychopathology, they are clearly not the leading causes of their poor replicability. Our frontrunner remains the bifactor model's complexity.

Violations of latent variable assumptions

Issues replicating general factors of psychopathology across studies point to another statistical concern regarding the nature of the *p*-factor. General factors of psychopathology are almost exclusively modelled using factor analysis, a form of latent variable modelling. Latent factors reflect the shared variance of their indicators, but they are often interpreted as causal entities whereby variation in the indicators is caused by variation in the latent factor (that is, a common-cause model; see ref. 94 for a more detailed description of why such a conclusion is misguided). Thus, the arrows in path diagrams point away from the latent factor and towards the indicators.

Although not all psychopathology researchers interpret factor models as causal[5,95], explicit causal interpretations of general factors of psychopathology have dominated the *p*-factor literature. For example, one study stated "…the latent variable analysis posits that the positive correlations between symptoms (as well as disorders) arise from a *g*-like causal factor…"[6]. Similarly, another research group[2] offered that the *p*-factor "may reflect the influence of etiologic factors that are shared by all mental disorders" and later explicitly integrated the *p*-factor into a causal hierarchical taxonomy of psychopathology[7].

Importantly, latent variable models impose assumptions on data that have implications for the replicability of general factors of psychopathology and the meaning of the *p*-factor. One critical assumption is that latent variables are composed of random as opposed to fixed sets of indicators[96,97]. With fixed variables, researchers assume that a construct is perfectly measured because it was measured with all possible indicators. Fixed latent variables are largely implausible, in part because the recognition that indicators are imperfect is embedded in the principles of factor analysis; latent variables are viewed as preferable because aggregating imperfect, unreliable indicators improves the latent construct's reliability.

Further, general factors of psychopathology cannot be considered fixed variables because few could defend the position that all relevant indicators of psychopathology have been exhaustively measured. For example, the seminal study of the *p*-factor was limited to modelling externalizing and internalizing disorders[2], whereas another general factor of psychopathology was subsequently expanded to include psychosis-related phenomena[1]. Neither study included other possible indicators of mental disorder (for example, post-traumatic stress disorder or autism). This issue is exemplified in the *p*-factor literature, where some aspects of psychopathology (such as anxiety and depression) are often used to model general factors of psychopathology and others (such as vocal and motor tics, autism and bipolar disorder) are not[5].

With random variables, researchers assume that indicators are imperfect and randomly sampled from a universe of possible indicators. For instance, a test of mathematical skills in children necessarily cannot contain an infinite set of items that requires them to add up all possible combinations of numbers. Instead, a test contains a mere subset of possible questions. When researchers assume that they are taking a random sample of indicators, they also assume that the indicators are interchangeable and equally relevant to the latent construct (Box 2). That is, we assume that the latent variable is "indifferent to its indicators"[29]. In the math test example, a researcher might assume that 2 + 4 and 1 + 3 are equally difficult, and so that including either item would produce the same score on that test. Thus, any latent factor's meaning should not vary across all possible subsamples of indicators, nor should its associations with external criteria[98].

Nevertheless, several empirical findings are further at odds with the assumption that general factors of psychopathology are random and indicated by interchangeable indicators, or put differently,

that general factors of psychopathology with different indicators reflect the same construct. First, general factors of psychopathology do not replicate strongly across studies with different indicators[5,24]. Second, one study showed that general factors of psychopathology are not always invariant to their contents within a given data set by extracting a variety of general factors of psychopathology in the same data and leaving out one of 15 indicators per model. The exclusion of either major depression or conduct disorder from the general factors of psychopathology caused their interpretation to shift dramatically[4]. Major depression's loadings on general factors of psychopathology ranged from −0.32 to 0.81 and conduct disorder's loadings ranged from −0.45 to 0.63. These findings suggest that the general factor of psychopathology composed of all indicators was so defined by depression and conduct problems that their exclusion changed its meaning.

In sum, general factors of psychopathology are extremely sensitive to their contents. This observation explains the poor replication of general factors of psychopathology across studies, and why some general factors of psychopathology correlate very highly with negative emotionality and others very little (Fig. 2). General factors of psychopathology composed of different indicators do not necessarily reflect the same construct.

Assumption of causality

A critical implication of violations of indicator interchangeability is that the *p*-factor is unlikely to reflect a latent cause, rendering substantive explanations of the *p*-factor implausible and general factors of psychopathology artefactual. In fact, there are numerous viable alternative explanations for general factors of psychopathology that do not posit that a unitary entity causes variation in all forms of psychopathology (Box 3). That is, there are numerous alternative models of psychopathology that do not contain a causal *p*-factor.

For example, general factors represent little more than the sum of scores[97]. Thus, general factors of psychopathology are often a literal index of diagnostic comorbidity or severity (sums of diagnoses or symptoms). This assertion is supported by evidence that sum scores of diagnoses exhibit near-perfect correlations (*r* values 0.87 to 1.00)[99] with the general factor of psychopathology modelled in classic *p*-factor studies[2,8,19]. These findings imply that identical information can be obtained from a very simple addition of all symptoms of a person, compared with highly sophisticated bifactor models that often contain dozens if not hundreds of estimated parameters[1,2]. These findings also imply that, at best, general factors of psychopathology might index the severity of one's psychopathology profile[1,5,6].

Ironically, total psychopathology scores have been available since the 1970s in widely used instruments[93], but they are rarely reported in empirical studies (for exceptions, see refs. 19,88). More generally, it is exceedingly rare to see total psychopathology score composites in published articles, raising the question of why general factors of psychopathology estimated using sophisticated models have become so popular in the past decade. In addition to focusing narrowly on model fit, we suspect that general factors of psychopathology became popular owing to the tendency of *p*-factor researchers to reify general factors of psychopathology as latent, causal entities[2,6]. That is, although a sum score composite tends not to encourage a reification fallacy or belief in an illusory essence[100], latent variables do.

An alternative explanation is that general factors of psychopathology are formative as opposed to reflective. In a reflective model, the indicators (for example, item responses or diagnoses) reflect the underlying latent variable. That is, variation in indicators is thought to be caused by the latent variable. By contrast, in a formative model the variable is formed by its indicators, and the particular composition of that variable will influence its interpretation (for example, the meaning of a formative variable of socioeconomic status will differ depending on whether neighbourhood factors are included; the likelihood that the variable's meaning will differ depending on the indicators included in the model is compatible with violations of item interchangeability in general factors of psychopathology). If general factors of psychopathology are formative as opposed to reflective, what is captured in general factors of psychopathology is a shared outcome (or outcomes) as opposed to a shared cause. Some *p*-factor researchers have arrived at this very conclusion, arguing that general factors of psychopathology most plausibly capture impairment[5,27].

We concur that most general factors of psychopathology probably capture impairment, given the way they are modelled. Researchers tend

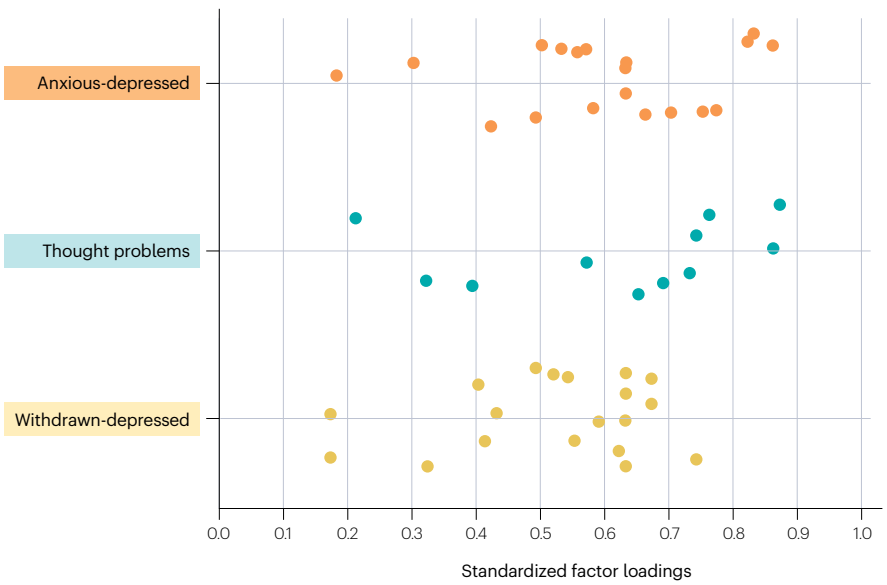
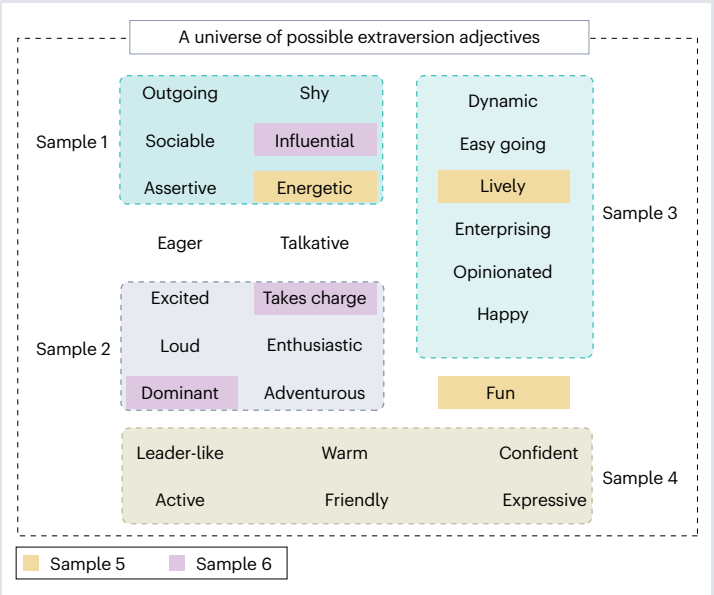


Fig. 4 | Standardized factor loadings on general factors of psychopathology constructed using identical measurement. Standardized factor loadings on 12 published general factors of psychopathology[9,13,45,56,57,79,81,88–92] for anxious-depressed, thought problems and withdrawn-depressed Achenbach School-based Empirically Based Assessment subscales. Each dot represents a factor loading from a study.

Box 2

Item interchangeability

Consider the personality trait extraversion (the tendency to experience positive affect and sociability). There are dozens if not hundreds of adjectives that might describe an extraverted individual. It is essentially impossible to construct a measure that contains every possible adjective relevant to extraversion. Instead, measures attempt to capture a range of relevant descriptors, but an incomplete set of them. If latent variables are not fixed, or composed of all possible indicators, the latent variable is necessarily random, containing a purportedly random sample of a universe of possible indicators[96,97]. One random sample (or measure) of extraversion adjectives could include outgoing, sociable, assertive, shy, influential and energetic (see the figure, sample 1) as indicators of extraversion, whereas another could include leader-like, active, warm, friendly, confident and expressive (see the figure, sample 4).



Researchers make two assumptions with random variables. First, researchers assume that indicators are imperfect. Indeed, the appeal of latent variable modelling is that latent variables aggregate indicators that are associated with unique measurement error, resulting in a more reliable composite than sum scores and removing indicator-specific measurement error. Second, researchers assume that indicators are interchangeable and equally relevant to the latent construct. Thus, samples 1, 2, 3 and 4 should be equivalent representations of extraversion. If they are not, it would suggest that extraversion should not be conceptualized as a latent variable.

Consider two other samples: sample 5 is composed of fun, lively and energetic, and sample 6 is composed of influential, takes charge and dominant. Although these are technically subsets of extraversion adjectives, they are more homogeneous than samples 1–4. In fact, arguably the adjectives in sample 5 reflect a narrower component of extraversion, something like ebullience. By contrast, the adjectives in sample 6 reflect leader-like qualities. Although these subsets of indicators are certainly relevant to extraversion, they are not sufficient to describe it. Research studies that correlate, for example, a composite of sample 5 indicators with an outcome (for example, popularity in school) might erroneously conclude that extraversion is associated with it, but that association might be driven exclusively by ebullience, particularly if other samples of items are not robustly correlated with the outcome[98].

to interpret general factors of psychopathology as reflecting some etiologic process (or processes)[6,7], but they model psychopathology symptoms and diagnoses. Setting aside biased response processes, psychopathology symptoms and diagnoses are the outcomes of some set of etiologic factors that produce psychopathology. Thus, general factors of psychopathology probably capture downstream sequelae of whatever causes psychopathology, further distancing latent factors from their purported causal processes. General factor of psychopathology models might be missing relevant etiologic components in part because modern taxonomies of psychopathology stick closely to DSM-defined phenomena[101], and DSM criteria are notoriously agnostic with respect to etiology[102,103].

Another model that might explain general factors of psychopathology is a network model[104,105], where the covariance between two symptoms or two disorders might not stem from a latent variable that causes both, but from direct causal associations among them (for example, insomnia might lead to fatigue and depression might lead to insomnia and fatigue). Alternatively, psychopathology might be best explained by a hybrid model that contains a mix of latent variables and individual symptoms (among other variables) that cause one another (for example, a residual network model)[26]. Realistically, there are as many statistical alternatives for modelling the data-generating process in psychopathology as there are hypotheses about what the *p*-factor might represent[106]. But assuming that the *p*-factor reflects some substantive causal process when other statistically equivalent models imply otherwise is a prime example of 'affirming the consequent'[26,48]: it is fallacious to conclude that variation in psychopathology symptoms must be caused by a latent construct simply because a latent variable of psychopathology symptoms was extracted.

Moving forward

Foremost, we encourage *p*-factor researchers to abandon the unrestricted bifactor model and to gauge the strength of their theories based on how well they stand up to risky tests[75]. Accordingly, researchers should drastically de-emphasize or disregard goodness-of-fit in bifactor modelling[68,74,107], as they do when evaluating exploratory models[108]. A riskier, more rigorous test of the appropriateness of a bifactor model requires incorporating plausible constraints, such as Bayesian constraints according to a priori theories. If one presupposes that the *p*-factor should equally influence all mental disorders — if it is,

Box 3

Model equivalence

Model fit statistics cannot tell us which model is correct, and numerous equally plausible, well-fitting models could represent the underlying data structure[106,135]. Model equivalence occurs when models "yield identical (a) implied covariance, correlation, and other moment matrices when fit to the same data, which in turn imply identical (b) residuals and fitted moment matrices, (c) fit functions and chi-square values, and (d) goodness-of-fit indices based on fit functions and chi-square"[136]. Within the context of the *p*-factor literature, there are many fit-equivalent models that differ in their functional form, and ultimately differ in their implications for *p*-factor theories. Further, although many researchers establish that their bifactor model of psychopathology fits well, few if any acknowledge that there might be alternative models that better describe the observed data.

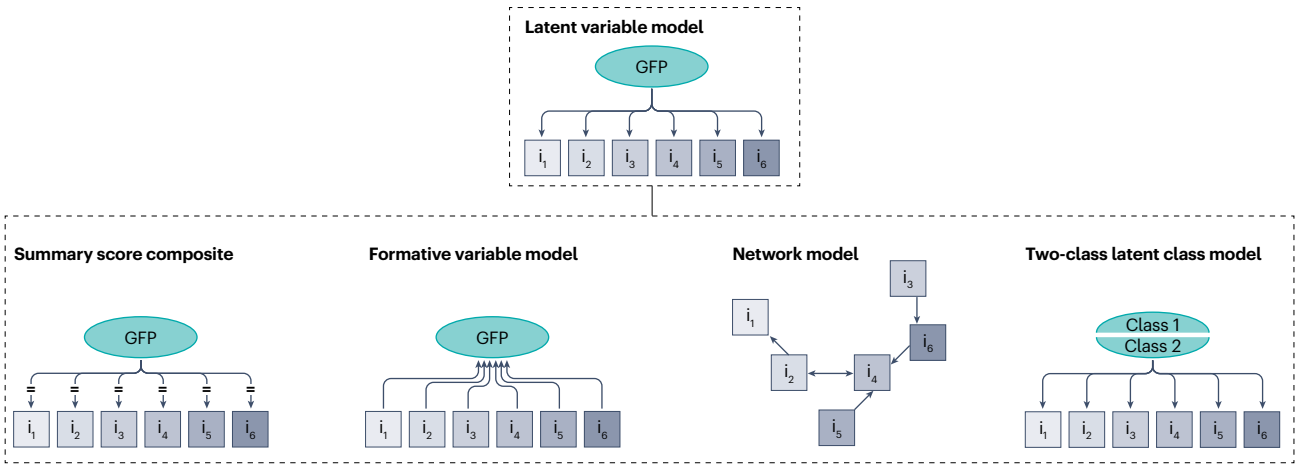
Here, we focus on four models (see the figure) that have been empirically demonstrated as equivalent to a single-factor latent variable model under many conditions (model equivalence will always depend on the particular arrangement of model parameters). We focus on equivalence between a single-factor confirmatory factor analysis model (latent variable) and alternative models (as opposed to a bifactor model) because general factors of psychopathology across single-factor and bifactor models are highly congruent with one another[19].

First, a summary score composite simply adds up items into a composite. It is equivalent to a single-factor model when the factor loadings on the single-factor model are equal (or tau-equivalent; typically, latent factors are treated as congeneric, which is when loadings are freely estimated and allowed to vary across indicators)[137]. In summary score composites, indicators are unit weighted (multiplied by one) and treated as equivalent, exchangeable indicators of the construct of interest. The same goes for a single-factor model, except that the items are weighted by the factor loading that is held constant across all indicators; thus, the indicators are weighted by different constants, which affects the scaling of their scores, but they are otherwise equivalent. This instance of model equivalence has two important implications: (1) that it is equally as plausible that the *p*-factor reflects some substantive mechanism as it is that the *p*-factor is simply a summary composite of diagnostic comorbidity or severity[2,5,6]; and (2) that the *p*-factor literature's tendency to reify latent variables as causal entities cannot be justified on the basis of simply extracting a latent variable.

Second, a formative model is equivalent to a single-factor model. In the formative model, the arrows point towards the construct as opposed to away from it[138,139]. That is, indicators form a construct, with no assumptions regarding the nature of the interrelatedness among the indicators. This instance of model equivalence is directly relevant to *p*-factor theories, in which substantive, causal interpretations are presumably supported by extracting a single-factor model, but an equivalent structural representation of the indicators supports an opposing perspective: that the *p*-factor reflects impairment[5,27].

Third, a single-factor model can be equivalent to a network model[140,141], whereby the associations of specific indicators are summarized by causal associations (or dynamic relationships over time) as opposed to a unitary latent variable. In contrast with latent variable conceptualizations of the *p*-factor, a network model conceptualization suggests that psychopathology is not caused by a major superordinate dimension, but that certain psychopathology symptoms cause others, and those might cause others[105].

Fourth, a two-class latent class model can be equivalent to a single-factor model. A latent class model identifies groups or subtypes of people (latent classes). A two-class solution indicates that there are two qualitatively different subpopulations of people with different profiles of psychopathology. For instance, one class might have high externalizing, and the other might have high internalizing. In fact, simulations suggest that researchers are more likely to find support for a bifactor model in the presence



(continued from previous page)
of population heterogeneity, suggesting that the bifactor model can mislead researchers into concluding that a single structure of psychopathology describes the entire population[142]. With respect to the *p*-factor, these findings suggest that spurious support for a general factor of psychopathology can arise in the presence of population heterogeneity. For more thorough discussions of model equivalence, see refs. 106,107. C, cluster; F, factor; GFP, general factor of psychopathology; i, item.

in fact, a general influence on psychopathology — one could impose constraints that test whether the factor loadings on the general factor of psychopathology are statistically equivalent[4]. Similarly, if one believes that the *p*-factor is caused by disordered thought, one might impose constraints on factor loadings such that the highest loadings correspond to the forms of psychopathology most imbued with disordered thought (such as psychosis)[6]. Otherwise, without deliberately reining in the complexity of a bifactor model, its good fit is meaningless[49,77]. The same strategy could be applied to construct validation efforts for general factors of psychopathology: Researchers could specify the range of associations one might expect between a general factor of psychopathology and a given outcome should it support or fail to support a particular *p*-factor theory. Generally, researchers who are keen to establish the general factor of psychopathology as a robust statistical phenomenon ought to rely on numerous statistical methods that place their favoured *p*-factor theory at risk, as opposed to relying on fallible indices from a single statistical model (as demonstrated in a preprint)[109].

Further, the *p*-factor literature's focus on factor models and between-subjects data limits interpretations of general factors of psychopathology to individual differences accounts. Current models are unable to distinguish the group level from the individual level because estimated model parameters summarize the group, not an individual person[110]. Statistical models that can evaluate how well a group-level model describes particular individuals in the data set require disentangling between-subjects from within-subjects variance, which requires data with multiple timepoints. We therefore call for models that differentiate nomothetic and idiographic processes with respect to approximating the so-called *p*-factor.

More generally, time and context are also relatively understudied in the *p*-factor literature. Most general factors of psychopathology are estimated on cross-sectional data, which provides crude insight into the dynamics of mental health problems and fails to articulate how psychopathology develops over time. These limitations of the modal study design in the *p*-factor literature are critical given that the *p*-factor is often promoted as a mechanism (or set of mechanisms) responsible for the cause, expression and maintenance of all forms of psychopathology[2,3,6]. Studying how psychopathology develops requires a study design that is equipped to inform change over time, but many features of studies in the *p*-factor literature preclude studying change. For instance, common assessment methods (such as assessing symptoms on a frequency scale) inform how cross-situationally consistent or trait-like the psychopathology is[111], as do structural equation models that aggregate multiple assessments over time into a composite. Indeed, even when longitudinal data are used in studies of general factors of psychopathology[1], they extract the shared variance in psychopathology over time.

Ultimately, these studies encourage the conclusion that psychopathology is chronic, potentially even trait-like. But psychopathology is far from static: many mental disorders (such as major depressive disorder or bipolar disorder) are often experienced in episodes[112–114], their expression can oscillate depending on biological and contextual factors[115], and they can develop in a stage-like manner or through dynamic, casual interactions among phenomena over time[105,116]. Because psychopathology cannot be fully understood or modelled without taking time into account, we encourage researchers to assess and model psychopathology in a manner that adequately models change (and lack thereof) over time. Longitudinal studies are not one-size-fits-all, and instead require careful consideration of the time-bound functional relationships among forms of psychopathology. For instance, if a researcher supposes that there may be a causal association between two phenomena, the study must follow people over a sufficiently informative window of time to track the dynamic unfolding of the phenomenon of interest and assess psychopathology at the proper increments. A study on whether negative mood elicits alcohol consumption would be uninformative if both are only assessed annually, and instead could involve repeated daily assessments. By contrast, a month-long study of phenomena that unfold over longer periods of time, such as alcohol withdrawal, is ill-suited to answer questions about their dynamics.

Nevertheless, even when general factors of psychopathology are tested in a more rigorous and constrained manner, they often fail to incorporate critical etiologic factors that are articulated in theories of psychopathology[117]. For instance, general factors of psychopathology often do not directly incorporate biopsychosocial and contextual factors that are much more proximal to the development of psychopathology, such as attachment styles, stress, adversity, early trauma and personality. We encourage moving away from modelling diagnoses or related variables (such as symptom counts), which are the consequences of disruption in biological and environmental systems that combine to produce mental dysfunction. Modelling outcomes of mental dysfunction distances psychopathology models from their putative mechanisms and causes.

Further, the omission of intersecting contextual factors[23,118–120] that clarify and make meaning of what might be contributing to a particular person's distress at a specific time and place, and why that person might respond to specific situations in more or less adaptive ways, limits the *p*-factor's validity and utility in clinical research and practice (for an alternative view, see ref. 121). For the *p*-factor literature to inform clinical practice, general factors of psychopathology ought to incorporate the very phenomena that are critical to case conceptualization. Otherwise, general factors of psychopathology will continue to function as mere proxies of clinical severity and impairment.

Considering the methodological issues that plague the *p*-factor literature, we encourage researchers to approach modelling general factors of psychopathology with increased skepticism, gathering sufficient empirical evidence — through rigorous as opposed to superficial tests — before reifying the *p*-factor as a construct that is implicated in the cause or development of psychopathology broadly construed. A high-profile example of the *p*-factor's reification is its adoption in the Hierarchical Taxonomy of Psychopathology (HiTOP). An instance of truth in advertising, HiTOP organizes psychopathology hierarchically,

from narrow signs and symptoms into a series of increasingly broad dimensions. A central tenet of HiTOP is that it is a data-driven, 'living' model that is subject to revisions (including incorporating novel constructs) based on evidence as appropriate following a formal and rigorous process[122].

In its inaugural publication in 2017 (ref. 123), HiTOP's model incorporated an agnostic label at its apex of 'higher-order dimensions'. This label acknowledged the likelihood that multiple higher-order dimensions (as opposed to one dimension) are positioned at the apex of the psychopathology hierarchy. In 2019, HiTOP papers that depicted the conceptual model altered 'higher-order dimensions' to 'general psychopathology'[124] and later to 'general factor of psychopathology (*p*-factor)'[39]. These changes occurred without a formal revision to the model. Because the *p*-factor was effectively grandfathered into the HiTOP model before the formal revisions process was developed, it was arguably not subjected to the same scrutiny it might be subjected to today. In fact, most other major dimensions in HiTOP, particularly externalizing and internalizing, have been studied for many decades and have a relatively strong empirical foundation, whereas other dimensions are regarded as provisional in the current model[39] even though they were studied long before the *p*-factor. For instance, the body of evidence for a somatoform dimension, whose specific position in the model also remains debated, dates back to 1999 (ref. 38). Ultimately, the *p*-factor's inclusion in HiTOP warrants revisiting, because it was arguably premature and based on weak evidence.

Conclusion

Central to the genesis of the theoretical *p*-factor was the superior fit of one statistical model — the bifactor model — over all others, not theory[5]. Favouring statistical models over theory is not inherently problematic; exploratory methods can reveal novel insights into the structure of psychopathology. Yet the *p*-factor literature is moored to statistical characteristics that cannot guarantee that researchers chose the optimal model. In fact, the bifactor model's fitting propensity increases the likelihood that researchers will choose the least generalizable model of the ones tested. Accordingly, meaningful aspects of general factors of psychopathology have failed to replicate well across studies and violate key model assumptions, which makes it difficult to synthesize the published literature and renders some favoured explanations of the *p*-factor exceedingly unlikely. Even worse, many post-hoc theories of the *p*-factor are so weak that they can be corroborated with virtually any model that extracts a general factor of psychopathology or any examination of a general factor of psychopathology's correlations with any external criterion. These critical limitations of the literature undermine confidence that the *p*-factor is a real, latent entity, or that general factors of psychopathology are useful statistical summaries of psychopathology.

As researchers embark on the second decade of research on the *p*-factor, two things are sure to advance the literature: subjecting theoretical and statistical models to riskier tests and bringing theory to the fore[48]. Preacher, a quantitative psychologist and psychometrician, once stated "Cherished models may have to be abandoned or replaced if their past successes can be ascribed more to [fitting propensity] than to any insight they lend into the process that actually generated the data"[69]. Indeed, we argue that the cherished *p*-factor might have to be abandoned because its past successes can be ascribed to fitting propensity, conflating theoretical and statistical models, weak theories, underspecified attempts to validate general factors of psychopathology, and the use of data that are underequipped to inform the etiology, development and maintenance of psychopathology. To be sure, if modal practices continue, another decade of *p*-factor research will leave researchers largely empty-handed, bereft of meaningful progress on the classification and treatment of psychopathology. As Menninger's provocative concept of a single mental illness enjoys a contemporary renaissance, it calls to mind the warning "Old beliefs die hard even when demonstrably false"[125].

Published online: 22 December 2023

References

1. Caspi, A. et al. The *p* factor: one general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.* **2**, 119–137 (2014).
2. Lahey, B. B. et al. Is there a general factor of prevalent psychopathology during adulthood? *J. Abnorm. Psychol.* **121**, 971–977 (2012).
3. Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D. & Zald, D. H. Validity and utility of the general factor of psychopathology. *World Psychiatry* **16**, 142–144 (2017).
4. Watts, A. L., Poore, H. E. & Waldman, I. D. Riskier tests of the validity of the bifactor model of psychopathology. *Clin. Psychol. Sci.* **7**, 1285–1303 (2019).
5. Watts, A. L., Lane, S. P., Bonifay, W., Steinley, D. & Meyer, F. A. C. Building theories on top of, and not independent of, statistical models: the case of the *p*-factor. *Psychol. Inq.* **31**, 310–320 (2020).
6. Caspi, A. & Moffitt, T. E. All for one and one for all: mental disorders in one dimension. *Am. J. Psychiatry* **175**, 831–844 (2018).
7. Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D. & Zald, D. H. A hierarchical causal taxonomy of psychopathology across the life span. *Psychol. Bull.* **143**, 142–186 (2017).
8. Greene, A. L. & Eaton, N. R. The temporal stability of the bifactor model of comorbidity: an examination of moderated continuity pathways. *Compr. Psychiatry* **72**, 74–82 (2017).
9. Snyder, H. R., Young, J. F. & Hankin, B. L. Strong homotypic continuity in common psychopathology-, internalizing-, and externalizing-specific factors over time in adolescents. *Clin. Psychol. Sci.* **5**, 98–110 (2017).
10. Murray, A. L., Eisner, M. & Ribeaud, D. The development of the general factor of psychopathology '*p* factor' through childhood and adolescence. *J. Abnorm. Child Psychol.* **44**, 1573–1586 (2016).
11. Tackett, J. L. et al. Common genetic influences on negative emotionality and a general psychopathology factor in childhood and adolescence. *J. Abnorm. Psychol.* **122**, 1142–1153 (2013).
12. Waldman, I. D., Poore, H., van Hulle, C., Rathouz, P. & Lahey, B. B. External validity of a hierarchical dimensional model of child and adolescent psychopathology: tests using confirmatory factor analyses and multivariate behavior genetic analyses. *J. Abnorm. Psychol.* **125**, 1053–1066 (2016).
13. Harden, K. P. et al. Genetic associations between executive functions and a general factor of psychopathology. *J. Am. Acad. Child Adolesc. Psychiatry* **59**, 749–758 (2020).
14. Allegrini, A. G. et al. The *p* factor: genetic analyses support a general dimension of psychopathology in childhood and adolescence. *J. Child Psychol. Psychiatry* **61**, 30–39 (2020).
15. Kaczkurkin, A. N. et al. Common and dissociable regional cerebral blood flow differences associate with dimensions of psychopathology across categorical diagnoses. *Mol. Psychiatry* **23**, 1981–1989 (2018).
16. Romer, A. L. et al. Replicability of structural brain alterations associated with general psychopathology: evidence from a population-representative birth cohort. *Mol. Psychiatry* **26**, 3839–3846 (2019).
17. O'Reilly, L. M. et al. The association between general childhood psychopathology and adolescent suicide attempt and self-harm: a prospective, population-based twin study. *J. Abnorm. Psychol.* **129**, 364–375 (2020).
18. Pettersson, E., Larsson, H., D'Onofrio, B. M., Bölte, S. & Lichtenstein, P. The general factor of psychopathology: a comparison with the general factor of intelligence with respect to magnitude and predictive validity. *World Psychiatry* **19**, 206–213 (2020).
19. Forbes, M. K. et al. Three recommendations based on a comparison of the reliability and validity of the predominant models used in research on the empirical structure of psychopathology. *J. Abnorm. Psychol.* **130**, 297–317 (2021).
20. Martel, M. M. et al. A general psychopathology factor (*p* factor) in children: structural model analysis and external validation through familial risk and child global executive function. *J. Abnorm. Psychol.* **126**, 137–148 (2017).
21. Lahey, B. B. et al. Criterion validity of the general factor of psychopathology in a prospective study of girls. *J. Child Psychol. Psychiatry* **56**, 415–422 (2015).
22. Pettersson, E., Larsson, H., D'Onofrio, B., Almqvist, C. & Lichtenstein, P. Association of fetal growth with general and specific mental health conditions. *JAMA Psychiatry* **76**, 536 (2019).
23. Watts, A. L. et al. How robust is the *p* factor? Using multitrait-multimethod modeling to inform the meaning of general factors of youth psychopathology. *Clin. Psychol. Sci.* **10**, 640–661 (2022).
24. Levin-Aspenson, H. F., Watson, D., Clark, L. A. & Zimmerman, M. What is the general factor of psychopathology? Consistency of the *p* factor across samples. *Assessment* **28**, 1035–1049 (2021).

25. Bonifay, W., Lane, S. P. & Reise, S. P. Three concerns with applying a bifactor model as a structure of psychopathology. *Clin. Psychol. Sci.* **5**, 184–186 (2017).
26. van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D. & van der Maas, H. L. J. What is the *p*-factor of psychopathology? Some risks of general factor modeling. *Theory Psychol.* **27**, 759–773 (2017).
27. Smith, G. T., Atkinson, E. A., Davis, H. A., Riley, E. N. & Oltmanns, J. R. The general factor of psychopathology. *Annu. Rev. Clin. Psychol.* **16**, 75–98 (2020).
28. Littlefield, A. K., Lane, S. P., Gette, J. A., Watts, A. L. & Sher, K. J. The "big everything": integrating and investigating dimensional models of psychopathology, personality, personality pathology, and cognitive functioning. *Pers. Disord. Theory Res. Treat.* **12**, 103–114 (2021).
29. Spearman, C. 'General intelligence,' objectively determined and measured. *Am. J. Psychol.* **15**, 201 (1904).
30. Horn, J. L. & McArdle, J. J. in *Factor Analysis at 100: Historical Developments and Future Directions* 1st edn Ch. 11 (eds Cudeck, R, & MacCallum, R. C.) 205–247 (Routledge, 2007).
31. Blum, J. M. *Pseudoscience and Mental Ability: The Origins and Fallacies of the IQ Controversy* (Monthly Review Press, 1978).
32. Gould, S. J. *The Mismeasure of Man* (Penguin, 1984).
33. Menninger, K., Ellenberger, H., Pruyser, P. & Mayman, M. The unitary concept of mental illness. *Pastor. Psychol.* **10**, 13–19 (1959).
34. McKusick, V. A. On lumpers and splitters, or the nosology of genetic disease. *Persp. Biol. Med.* **12**, 298–312 (1969).
35. Achenbach, T. M. The classification of children's psychiatric symptoms: a factor-analytic study. *Psychol. Monogr. Gen. Appl.* **80**, 1–37 (1966).
36. Krueger, R. F. & Markon, K. E. Reinterpreting comorbidity: a model-based approach to understanding and classifying psychopathology. *Annu. Rev. Clin. Psychol.* **2**, 111–133 (2006).
37. Krueger, R. F. et al. Validity and utility of hierarchical taxonomy of psychopathology (HiTOP): II. Externalizing superspectrum. *World Psychiatry* **20**, 171–193 (2021).
38. Watson, D. et al. Validity and utility of hierarchical taxonomy of psychopathology (HiTOP): III. Emotional dysfunction superspectrum. *World Psychiatry* **21**, 26–54 (2022).
39. Kotov, R. et al. The hierarchical taxonomy of psychopathology (HiTOP): a quantitative nosology based on consensus of evidence. *Annu. Rev. Clin. Psychol.* **17**, 83–108 (2021).
40. Kotov, R. New dimensions in the quantitative classification of mental illness. *Arch. Gen. Psychiatry* **68**, 1003–1011 (2011).
41. Ringwald, W. R., Forbes, M. K. & Wright, A. G. C. Meta-analysis of structural evidence for the Hierarchical Taxonomy of Psychopathology (HiTOP) model. *Psychol. Med.* **53**, 533–546 (2021).
42. Olino, T. M. et al. The development of latent dimensions of psychopathology across early childhood: stability of dimensions and moderators of change. *J. Abnorm. Child Psychol.* **46**, 1373–1383 (2018).
43. Hoertel, N. et al. Mental disorders and risk of suicide attempt: a national prospective study. *Mol. Psychiatry* **20**, 718–726 (2015).
44. Laceulle, O. M., Chung, J. M., Vollebergh, W. A. M. & Ormel, J. The wide-ranging life outcome correlates of a general psychopathology factor in adolescent psychopathology. *Pers. Ment. Health* **14**, 9–29 (2020).
45. Brandes, C. M., Herzhoff, K., Smack, A. J. & Tackett, J. L. The *p* factor and the *n* factor: associations between the general factors of psychopathology and neuroticism in children. *Clin. Psychol. Sci.* **7**, 1266–1284 (2019).
46. Carver, C. S., Johnson, S. L. & Timpano, K. R. Toward a functional view of the *p* factor in psychopathology. *Clin. Psychol. Sci.* **5**, 880–889 (2017).
47. Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I. & Waldorp, L. J. Invisible hands and fine calipers: a call to use formal theory as a toolkit for theory construction. *Persp. Psychol. Sci.* **16**, 725–743 (2021).
48. Fried, E. I. Lack of theory building and testing impedes progress in the factor and network literature. *Psychol. Inq.* **31**, 271–288 (2020).
49. Bonifay, W., Winter, S. D., Skoblow, H. F. & Watts, A. L. Good fit is weak evidence of replication: increasing rigor through prior predictive similarity checking. *OSF* https://doi.org/10.17605/OSF.IO/Q6RVF (2023).
50. Lahey, B. B. et al. Measuring the hierarchical general factor model of psychopathology in young adults. *Int. J. Methods Psychiat. Res.* **27**, e1593 (2018).
51. Kovas, Y. & Plomin, R. Learning abilities and disabilities: generalist genes, specialist environments. *Curr. Dir. Psychol. Sci.* **16**, 284–288 (2007).
52. Lahey, B. B., Van Hulle, C. A., Singh, A. L., Waldman, I. D. & Rathouz, P. J. Higher-order genetic and environmental structure of prevalent forms of child and adolescent psychopathology. *Arch. Gen. Psychiatry* **68**, 181 (2011).
53. Castellanos-Ryan, N. et al. The structure of psychopathology in adolescence and its common personality and cognitive correlates. *J. Abnorm. Psychol.* **125**, 1039–1052 (2016).
54. Southward, M. W., Cheavens, J. S. & Coccaro, E. F. Defining the *p*-factor: an empirical test of five leading theories. *Psychol. Med.* 53, 2732–2743 (2022).
55. Widiger, T. A. & Oltmanns, J. R. Neuroticism is a fundamental domain of personality with enormous public health implications. *World Psychiatry* **16**, 144–145 (2017).
56. Hankin, B. L. et al. Temperament factors and dimensional, latent bifactor models of child psychopathology: transdiagnostic and specific associations in two youth samples. *Psychiatry Res.* **252**, 139–146 (2017).
57. Neumann, A. et al. Single nucleotide polymorphism heritability of a general psychopathology factor in children. *J. Am. Acad. Child Adolesc. Psychiatry* **55**, 1038–1045.e4 (2016).
58. Watts, A. L., Poore, H. E., Lilienfeld, S. O. & Waldman, I. D. Clarifying the associations between Big Five personality domains and higher-order psychopathology dimensions in youth. *J. Res. Pers.* **82**, 103844 (2019).
59. Olino, T. M., Dougherty, L. R., Bufferd, S. J., Carlson, G. A. & Klein, D. N. Testing models of psychopathology in preschool-aged children using a structured interview-based assessment. *J. Abnorm. Child Psychol.* **42**, 1201–1211 (2014).
60. Class, Q. A. et al. Socioemotional dispositions of children and adolescents predict general and specific second-order factors of psychopathology in early adulthood: a 12-year prospective study. *J. Abnorm. Psychol.* **128**, 574–584 (2019).
61. Avinun, R., Romer, A. L. & Israel, S. Vitamin D polygenic score is associated with neuroticism and the general psychopathology factor. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **100**, 109912 (2020).
62. Berg, J. M., Latzman, R. D., Bliwise, N. G. & Lilienfeld, S. O. Parsing the heterogeneity of impulsivity: a meta-analytic review of the behavioral implications of the UPPS for psychopathology. *Psychol. Assess.* **27**, 1129–1146 (2015).
63. King, J. A. et al. Cognitive overcontrol as a trait marker in anorexia nervosa? Aberrant task- and response-set switching in remitted patients. *J. Abnorm. Psychol.* **128**, 806–812 (2019).
64. Pinto, A., Steinglass, J. E., Greene, A. L., Weber, E. U. & Simpson, H. B. Capacity to delay reward differentiates obsessive–compulsive disorder and obsessive–compulsive personality disorder. *Biol. Psychiatry* **75**, 653–659 (2014).
65. Reise, S. P., Kim, D. S., Mansolf, M. & Widaman, K. F. 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. *Multivar. Behav. Res.* **51**, 818–838 (2016).
66. Greene, A. L. et al. Are fit indices used to test psychopathology structure biased? A simulation study. *J. Abnorm. Psychol.* **128**, 740–764 (2019).
67. Morgan, G. B., Hodge, K. J., Wells, K. E. & Watkins, M. W. 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. *J. Intell.* **3**, 2–20 (2015).
68. Bonifay, W. & Cai, L. On the complexity of item response theory models. *Multivar. Behav. Res.* **52**, 465–484 (2017).
69. Preacher, K. J. Quantifying parsimony in structural equation modeling. *Multivar. Behav. Res.* **41**, 227–259 (2006).
70. Bader, M. & Moshagen, M. Assessing the fitting propensity of factor models. *Psychol. Methods* https://doi.org/10.1037/met0000529 (2022).
71. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* **19**, 716–723 (1974).
72. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).
73. Falk, C. F. & Muthukrishna, M. Parsimony in model selection: tools for assessing fit propensity. *Psychol. Methods* **28**, 123–136 (2021).
74. Greene, A. L. et al. Misbegotten methodologies and forgotten lessons from Tom Swift's electric factor analysis machine: a demonstration with competing structural models of psychopathology. *Psychol. Methods* https://doi.org/10.1037/met0000465 (2022).
75. Waller, N. G. & Meehl, P. E. Risky tests, verisimilitude, and path analysis. *Psychol. Methods* **7**, 323–337 (2002).
76. Meehl, P. E. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *J. Consult. Clin. Psychol.* **46**, 806–834 (1978).
77. Roberts, S. & Pashler, H. How persuasive is a good fit? A comment on theory testing. *Psychol. Rev.* **107**, 358–367 (2000).
78. Campbell, D. T. & Fiske, D. W. Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychol. Bull.* **56**, 81–105 (1959).
79. Noordhof, A., Krueger, R. F., Ormel, J., Oldehinkel, A. J. & Hartman, C. A. Integrating autism-related symptoms into the dimensional internalizing and externalizing model of psychopathology. The TRAILS study. *J. Abnorm. Child Psychol.* **43**, 577–587 (2015).
80. Snyder, H. R., Young, J. F. & Hankin, B. L. Chronic stress exposure and generation are related to the *p*-factor and externalizing specific psychopathology in youth. *J. Clin. Child Adolesc. Psychol.* **48**, 306–315 (2019).
81. Laceulle, O. M., Vollebergh, W. A. M. & Ormel, J. The structure of psychopathology in adolescence: replication of a general psychopathology factor in the TRAILS study. *Clin. Psychol. Sci.* **3**, 850–860 (2015).
82. Gluschkoff, K., Jokela, M. & Rosenström, T. The general psychopathology factor: structural stability and generalizability to within-individual changes. *Front. Psychiatry* **10**, 594 (2019).
83. Conway, C. C., Mansolf, M. & Reise, S. P. Ecological validity of a quantitative classification system for mental illness in treatment-seeking adults. *Psychol. Assess.* **31**, 730–740 (2019).
84. Funkhouser, C. J. et al. Evaluating the criterion validity of hierarchical psychopathology dimensions across models: familial aggregation and associations with research domain criteria (sub)constructs. *J. Abnorm. Psychol.* **130**, 575–586 (2021).
85. Pitt, M. A., Kim, W. & Myung, I. J. Flexibility versus generalizability in model selection. *Psychon. Bull. Rev.* **10**, 29–44 (2003).
86. Bonifay, W. Increasing generalizability via the principle of minimum description length. *Behav. Brain Sci.* **45**, e5 (2022).
87. Myung, J. I., Pitt, M. A. & Kim, W. in *Handbook of Cognition* Ch. 19 (eds Lamberts, K. & Goldstone, R. L.) 422–436 (Sage, 2005).
88. Clark, D. A. et al. The general factor of psychopathology in the Adolescent Brain Cognitive Development (ABCD) study: a comparison of alternative modeling approaches. *Clin. Psychol. Sci.* **9**, 169–182 (2021).

89. Bloemen, A. J. P. et al. The association between executive functioning and psychopathology: general or specific? *Psychol. Med.* **48**, 1787–1794 (2018).
90. Liu, J., Mustanski, B., Dick, D., Bolland, J. & Kertes, D. A. Risk and protective factors for comorbid internalizing and externalizing problems among economically disadvantaged African American youth. *Dev. Psychopathol.* **29**, 1043–1056 (2017).
91. Shields, A. N., Reardon, K. W., Brandes, C. M. & Tackett, J. L. The *p* factor in children: relationships with executive functions and effortful control. *J. Res. Pers.* **82**, 103853 (2019).
92. Neumann, A. et al. White matter microstructure and the general psychopathology factor in children. *J. Am. Acad. Child Adolesc. Psychiatry* **59**, 1285–1296 (2020).
93. Achenbach, T. M., Ivanova, M. Y. & Rescorla, L. A. Empirically based assessment and taxonomy of psychopathology for ages 1½–90+ years: developmental, multi-informant, and multicultural findings. *Compr. Psychiatry* **79**, 4–18 (2017).
94. Van Bork, R., Wijsen, L. D. & Rhemtulla, M. Toward a causal interpretation of the common factor model. *Disputatio* **9**, 581–601 (2017).
95. DeYoung, C. G. & Krueger, R. F. To wish impossible things: on the ontological status of latent variables and the prospects for theory in psychology. *Psychol. Inq.* **31**, 289–296 (2020).
96. Steyer, R. Models of classical psychometric test theory as stochastic measurement models: representation, uniqueness, meaningfulness, identifiability, and testability. *Methodika* **3**, 25–60 (1989).
97. Cattell, R. B. *Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist* (Greenwood, 1973).
98. Vainik, U., Mõttus, R., Allik, J., Esko, T. & Realo, A. Are trait–outcome associations caused by scales or particular items? Example analysis of personality facets and BMI. *Eur. J. Personal.* **29**, 622–634 (2015).
99. Fried, E. I., Greene, A. L. & Eaton, N. R. The *p* factor is the sum of its parts, for now. *World Psychiatry* **20**, 69–70 (2021).
100. Brick, C., Hood, B., Ekroll, V. & de-Wit, L. Illusory essences: a bias holding back theorizing in psychological science. *Perspect. Psychol. Sci.* **17**, 491–506 (2022).
101. McWilliams, N. Diagnosis and its discontents: reflections on our current dilemma. *Psychoanal. Inq.* **41**, 565–579 (2021).
102. Whooley, O. Nosological reflections: the failure of *DSM*-5, the emergence of RDoC, and the decontextualization of mental distress. *Soc. Ment. Health* **4**, 92–110 (2014).
103. Cuthbert, B. N. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology: forum — The Research Domain Criteria Project. *World Psychiatry* **13**, 28–35 (2014).
104. Borsboom, D. A network theory of mental disorders. *World Psychiatry* **16**, 5–13 (2017).
105. Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R. & Borsboom, D. The network approach to psychopathology: a review of the literature 2008–2018 and an agenda for future research. *Psychol. Med.* **50**, 353–366 (2020).
106. Tomarken, A. J. & Waller, N. G. Potential problems with 'well fitting' models. *J. Abnorm. Psychol.* **112**, 578–598 (2003).
107. Greene, A. L. et al. Model fit is a fallible indicator of model quality in quantitative psychopathology research: a reply to Bader and Moshagen. *J. Psychopathol. Clin. Sci.* **131**, 696–703 (2022).
108. Montoya, A. K. & Edwards, M. C. The poor fit of model fit for selecting number of factors in exploratory factor analysis for scale evaluation. *Educ. Psychol. Meas.* **81**, 413–440 (2021).
109. Watts, A. L., Sher, K. J., Heath, A. C., Steinley, D. & Brusco, M. 'General addiction liability' revisited. *OSF* https://doi.org/10.17605/OSF.IO/XDSNP (2023).
110. Piccirillo, M. L. & Rodebaugh, T. L. Foundations of idiographic methods in psychology and applications for psychotherapy. *Clin. Psychol. Rev.* **71**, 90–100 (2019).
111. Fleeson, W. Toward a structure- and process-integrated view of personality: traits as density distributions of states. *J. Pers. Soc. Psychol.* **80**, 1011–1027 (2001).
112. Lorenzo-Luaces, L. Heterogeneity in the prognosis of major depression: from the common cold to a highly debilitating and recurrent illness. *Epidemiol. Psychiat. Sci.* **24**, 466–472 (2015).
113. Monroe, S. M. & Harkness, K. L. Is depression a chronic mental illness? *Psychol. Med.* **42**, 899–902 (2012).
114. Vriends, N., Bolt, O. C. & Kunz, S. M. Social anxiety disorder, a lifelong disorder? A review of the spontaneous remission and its predictors. *Acta Psychiat. Scand.* **130**, 109–122 (2014).
115. De Los Reyes, A. & Makol, B. A. in *The Oxford Handbook of Personality and Psychopathology Assessment* 2nd edn (Oxford Academic, 2021).
116. Olthof, M., Hasselman, F., Oude Maatman, F., Bosman, A. M. T. & Lichtwarck-Aschoff, A. Complexity theory of psychopathology. *J. Psychopathol. Clin. Sci.* **132**, 314–323 (2023).
117. Cicchetti, D. & Cohen, D. J. (eds) *Developmental Psychopathology* (Wiley, 2016).
118. De Los Reyes, A. & Kazdin, A. E. Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychol. Bull.* **131**, 483–509 (2005).
119. De Los Reyes, A., Thomas, S. A., Goodman, K. L. & Kundey, S. M. A. Principles underlying the use of multiple informants' reports. *Annu. Rev. Clin. Psychol.* **9**, 123–149 (2013).
120. De Los Reyes, A., Tyrell, F. A., Watts, A. L. & Asmundson, G. J. G. Conceptual, methodological, and measurement factors that disqualify use of measurement invariance techniques to detect informant discrepancies in youth mental health assessments. *Front. Psychol.* **13**, 931296 (2022).
121. Meier, M. A. & Meier, M. H. Clinical implications of a general psychopathology factor: a cognitive–behavioral transdiagnostic group treatment for community mental health. *J. Psychother. Integr.* **28**, 253–268 (2018).
122. Forbes, M. K. et al. Principles and procedures for revising the hierarchical taxonomy of psychopathology. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/xr48p (2023).
123. Kotov, R. et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. *J. Abnorm. Psychol.* **126**, 454–477 (2017).
124. Conway, C. C. et al. A hierarchical taxonomy of psychopathology can transform mental health research. *Persp. Psychol. Sci.* **14**, 419–436 (2019).
125. Wilson, E. O. *Consilience: The Unity of Knowledge* (Vintage Books, 1999).
126. Vitanyi, P. M. B. & Ming, L. Minimum description length induction, Bayesianism, and Kolmogorov complexity. *IEEE Trans. Inf. Theory* **46**, 446–464 (2000).
127. Reise, S. P. The rediscovery of bifactor measurement models. *Multivar. Behav. Res.* **47**, 667–696 (2012).
128. Forbes, M. K. Improving hierarchical models of individual differences: an extension of Goldberg's bass-ackward method. *Psychol. Methods* https://doi.org/10.1037/met0000546 (2023).
129. Mirkin, B. Choosing the number of clusters. *WIREs Data Mining Knowl. Discov.* **1**, 252–260 (2011).
130. Milligan, G. W. & Cooper, M. C. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**, 159–179 (1985).
131. Pezzoli, P., Antfolk, J. & Santtila, P. Phenotypic factor analysis of psychopathology reveals a new body-related transdiagnostic factor. *PLoS ONE* **12**, e0177674 (2017).
132. He, Q. & Li, J. J. Factorial invariance in hierarchical factor models of mental disorders in African American and European American youths. *J. Child Psychol. Psychiatry* **62**, 289–298 (2021).
133. Pitt, M. A. & Myung, I. J. When a good fit can be bad. *Trends Cogn. Sci.* **6**, 421–425 (2002).
134. Vanpaemel, W. Strong theory testing using the prior predictive and the data prior. *Psychol. Rev.* **127**, 136–145 (2020).
135. Raykov, T. & Marcoulides, G. A. Can there be infinitely many models equivalent to a given covariance structure model? *Struct. Equ. Modeling* **8**, 142–149 (2001).
136. Hershberger, S. L. & Marcoulides, G. A. in *Structural Equation Modeling: A Second Course* 2nd edn (eds Hancock, G. & Mueller, R.) 3–39 (Information Age, 2013).
137. McNeish, D. & Wolf, M. G. Thinking twice about sum scores. *Behav. Res. Methods* **52**, 2287–2305 (2020).
138. Schuberth, F. The Henseler–Ogasawara specification of composites in structural equation modeling: a tutorial. *Psychol. Methods* https://doi.org/10.1037/met0000432 (2021).
139. Edwards, J. R. & Bagozzi, R. P. On the nature and direction of relationships between constructs and measures. *Psychol. Methods* **5**, 155–174 (2000).
140. van Bork, R. et al. Latent variable models and networks: statistical equivalence and testability. *Multivar. Behav. Res.* **56**, 175–198 (2021).
141. Van Der Maas, H. L. J. et al. A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychol. Rev.* **113**, 842–861 (2006).
142. Raykov, T., Marcoulides, G. A., Menold, N. & Harrison, M. Revisiting the bi-factor model: can mixture modeling help assess its applicability? *Struct. Equ. Modeling* **26**, 110–118 (2019).

Acknowledgements
A.L.W. is funded through grant K99/R00AA028306. A.L.G. is supported by the Office of Academic Affiliations, Advanced Fellowship Program in Mental Illness Research and Treatment, Department of Veterans Affairs. W.B. is supported by the Institute of Education Sciences, US Department of Education, through grant R305D210032.

Author contributions
A.L.W. wrote the first full draft, and all authors contributed to the manuscript's conceptualization and revision.

Competing interests
The authors declare no competing interests.

Additional information
Peer review information *Nature Reviews Psychology* thanks Benjamin Hankin, Tyler Moore and Matthew Southward for their contribution to the peer review of this work.