Supplementary Information for "Answering questions about the Hierarchical Taxonomy of Psychopathology (HiTOP): Analogies to whales and sharks miss the boat," by DeYoung et al., published in *Clinical Psychological Science.*

Here we present corrections and rebuttals to errors and misconceptions in "Folk classification and factor rotations: Whales, sharks, and the problems with HiTOP," by Haeffel et al., which were not fully covered in our commentary. Problematic passages are listed in the order in which they appear in the target article, and section headings from the target article are noted for ease of reference. Quotations from Haeffel et al. are presented in bold, followed by our discussion of each (page numbers were not available at the time of preparing this supplement).

Abstract

"The Hierarchical Taxonomy of Psychopathology (HiTOP) uses factor analysis to group people with similar self-reported symptoms (i.e., like-goes-with-like)."

This is the first of many instances in which Haeffel et al. mischaracterize what HiTOP is classifying. HiTOP does not attempt to group people. HiTOP is a classification of features of psychopathology, not of people, and it groups those features according to their tendency to co-occur within people, taking a variable-centered, rather than a person-centered, approach to classification. Also, HiTOP includes other assessment modalities in addition to self-report, such as clinician ratings.

"Claim 1. Symptom Correlations Carve Nature at its Joints"

"In the Linnaean system, classification decisions are not based on total levels of 'likeness' (i.e., their covariation) as in HiTOP, but rather on a subgroup of highly meaningful features as determined by evolutionary theory (i.e., phylogeny, e.g., Nickels & Nelson, 2005). To this end, the Linnaean system distinguishes between homology and analogy (Petto & Mead, 2009)."

This description of the Linnaean taxonomic system is misleading because it depicts the Linnaean system as grounded in evolutionary theory. In fact, the Linnaean system predates the theory of evolution by a century. The manner in which Haeffel et al. describe biological classification corresponds to the modern evolutionary taxonomy of species, rather than to the structural taxonomy of Carl Linnaeus. The first work cited in the quoted passage describes how Darwin's evolutionary ideas about biological classification "were a major advance over Linnaeus' ideas" (Nickels & Nelson, 2005, p. 283), and the second describes the modern evolutionary taxonomic system without mentioning the Linnaean system at all.

"This process would likely lead to an overarching factor of 'animal' (the A-factor), which might then break down into a bifactor model of "land" and "water" animals. An examination of the subgroups of animals organized within these two levels starts to reveal the problems with HiTOP. For example, whales and sharks would be incorrectly classified together given the high correlations among their shared features (e.g., ocean dwellers, fins for locomotion, fish and crustacean eaters, similar life spans, can adapt to multiple aquatic habitats, both largest of their family). This is

because in HiTOP, features such as being warm blooded and having hair do not carry special importance."

This passage reflects Haeffel et al.'s misunderstanding of what it is that HiTOP classifies (features of psychopathology, not people). It is true that the biological taxonomy of species is not derived using methods, such as factor analysis or taxometrics, that HiTOP uses. However, if biological classification were to be constructed using the same methods as HiTOP, it would not be a categorization of individual species (analogous to individual people) but rather of the features of organisms (analogous to features of psychopathology). Thus, it would not resemble what is described in the quoted passage or depicted in Haeffel et al.'s Figure 1. Further, if one did apply factor analysis to the features of organisms, we see no reason to assume that living on land or water would necessarily be the defining indicators of resulting factors, because those would just be two variables along with many others. Nonetheless, if factors related to piscine water-dwelling (fins, scales, gills, etc.) and mammalian land-dwelling (legs, hair, lungs, etc.) did appear, it could easily be the case that whales had moderately high scores on the mammalian factor and also some elevation on the piscine factor. Importantly, each species would have scores on each factor. Clearly, there would be more to discover, based on genetics, after such an outcome for whales, but the initial findings might not be so misleading as Haeffel et al. suggest. (Note that we are not recommending this sort of factor analysis as an approach to biological classification; we are merely illustrating the flaws in Haeffel et al.'s analogy.) We note also that, although Figure 1 shows water-dwelling animals divided into "Finned" and "Flippers," with whales in the former, in reality whales have both fins, on their backs, and flippers, as their forelimbs.

"This calls into question HiTOP's most fundamental assumption, namely that individuals who report similar patterns of symptoms have the same form of psychopathology (which can be targeted by the same treatment due to shared etiology; Ruggero et al., 2019)."

HiTOP does not assume that a given symptom or set of symptoms always has the same etiology, and it is never *necessary* that they would have the same etiology, no matter how similar their symptoms. Nonetheless, the more features two cases have in common, the more *likely* it becomes that they share some etiology as well. To be maximally similar in HiTOP, two people would have to have similar levels, not just of the six spectra, but also of all 100+ of its lower-level dimensions. In such an extreme case, it would be surprising if at least some etiological factors were not shared. Additionally, although structural models themselves do not necessarily reflect shared etiology (Jonas & Markon, 2016), nonetheless, if the correct causal model is among the structural models being compared, it will be favored by information-theoretical approaches (Grunwald, 2007; Vrieze, 2012). Even if the true causal model is not among those considered, structural models can increase the amount of causal information available. For a fuller description of the relationship between structure and causation, see Markon and Jonas (2016).

"Similarly, it is untenable to assume that people with depression and people with PTSD should be grouped together (because of shared "distress" symptoms) without understanding their etiology."

Again, HiTOP does not group people; it groups symptoms and other features of psychopathology according to their patterns of covariation. HiTOP's hierarchical structure makes clear that traditional diagnoses of depression and PTSD share some features but also include other features that are not shared.

"Claim 2. HiTOP Will Solve the Problems of Comorbidity and Heterogeneity"

"That said, let us assume that comorbidity in the currently used diagnostic system (DSM) *does* reflect redundancies and inaccuracies. Does HiTOP solve the problem as promised by Conway and colleagues (2019)? The HiTOP solution is to lump diagnoses together and then give them a new label. ... [HiTOP] gives new labels to the same collection of symptoms. This creates larger more heterogeneous groupings, which may not be clinically useful and can hinder our understanding of the etiology of mental illness."

HiTOP does not lump diagnoses together because it does not include diagnostic categories, due to their empirically demonstrated unreliability and limited validity. Instead, HiTOP provides a dimensional system in which symptoms are grouped together at upper levels based on their tendencies to covary, while being distinguished at lower levels based on their unshared variance. The inclusion of some traditional diagnostic labels at the level of "Syndromes" in diagrams of the HiTOP system has sometimes led to the unfortunate misconception that these DSM constructs are included as HiTOP constructs. In fact, those labels were included for illustrative purposes only, to allow mapping of existing nosologies onto HiTOP (and those with the most prominent cross-loadings are listed in multiple places). The HiTOP consortium has used information about covariance of diagnoses only as a way to bootstrap toward understanding the broader dimensional structure of psychopathology, not in order to include the diagnoses themselves within HiTOP. HiTOP often does not retain the same grouping of symptoms from each DSM diagnosis, but rather redistributes them among different dimensions based on their patterns of covariance with other symptoms. Nor does HiTOP create more heterogeneous groupings because, at its lowest levels, it provides extremely homogeneous, narrow symptom dimensions, while at higher levels it shows how these tend to co-occur. Thus, at each level, features are designed to have maximum homogeneity given the level of resolution in question.

"Moreover, an implicit assumption of HiTOP is that people will fit neatly into one spectrum and a line of subfactors. However, research indicates that this is unlikely. Instead, people will 'score high' on multiple subfactors and spectra (e.g., the co-occurrence of internalizing and externalizing problems is substantial in both clinical and epidemiological studies; Pesenti-Gritti et al., 2008). Thus, people categorized using HiTOP are still going to carry an abundance of labels, as a person might report internalizing, externalizing, substance use, distress, and antisocial behavior symptoms."

HiTOP does not make this assumption, either explicitly or implicitly. Instead, HiTOP indicates (based on empirical evidence like that described by Pesenti-Gritti et al., 2008, or Caspi et al., 2020) that many people will exhibit symptoms from multiple spectra. It is a feature, not a bug, of the system that people can have elevated scores on multiple dimensions. Consortium publications state that psychopathology should be characterized by scores from across HiTOP's multiple spectra. For example, Ruggero et al., (2019, p. 1075) wrote, "HiTOP permits a flexible, stepwise approach to assessment, beginning with brief screening of higher order spectra, and then—based on time and need—progressing to more focused assessments to characterize the subfactors, syndromes, and symptoms/traits within each spectrum more fully." The dimensional conceptualization of psychopathology entails that every person in the population has a position (potentially at zero) on every dimension of the HiTOP model. The person is not represented by one dimension but by a profile across all dimensions. Similar assertions have been made in many consortium publications.

"These are just a few examples (others include COVID-19, hyperthyroidism, irritable bowel syndrome, etc.) that illustrate how people can express completely different symptom profiles without overlapping symptoms, and yet suffer from the same underlying problem. HiTOP would miss these cases because the symptom profiles do not covary; it cannot deal with this kind of natural complexity (Kendler et al., 2011)."

HiTOP would not "miss" these cases because it would describe each of them in terms of the symptoms with which they present. Instead, what the current HiTOP system might miss is the shared etiology of such symptoms. Importantly, HiTOP is designed to evolve based on etiological information, as that information becomes known. Thus, a future version of HiTOP might include description of the fact that the two symptom profiles hypothetically in question here could share the same etiology, despite their lack of symptom overlap. HiTOP does not include such descriptions currently, as there are no established psychopathological constructs that follow this pattern (i.e., produce non-overlapping symptom profiles).

"According to these authors, neuroticism consists of six correlated but distinct constructs. Thus, it is possible for two people to have the exact same score on a general measure of neuroticism but for different reasons (e.g., one person may score high on hostility and low on self-consciousness, whereas another person may score low on hostility and high on self-consciousness). They argue that this kind of heterogeneity makes a total score on neuroticism imprecise, ambiguous, and an obstacle to theory testing."

The final sentence here reveals a fundamental misconception about hierarchical constructs such as neuroticism or HiTOP spectra. In fact, the neuroticism score is intended to reflect the shared variance of its lower level "correlated but distinct constructs." Genetically informative research indicates that there is unique valid variance at each level of the personality trait hierarchy (Jang et al., 1998, 2002; McCrae et al., 2008; Mõttus et al., 2019). Similarly, there is unique valid variance at each level of the hierarchy of psychopathology (Kotov et al., 2016; Michelini et al., 2019; Zald & Lahey, 2017). Thus, there are likely to be some causes specific to neuroticism as a general tendency, regardless of how that tendency is manifested in terms of the elevation of its constituent lower-level constructs. Then there will be other causes that are specific to the lower-level constructs. Theories must be developed regarding the causes of covariance at each level of the hierarchy. This makes it advantageous to work with a hierarchical descriptive model like HiTOP.

"Depression appears to be a heterogeneous construct, likely reflecting multiple disorders with distinct etiologies (McGrath, 2005; Smith et al., 2009). Thus, an overall depression score is imprecise and may lead to uninterpretable findings. HiTOP compounds the problem by creating even larger groupings such as 'distress,' which includes not only depression, but also syndromes like Post Traumatic Stress Disorder and Generalized Anxiety Disorder. Distress is then combined with other heterogeneous groupings (e.g., fear, eating pathology, mania, sexual problems) under the umbrella of 'internalizing.' As one moves up the hierarchy, the scores become less and less useful."

This passage begins by disregarding the fact that HiTOP specifies narrower dimensions below depression, including anhedonia, dysphoria, lassitude, suicidality, agitation, etc. The lower levels of the HiTOP hierarchy provide a finer level of resolution that is lacking in DSM or ICD diagnoses. Further, note that HiTOP syndromes are not identical to DSM diagnoses, despite the fact that some DSM labels have been included in diagrams of the HiTOP system to aid in translating from DSM to HiTOP. The fact that symptoms associated with distress, fear, eating pathology, etc. tend to covary is indicated by the existence

of the internalizing dimension, which reflects, in part, that the lower-order constructs share some genetic influences (Patterson et al., 2018; Waldman et al., 2020). To claim that broader dimensions are generally less useful than the narrower dimensions beneath them is an unwarranted assertion, for which no good evidence exists. Further, utility, or lack thereof, cannot be asserted or evaluated in a vacuum, it must reference some purpose (e.g., prognosis, treatment selection, public health decisions). In any case, HiTOP includes both broad and narrow dimensions, thereby capitalizing on the potential usefulness of multiple levels of description.

"Despite the different symptom expressions, the DSM can identify these people as having the same problem [i.e., PTSD], in part, by requiring the presence of a common contributory cause (life threatening event)."

Here the authors claim that DSM can validly use the presence of a common cause (trauma) as evidence in identifying people with completely different symptoms as having the same disorder (PTSD). This argument fails to take into account the fact that trauma is a broad risk factor for psychopathology generally, including everything from depression to psychosis (e.g., Gibson et al., 2016). Moreover, there is evidence that non-traumatic life events elicit PTSD symptoms (e.g., Gold et al., 2005; Larsen & Pacella, 2016), leading to questions regarding whether DSM's approach to defining PTSD based on trauma exposure is valid (Rosen & Lilienfeld, 2008). Haeffel et al. appear to confuse convention (the decision of DSM committees to define a category of disorders based on trauma exposure) with empirical evidence that PTSD is a valid nosologic entity. Currently, the field has not reached consensus regarding the validity of PTSD and many concerns about it are actively being debated. We suspect that merely sharing a common contributory cause from an environmental exposure is not a particularly useful type of evidence for classification of psychopathology. (Also, note that, earlier, Haeffel et al. wrote, "It is not necessarily appropriate to conclude that the existence of common risk factors means that the disorders they influence should be considered 'the same,'" which directly contradicts their argument here.)

"Claim 3. HiTOP is Empirical and Objective"

"this is the same approach used by its predecessor, the five-factor model of personality."

HiTOP is not a descendent of the five factor model (FFM), which is a model of normal personality variation. HiTOP is a model of psychopathology, not normal personality. HiTOP was developed based on studies that did not consider normal personality. Parallels between HiTOP and the FFM are discussed in several consortium publications (e.g., Kotov et al., 2017; Widiger et al., 2019), but nowhere is HiTOP described as derivative of the FFM. HiTOP and the FFM simply share the use of statistical methods designed to understand covariance structure.

"multiple studies show that the complexity of human personality descriptors may be better represented by a spherical three-dimensional model than the more widely endorsed five factor model (e.g., Markey & Markey, 2006; Turkheimer et al., 2014).

This is a misrepresentation of the work by Markey and Markey (2006) and Turkheimer et al. (2014). Markey and Markey simply use three of the FFM dimensions (the two taken from the interpersonal circumplex are noted to be rotations of Agreeableness and Extraversion). They use these three dimensions to map constructs in a three dimensional space in a manner analogous to that traditionally used in

circumplex models in two dimensions (hence the "sphere"). This approach is not claimed by Markey and Markey to be either distinct from or better than the FFM. Turkheimer et al. (2014) similarly make no claim that spherical or three-dimensional models of personality provide a better representation of personality than the FFM.

"HiTOP may ultimately be a useful heuristic, but it is false to claim that it is an empirically validated or a data-driven realization of the structure of the symptoms of psychopathology."

This passage misrepresents the nature of scientific validation, which is not an all-or-nothing phenomenon. Rather, validity is a matter of degree. HiTOP is a model of constructs and their relations, and the validity of such a model reflects the degree of accuracy with which it describes dimensions of psychopathology and patterns of covariation that exist in reality, and this accuracy will partially determine its ability to facilitate scientific discovery and clinical intervention. Thus, it is meaningless to deny that HiTOP is "empirically validated" rather than to discuss the extent of its empirical validation (which is extensive, though far from complete; Kotov et al., 2020; Krueger et al., 2021; Watson et al., in press). Similarly, it is wrong to claim that HiTOP is not "data-driven," given that it is based on extensive modeling of data. Instead, Haeffel et al. may mean to take issue with the extent to which HiTOP's data-driven nature has yielded valid results. Again, we point to the extensive evidence that has been amassed for the system's validity (while acknowledging that many validity-related questions still remain unanswered).

"Data decisions are easy when there is a well-defined and circumscribed body of data. For example, input decisions for the five-factor model of personality, from which HiTOP was derived, are based on the lexical hypothesis."

HiTOP was in no way derived from the FFM. Further, although it is true that person-descriptive adjectives in the dictionary made input decisions relatively easy for lexical research on normal personality, it is similarly possible, in principle, to attempt to include every sign and symptom of psychopathology regularly encountered in clinical settings. HiTOP has not completed the task of including every such sign and symptom, but this is a core goal of HiTOP, and ongoing efforts are underway to ensure reasonably comprehensive coverage.

"Are the self-reported symptoms used to create the HiTOP factors all meaningful indicators of psychopathology (e.g., McGrane & Maul, 2020; Michell, 2000)? Further, how many important indicators are missing from the model?"

With reference to the first question, much of the research involved in creating HiTOP involved clinical assessments, clinician ratings, and diagnoses, rather than self-report per se. In relation to the second, the consortium is making ongoing efforts to ensure that the coverage of signs and symptoms modeled in HiTOP is as comprehensive as possible and to expand the HiTOP model as necessary to encompass them. Indeed, HiTOP has a Measurement Work Group whose primary task is the creation of a more comprehensive, fine-grained measure of psychopathology than currently exists. We welcome any communications that help to identify specific gaps in HiTOP's coverage.

"we already know that the data used by HiTOP are biased in terms of culture, race, age, and gender, as they come from studies using samples of Western, Educated, Industrial, Rich, Democratic (WEIRD) participants"

Similar to the vast majority of extant research in psychology and psychopathology, it is indeed the case that much of the research contributing to the HiTOP framework has been conducted on WEIRD participants. It is certainly true that more work must be done to determine the applicability of HiTOP, and psychopathology models generally, in diverse populations. Nonetheless, it is important to note that Kotov et al. (2017) summarized a diverse corpus of literature that provided the initial evidence for HiTOP and that included numerous cross-cultural studies (e.g., Ivanova et al., 2007, 2015; Kessler et al., 2011; Krueger et al., 2003; Rescorla et al., 2013). Thus, it is inaccurate to claim that evidence for HiTOP is limited to WEIRD participants. Further, HiTOP holds the possibility of encompassing various aspects of diversity within its model—for example, studies have been published examining the structure of the model in various sociodemographic groups (Eaton, 2014; Rodriguez-Seijas et al., 2019, 2020, 2021). The flexibility of HiTOP, as well as ongoing refinement based on novel evidence, positions it favorably to describe psychopathology among typically understudied populations. Additionally, HiTOP is useful for exploring and understanding various ways in which psychopathology manifests among understudied populations. For example, HiTOP-consistent measurement has been used to disentangle biases in the assignment of Borderline Personality Disorder diagnoses among sexual and gender minority persons (Rodriguez-Seijas et al., 2020; 2021).

"The lack of representation in psychological research is a problem for all taxonomies. However, it may be significantly more difficult for data-driven models like HiTOP to capture cultural nuance than it is for other approaches"

We believe it will be easier for HiTOP to capture cultural nuance than it is for the current psychiatric nosology or other taxonomic approaches like it. All that is required is the collection of data in diverse populations, followed by quantitative evaluation of differences in structure between populations.

"Claim 4. HiTOP Will Lead to Genetic Discovery"

"First, HiTOP probably is not valid; it is a descriptive taxonomy based on symptom correlations. There is little reason to believe that these groupings reflect any natural kinds for which causal genetic variants can be discovered."

First, the validity of HiTOP dimensions can be established through construct validation, just like any other psychological construct (Cronbach & Meehl, 1955; Loevinger, 1957; Watson, 2012). Considerable validity evidence already exists for many HiTOP dimensions (e.g., Kotov et al., 2020, 2021; Krueger et al, in press; Watson et al., 2021). Second, it is important to understand that correlations that are reliably nonzero do imply a causal connection between the variables involved, although the nature of that causal connection is indeterminate without further evidence. One variable may cause the other, they may mutually cause each other, or they may be correlated because some third variable or set of variables affects both, but something must link them causally. One reasonable possibility to investigate is that shared genetic influences cause their covariance.

"Second, there is the 'gloomy prospect' (Plomin & Daniels, 1987; Turkheimer & Waldron, 2000). Even if HiTOP somehow got everything right, it still would not lead to the identification of any genetic mechanisms. That is because there are no specific genetic mechanisms to be found (i.e., no 'mental illness genes'). Mental illness is too complex. Researchers are converging on the conclusion that complex behavioral phenotypes are likely the result of thousands of genes, each with a negligible effect (Turkheimer, 2016; Visscher et al., 2010)."

There are several problems with this passage. First, the references cited in the first sentence use the phrase "gloomy prospect" to describe the great difficulty in identifying systematic non-shared *environmental* causes of phenotypic variation, not to describe the difficulty of identifying specific genes involved in mental illness. Second, pessimism regarding the identification of specific genes has been largely invalidated, as large scale consortia for genetic research in the last decade have been remarkably successful in using genome wide association studies (GWAS) to identify replicable genetic variants associated with complex phenotypes, including diagnoses of psychopathology (Smoller et al., 2019). Although it is true that complex behavioral phenotypes are likely the result of variation in very large numbers of genes, sufficiently large samples have begun to enable the identification of these genes, and bioinformatic approaches have begun to allow the identification of the biological systems and pathways that these genes influence (Hyman, 2021). This research process begins to connect genetic variation with behavioral phenotypes, via the biological systems that govern the phenotypes. As but one example, large-scale GWAS of schizophrenia helped lead to the discovery and elucidation of *C4* as an important and heretofore unknown underlying risk mechanism that is involved in excess synaptic pruning in adolescence (Sekar et al., 2016). Currently, the field exhibits cautious optimism that psychiatric genetic research will inform other basic and social sciences, as well as clinical practice (Harden & Koellinger, 2020; Lewis & Vassos, 2020; Smoller, 2017). Already, polygenic risk scores have been used successfully to predict treatment responses, etc (Zhang, 2019). HiTOP is well poised to assist in the accelerating discoveries of modern genetics (Waszczuk et al., 2020).

"If the slow progress in this area was caused by poor DSM phenotypes, as claimed by the HiTOP consortium, then we should see success in other areas of social science that have better theories and measurement tools. This is not the case; researchers have yet to discover the genetic mechanism for *any* complex human phenotype (intelligence, personality, etc.; Matthews & Turkheimer, 2019)."

It is unreasonable to expect that science would already have discovered *the* genetic mechanism for complex human phenotypes, first, because there is no single genetic mechanism for the phenotypes in question but rather a great many, and second because the field is still so young. However, it is not the case that there has been no success; knowledge regarding genetic variants and associated biological pathways is increasing rapidly in relation to various complex human phenotypes, including intelligence and personality (e.g., Nagel et al., 2018; Savage et al., 2018). Haeffel et al. cite Turkheimer extensively, but Turkheimer is unusual in the field in his degree of pessimism, and current progress suggests that it is unwarranted. The benchmark that Haeffel et al. have in mind for genetic discovery seems to be a fully elucidated mechanism stretching from DNA to phenotype, akin to discoveries in Mendelian disorders. However, the HiTOP consortium's review of genetic research (Waszczuk et al., 2020) posited only that HiTOP phenotypes can aid in the identification of biological pathways associated with identified genetic variants, which is in line with the current state of the field.

"Turkheimer (2014) reviewed the literature on personality and heritability and concluded 'that in the genetics of personality, a paradoxical outcome that has been looming for a long time has finally come to pass: personality is heritable, but it has no genetic mechanism.'"

In the seven years since Turkheimer published this pessimistic conclusion, research on personality genetics has contradicted it. Hundreds of independent genetic loci associated with personality dimensions have been discovered, and bioinformatic approaches have identified biological pathways (i.e., mechanisms) associated with the identified genes (Montag et al., 2020; Nagel et al., 2018).

"It is also important to address the claim that heritability estimates and genetic correlations can be used to validate the HiTOP hierarchy (Waszczuk et al., 2019)" ... "It is not appropriate to use heritability estimates as a method for corroborating a taxonomy."

Waszczuk et al. (2020) did not state that heritability estimates can be used to validate HiTOP. (The correct year for the published article in question is 2020, not 2019.)

"This is because everything is heritable (Turkheimer's [2000] first law of behavioral genetics). All measurable human differences have genetic correlations."

Although Turkheimer's first law (the assertion that all behavioral variables are heritable) does indeed seem to be true, this is distinct from the question of whether all such variables show genetic correlations (in other words, correlations between the genetic influences on different variables). Haeffel et al. appear to conflate heritability estimates, which are univariate, with genetic correlations among multiple variables. Only the latter were reviewed by Waszczuk et al. (2020). It is not a given that all dimensions of psychopathology will be genetically correlated, and, indeed, the range of magnitudes of the genetic correlations varies greatly. These differences in the magnitudes of genetic correlations have already led to research that supports aspects of the HiTOP model at the genomic level (Waldman et al., 2020), offering further evidence that HiTOP dimensions are good targets for genetic research. Additionally, some phenotypic associations can be fully explained by environmental correlations, especially in developmental samples, meaning that they do *not* have associated genetic correlations. For example, eating pathology symptoms of emotional overeating and undereating are moderately correlated in children, and this association appears to be explained by common shared environmental influences (Herle et al., 2017, 2018). In short, genetic correlations are a valuable source of validity evidence for HiTOP.

"In summary, it will be difficult for HiTOP to fulfill its promise to accelerate genetic discovery (Waszczuki et al., 2019). It is another descriptive taxonomy that lumps people based on similar symptom presentations. It proposes a unique hierarchy, but the symptom heterogeneity in the upper-level spectra will likely hinder genetic discovery (Smith et al., 2009)."

Once again, Haeffel et al. repeat their mistaken claim that HiTOP "lumps people based on similar symptom presentations," when in fact it groups symptoms based on their tendency to co-occur within people. The resulting hierarchical structure actually facilitates genetic discovery because genetic variants can be associated either with upper-level spectra (indicating that they are likely to influence all of the lower-level dimensions within the given spectrum) or they can be specific to lower-level dimensions. Using HiTOP, level of the hierarchy at which the genetic variant operates can be formally tested.

"Claim 5. HITOP is Ready to Use Today"

"Patients are going to score high on multiple spectra, subfactors, and disorders. How will a clinician interpret all of these scores?"

Evidence exists that clinicians find the use of assessments of HiTOP dimensions helpful in their clinical practice (e.g., Bornstein et al., 2019; McGrath, Rashid, Hayman, & Pogge, 2002; Morey, Skodol, & Oldham, 2014). Certainly, more research needs to be done to ensure that assessments of HiTOP dimensions are maximally useful for clinicians, but in some ways HiTOP is simply a formalization of what many clinicians do already when they attempt to understand the individual patient and to tailor their intervention to suit the person's individual personality and symptom profile, regardless of the formal diagnosis. Hence, it is not surprising that they find the approach reasonable and useful. Further, clinicians have been using similar profiles from instruments like the MMPI and CBCL/ASEBA for decades, so the basic approach underlying clinical application of HiTOP is not new or untested.

"Currently, there are no established norms or clinical cutoffs, no information for identifying primary versus secondary problems, no interpretation or treatment guidelines, etc. To date, there is not even a standardized measure that can assess the entire HiTOP taxonomy, which means clinicians are on their own to piece together an assessment and then somehow interpret the patchwork of results."

The HiTOP consortium has recommended specific measures that have norms, cutoffs, and interpretive guidelines and are currently in use clinically (see Kotov et al., 2017). These can be used to assess any part of the HiTOP model. In fact, a specific battery of measures is recommended in the resource cited by Ruggero et al. (2019), namely: https://hitop.unt.edu/clinical-tools/hitop-friendly-measures. In HiTOP, diagnosis would consist of a profile of scores across psychopathology dimensions. These scores are referenced to the general population, akin to cognitive tests or psychopathology inventories such as the Child Behavior Checklist (Achenbach & Rescorla, 2003). The profile can be succinctly summarized as elevations of certain extremity (e.g., two standard deviations above population mean). HiTOP includes several broad spectra and subfactors as well as many narrow symptom dimensions within them. However, many referral questions do not require an exhaustive assessment and can focus on the most relevant HiTOP constructs. In particular, some diagnostic questions can be answered by characterizing the patient just in terms of the six spectra. Other questions may focus on one spectrum and assess its constituent dimensions. Profiles have a long and successful track record in neuropsychology and psychopathology assessment, and HiTOP aims to expand this strategy to all forms of psychopathology.

"As noted by Conway and colleagues (2019) 'many of the analyses that we have reviewed were carried out using datasets that were not assembled with HiTOP in mind' (p. 428). In other words, support for HiTOP has not actually come from using HiTOP."

The second sentence here appears to confuse HiTOP as a taxonomic model with HiTOP as it could be assessed by a dedicated assessment tool. "Using HiTOP" means using the taxonomy of dimensions. Many existing instruments assess the dimensions of HiTOP and therefore can be used to provide evidence in support of the HiTOP model, regardless of whether they were administered with HiTOP in mind. The consortium is working to build a dedicated assessment tool for all of HiTOP, but HiTOP can already be used without it.

"A Comparison of Taxonomies"

"Similarly, in factor analysis, there are decisions about mode of representation and how to deal with rotational indeterminacy; the consequence being that HiTOP is not anymore 'empirical' or 'truthful' than the DSM approach."

HiTOP's approach to classification using analysis of covariance structure is certainly more quantitative than DSM's approach using expert consensus. Whether that makes it more "truthful" cannot be prejudged on the grounds there are subjective decisions involved in factor analysis (as there are in any statistical analysis). Judging truthfulness or verisimilitude requires validity evidence, and considerable evidence exists to show that HiTOP dimensions are typically more valid and, as noted above, more reliable and predictive (i.e. useful) than DSM diagnoses (Kotov et al., 2020, 2021; Krueger et al., in press; Watson et al., 2021). Further, structural analyses, even when simply correlational, provide evidence about possible causal structure in a manner that expert consensus cannot (Markon & Jonas, 2016).

"It is important to underscore that the decision to parse the landscape of psychopathology into categories or facets is based more on expedience than empirical evidence (Turkheimer, 2017)."

We disagree with this assertion. The science of taxometrics was developed specifically to test quantitatively whether variables are better characterized as categories or continua. As taxometric methods have improved, no common form of psychopathology has yet shown evidence of being categorical rather than continuous (Haslam, 2019; Haslam et al., 2020). The decision is therefore based on empirical evidence.

"HiTOP, on the other hand, exhibits few, if any, of the features found in a useful taxonomy. Its classification system is an interpretation of factor analytic results. It is a single picture. Absent our knowledge and previous experience with DSM descriptions and disorders, HiTOP contains no additional information. It contains no explanations, no descriptive information (other than symptom labels and lists), no necessary symptoms, no inclusion or exclusion criteria, no information about how to integrate impairment severity, no information about prevalence, no information on underlying developmental processes, and it ignores differences in culture, age and/or gender."

This characterization of HiTOP is seriously misleading because it ignores the enormous body of existing research on dimensional features of psychopathology. Because those dimensions are classified and organized by HiTOP, HiTOP can be used to synthesize a large amount of information about nearly every feature of psychopathology within it. That information is partially descriptive (e.g., behavioral and biological correlates) and partially explanatory (e.g., findings on treatment response in randomized clinical trails), based on existing scientific knowledge about the nature and sources of various dimensions. (As just one example, knowledge exists about the role of the extended amygdala in the internalizing spectrum, based on many different modalities of evidence; Hur et al., 2019.) In short, HiTOP allows integration of a great deal of existing scientific knowledge, as well as a strategy for building systematically on that knowledge and facilitating translation to clinical practice (Kotov et al., 2020; Krueger et al., in press; Watson et al., 2021; Ruggero et al., 2019). Perhaps the only thing that is reasonably accurate in the quoted passage is that HiTOP does not provide "inclusion or exclusion criteria," but that is because it is a dimensional rather than a categorical system, and inclusion and

exclusion make sense only in relation to categories. Rather than trying to find the right category for a person (as in DSM), HiTOP characterizes a person's psychopathology across all spectra. The consortium has an ongoing project to integrate impairment severity into HiTOP, and information about prevalence and differences in culture, age, and gender are already being investigated, as noted above (Eaton, 2014; Eaton et al., 2011, 2012; Rodriguez-Seijas et al., 2019).

"Meehlian taxometrics is not usable in clinical settings, but it is more scientifically progressive than HiTOP and DSM."

Taxometrics cannot properly be compared to either HiTOP or DSM because it is not itself a nosology or taxonomy. Rather, it is a statistical method for testing hypotheses about the existence of categorical entities ("taxa"). Further, taxometrics has provided strong evidence to favor a taxonomy of continuous dimensions (such as HiTOP) in which features of psychopathology are classified, rather than a taxonomy of discrete categories (such as DSM) in which people are classified (Haslam, 2019; Haslam et al., 2020). Indeed, the taxometrics literature is the main reason that HiTOP does not contain any categorical taxa at present, and this contribution of taxometrics to HiTOP has been described in consortium publications (e.g., Kotov et al., 2017).

"Recommendation"

"It is possible that HiTOP could also meet this need at some point, but it is ultimately handcuffed by its inability to evolve over time."

HiTOP is explicitly designed to evolve over time, as mentioned in multiple consortium publications, including the first (Kotov et al., 2017). The consortium has recently developed a formal revisions process to facilitate continual revision of the system (Kotov et al., 2021).

"Conclusion"

"Unless psychopathology plays by a different set of rules than nearly every other realm of nature, the result of pushing the factor analysis button is an incorrect answer."

Factor analysis does not necessarily or always produce a model that corresponds exactly to the underlying data-generating process, but simulation studies have shown that it can, and it can be used to produce evidence regarding possible causal structure (Markon & Jonas, 2016). Thus, the answers it provides are not necessarily incorrect. Nor is factor analysis by any means the only statistical procedure used in the construction of HiTOP. Statistical analyses such as factor analysis are just the first step, and results must be validated by subsequent research. The HiTOP consortium is committed to precisely this sort of research.

"In order for HiTOP to be valid, it would mean that: 1) self-reported symptom expressions are meaningful indicators of development processes and the etiology of psychopathology; 2) all of the symptom indicators are equally important (deserve equal weighting) for classifying psychopathology; 3) equifinality and multifinality do not apply to psychopathology; 4) the

expression and reporting of symptoms are not influenced by sex, culture, or age (and failing to account for them does not lead to algorithmic bias); and 5) a dimensional interpretation/simple structure approach represents the structure of psychopathology symptom data."

HiTOP's validity requires none of these assumptions. 1) HiTOP relies on clinician, observer, and peer judgments as well as self-reports (further, we note that, despite some limitations, self-reports have considerable evidence for their validity in general, and there is no reason to think that they cannot therefore be informative in the etiology of psychopathology). 2) Factor analyses yield results in which different variables are weighted differently (by having different factor loadings); thus, it does not assume that all symptom indicators are equally important. 3) HiTOP recognizes that different causal processes can produce similar symptoms (equifinality) and that similar causal processes can produce different causal outcomes in different people (multifinality); it makes no claim for a unitary etiology of each dimension. 4) HiTOP recognizes and incorporates information regarding structural differences in different sociodemographic groups. 5) HiTOP follows the data, which, to date, suggest that most features of psychopathology are dimensional rather than categorical. Should evidence arise for nondimensional phenomena in psychopathology, HiTOP will include them.

"Psychology's statistical-driven approach to classification seems to fail this critical requirement, as it is difficult to 'be wrong' in the absence of any specific theoretical hypotheses while reporting the output of factor analyses."

This claim overlooks the existence of confirmatory factor analysis (a technique used extensively in the development of HiTOP) as well as various related statistical approaches, all of which allow testing the fit of individual models to the data and comparing the fit of different models to each other. Thus, it is entirely possible for a particular factor model to be wrong and for quantitative, empirical evidence to show that it is wrong. The development of HiTOP has involved a substantial amount of testing specific hypotheses regarding the covariance structure of psychopathology. Additionally, it is an explicit goal of HiTOP to seek external validation of the resulting structure via convergent and discriminant associations with many other existing variables and outcomes, consistent with well-established construct validation approaches (Cronbach & Meehl, 1955; Watson, 2012).

References (excluding those cited only in quotation)

Bornstein, R. F., & Natoli, A. P. (2019). Clinical utility of categorical and dimensional perspectives on personality pathology: A meta-analytic review. Personality Disorders: Theory, Research, and Treatment, 10(6), 479.

Caspi, A., Houts, R. M., Ambler, A., Danese, A., Elliott, M. L., Hariri, A., ... & Moffitt, T. E. (2020). Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the Dunedin Birth Cohort Study. *JAMA network open*, *3*(4), e203221-e203221.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, *52*(4), 281.

Eaton, N. R. (2014). Transdiagnostic psychopathology factors and sexual minority mental health: evidence of disparities and associations with minority stressors. *Psychology of Sexual Orientation and Gender Diversity*, *1*(3), 244-254.

Eaton, N. R., Keyes, K. M., Krueger, R. F., Balsis, S., Skodol, A. E., Markon, K. E., ... & Hasin, D. S. (2012). An invariant dimensional liability model of gender differences in mental disorder prevalence: evidence from a national sample. *Journal of abnormal psychology*, *121*(1), 282-288.

Eaton, N. R., Krueger, R. F., & Oltmanns, T. F. (2011). Aging and the structure and long-term stability of the internalizing spectrum of personality and psychopathology. *Psychology and Aging*, *26*(4), 987-993.

Gibson, L. E., Alloy, L. B., & Ellman, L. M. (2016). Trauma and the psychosis spectrum: A review of symptom specificity and explanatory mechanisms. *Clinical Psychology Review*, *49*, 92-105.

Gold, S. D., Marx, B. P., Soler-Baillo, J. M., & Sloan, D. M. (2005). Is life stress more traumatic than traumatic stress?. *Journal of Anxiety Disorders*, *19*(6), 687-698.

Grunwald, P. D. (2007). *The minimum description length principle*. Cambridge, MA: MIT Press.

Harden, K. P., & Koellinger, P. D. (2020). Using genetics for social science. *Nature Human Behaviour, 4*(6), 567-576.

Haslam, N. (2019). Unicorns, snarks, and personality types: A review of the first 102 taxometric studies of personality. *Australian Journal of Psychology*, *71*(1), 39-49.

Haslam, N., McGrath, M. J., Viechtbauer, W., & Kuppens, P. (2020). Dimensions over categories: A meta-analysis of taxometric research. *Psychological Medicine*, *50*(9), 1418-1432.

Herle, M., Fildes, A., & Llewellyn, C. (2018). Emotional eating is learned not inherited in children, regardless of obesity risk. *Pediatric obesity, 13*(10), 628-631.

Herle, M., Fildes, A., Steinsbekk, S., Rijsdijk, F., & Llewellyn, C. H. (2017). Emotional over-and under-eating in early childhood are learned not inherited. *Scientific reports, 7*(1), 1-9.

Hur, J., Stockbridge, M. D., Fox, A. S., & Shackman, A. J. (2019). Dispositional negativity, cognition, and anxiety disorders: An integrative translational neuroscience framework. *Progress in brain research*, *247*, 375–436.

Hyman, S. E. (2021). Psychiatric Disorders: Grounded in Human Biology but Not Natural Kinds. *Perspectives in Biology and Medicine*, *64*(1), 6-28.

Ivanova, M. Y., Achenbach, T. M., Dumenci, L., Rescorla, L. A., Almqvist, F., Weintraub, S., ... Döpfner, M. (2007). Testing the 8-syndrome structure of the child behavior checklist in 30 societies. Journal of Clinical Child and Adolescent Psychology, 36, 405–417.

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Turner, L. V., Ahmeti-Pronaj, A., Au, A., ... Zasepa, E. (2015). Syndromes of self-reported psychopathology for ages 18–59 in 29 societies. Journal of Psychopathology and Behavioral Assessment, 37, 171–183.

Jang, K. L., Hu, S., Livesley, W. J., Angleitner, A., Riemann, R., & Vernon, P. A. (2002). Genetic and environmental influences on the covariance of facets defining the domains of the five-factor model of personality. *Personality and Individual Differences, 33,* 83–101.

Jang, K. L., McCrae, R. R., Angleitner, A., Riemann, R., & Livesley, W. J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology, 74,* 1556–1565. doi:10.1037/0022-3514.74.6.1556

Jonas, K. G., & Markon, K. E. (2016). A descriptivist approach to trait conceptualization and inference. *Psychological Review*, *123*(1), 90.

Kessler, R. C., Ormel, J., Petukhova, M., McLaughlin, K. A., Green, J. G., Russo, L. J., ... & Uestuen, T. B. (2011). Development of lifetime comorbidity in the World Health Organization world mental health surveys. *Archives of General Psychiatry*, *68*(1), 90-100.

Kotov, R., Foti, D., Li, K., Bromet, E. J., Hajcak, G., & Ruggero, C. J. (2016). Validating dimensions of psychosis symptomatology: Neural correlates and 20-year outcomes. Journal of Abnormal Psychology, 125, 1103-1119.

Kotov, R., Jonas, K. G., Carpenter, W. T., Dretsch, M. N., Eaton, N. R., Forbes, M. K., ... & HiTOP Utility Workgroup. (2020). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): I. Psychosis superspectrum. *World Psychiatry*, *19*(2), 151-172.

Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, et al. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4):454–77.

Kotov, R., Krueger, R. F., Watson, D., Cicero, D. C., Conway, C. C., DeYoung, C. G., ... & Wright, A. G. (2021). The Hierarchical Taxonomy of Psychopathology (HiTOP): A quantitative nosology based on consensus of evidence. *Annual Review of Clinical Psychology*, *17*, 19.1–19.26.

Krueger, R. F., Chentsova-Dutton, Y. E., Markon, K. E., Goldberg, D., & Ormel, J. (2003). A cross-cultural study of the structure of comorbidity among common psychopathological syndromes in the general health care setting. *Journal of Abnormal Psychology, 112*, 437–447.

Krueger, R.F., Hobbs, K.A., Conway, C.C., Dick, D.M., Dretsch, M.N., Eaton, N.R., Forbes, M.K., Forbush, K.T., Keyes, K.M., Latzman, R.D., Michelini, G. Patrick, C.J., Sellbom, M., Slade, T., South, S.C., Sunderland, M., Tackett, J., Waldman, I., Waszczuk, M.A., Wright, A.G.C., Zald, D.H., Watson, D., Kotov, R., & HiTOP Utility Workgroup. (2021). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): II. Externalizing superspectrum. *World Psychiatry, 20,* 171–193.

Larsen, S. E., & Pacella, M. L. (2016). Comparing the effect of DSM-congruent traumas vs. DSM-incongruent stressors on PTSD symptoms: A meta-analytic review. *Journal of Anxiety Disorders, 38*, 37-46.

Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome medicine, 12*, 1-11.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological reports*, *3*(3), 635-694.

Markey, P.M. & Markey, C.N. (2006). A spherical conceptualization of personality traits. *European Journal of Personality, 20,* 169-193. https://doi.org/10.1002%2Fper.582

Markon, K. E., & Jonas, K. G. (2016). Structure as cause and representation: Implications of descriptivist inference for structural modeling across multiple levels of analysis. *Journal of Abnormal Psychology*, *125*(8), 1146.

Michelini, G., Barch, D. M., Tian, Y., Watson, D., Klein, D. N., & Kotov, R. (2019). Delineating and validating higher-order dimensions of psychopathology in the Adolescent Brain Cognitive Development (ABCD) study. *Translational Psychiatry*, *9*, 1-15.

Montag, C., Ebstein, R. P., Jawinski, P., & Markett, S. (2020). Molecular genetics in psychology and personality neuroscience: On candidate genes, genome wide scans, and new research strategies. *Neuroscience & Biobehavioral Reviews*.

McCrae, R. R., Jang, K. L., Ando, J., Ono, Y., Yamagata, S., Riemann, R., … Spinath, F. M. (2008). Substance and artifact in the higherorder factors of the big five. *Journal of Personality and Social Psychology, 95*, 442–455. doi:10.1037/0022-3514.95.2.442

McGrath, R. E., Rashid, T., Hayman, J., & Pogge, D. L. (2002). A comparison of MMPI-2 high-point coding strategies. *Journal of Personality Assessment*, *79*, 243–256.

Morey, L. C., Skodol, A. E., & Oldham, J. M. (2014). Clinician judgments of clinical utility: A comparison of DSM-IV-TR personality disorders and the alternative model for DSM-5 personality disorders. *Journal of Abnormal Psychology*, *123*, 398–405.

Mõttus, R., Sinick, J., Terracciano, A., Hřebíčková, M., Kandler, C., Ando, J., ... & Jang, K. L. (2019). Personality characteristics below facets: A replication and meta-analysis of cross-rater agreement, rank-order stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *117*(4), e35.

Nagel, M., Jansen, P. R., Stringer, S., Watanabe, K., De Leeuw, C. A., Bryois, J., ... & Posthuma, D. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature genetics*, *50*(7), 920–927.

Nickels, M. K., & Nelson, C. E. (2005). Beware of nuts & bolts: Putting evolution into the teaching of biological classification. *The American Biology Teacher*, *67*(5), 283-289.

Patterson, M. W., Mann, F. D., Grotzinger, A. D., Tackett, J. L., Tucker-Drob, E. M., & Harden, K. P. (2018). Genetic and environmental influences on internalizing psychopathology across age and pubertal development. *Developmental psychology*, *54*(10), 1928.

Pesenti-Gritti, P., Spatola, C.A.M., Fagnani, C. *et al.* (2008). The co-occurrence between internalizing and externalizing behaviors. *Eur Child Adolesc Psychiatry* 17, 82–92.

Rescorla, L. A., Ginzburg, S., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., ... & Verhulst, F. C. (2013). Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *Journal of Clinical Child & Adolescent Psychology, 42*(2), 262-273.

Rodriguez-Seijas, C., Eaton, N. R., & Pachankis, J. E. (2019). Prevalence of psychiatric disorders at the intersection of race and sexual orientation: Results from the National Epidemiologic Survey of Alcohol and Related Conditions-III. *Journal of Consulting and Clinical Psychology*, *87*(4), 321-331.

Rodriguez-Seijas, C., Morgan, T. A., & Zimmerman, M. (2020). Is There a Bias in the Diagnosis of Borderline Personality Disorder Among Lesbian, Gay, and Bisexual Patients?. *Assessment*, *28*(3), 724-738.

Rodriguez-Seijas, C., Morgan, T. A., & Zimmerman, M. (2021). A Population-Based Examination of Criterion-Level Disparities in the Diagnosis of Borderline Personality Disorder Among Sexual Minority Adults. *Assessment*.

Rosen, G. M., & Lilienfeld, S. O. (2008). Posttraumatic stress disorder: An empirical evaluation of core assumptions. *Clinical Psychology Review, 28*(5), 837-868.

Ruggero CJ, Kotov R, Hopwood CJ, First M, Clark LA, et al. 2019. Integrating the Hierarchical Taxonomy of Psychopathology (HiTOP) into clinical practice. *J. Consult. Clin. Psychol.* 87(12):1069–84

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C. A., ... & Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature genetics*, *50*(7), 912-919.

Sekar, A., Bialas, A. R., De Rivera, H., Davis, A., Hammond, T. R., Kamitaki, N., ... & McCarroll, S. A. (2016). Schizophrenia risk from complex variation of complement component 4. *Nature*, *530*(7589), 177-183.

Smoller, J. W. (2017). A Quarter Century of Progress in Psychiatric Genetics. *Harvard review of psychiatry, 25*(6), 256-258.

Smoller, J. W., Andreassen, O. A., Edenberg, H. J., Faraone, S. V., Glatt, S. J., & Kendler, K. S. (2019). Psychiatric genetics and the structure of psychopathology. *Molecular psychiatry*, *24*(3), 409-420.

Turkheimer, E., Pettersson, E., & Horn, E.E. (2014). A Phenotypic Null Hypothesis for the Genetics of Personality. *Annual Review of Psychology,65*, 515–40.

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological methods*, *17*(2), 228.

Waldman, I. D., Poore, H. E., Luningham, J. M., & Yang, J. (2020). Testing structural models of psychopathology at the genomic level. *World Psychiatry*, *19*(3), 350-359.

Waszczuk, M. A., Eaton, N. R., Krueger, R. F., Shackman, A. J., Waldman, I. D., Zald, D. H., ... & Kotov, R. (2020). Redefining phenotypes to advance psychiatric genetics: Implications from hierarchical taxonomy of psychopathology. *Journal of abnormal psychology*, *129*(2), 143.

Watson, D. (2012). *Objective tests as instruments of psychological theory and research.* In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbooks in psychology. APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics* (p. 349–369). Washington, D.C.: American Psychological Association.

Watson, D., Levin-Aspenson, H. F., Waszczuk, M. A., Conway, C. C., Dalgleish, T., Dretsch, M. N., & Workgroup, H. U. (2021). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): III. Emotional dysfunction superspectrum. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*.

Widiger, T. A., Sellbom, M., Chmielewski, M., Clark, L. A., DeYoung, C. G., Kotov, R., ... & Wright, A. G. (2019). Personality in a hierarchical model of psychopathology. Clinical Psychological Science, 7(1), 77-92.

Zald, D. H., & Lahey, B. B. (2017). Implications of the hierarchical structure of psychopathology for psychiatric neuroimaging. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *2*(4), 310-317.

Zhang, J. P., Robinson, D., Yu, J., Gallego, J., Fleischhacker, W. W., Kahn, R. S., ... & Lencz, T. (2019). Schizophrenia polygenic risk score as a predictor of antipsychotic efficacy in first-episode psychosis. *American Journal of Psychiatry*, *176*(1), 21-28.