Principles and procedures for revising the Hierarchical Taxonomy of Psychopathology

Miriam K. Forbes, School of Psychological Sciences, Macquarie University;
ORCiD: 0000-0002-6954-3818

Whitney R. Ringwald, University of Pittsburgh;
ORCiD: 0000-0002-8883-5963

Timothy Allen, University of Pittsburgh;
ORCiD: 0000-0002-9487-9233

David C. Cicero, Department of Psychology, University of North Texas,
ORCiD: 0000-0002-5666-9139

Lee Anna Clark, Department of Psychology, University of Notre Dame,
ORCiD: 0000-0002-2571-4347

Colin G. DeYoung, Department of Psychology, University of Minnesota;
ORCID: 0000-0001-5621-1091

Nicholas Eaton, Department of Psychology, Stony Brook University, Stony Brook, NY, USA

Roman Kotov, Department of Psychiatry, Stony Brook University

Robert F. Krueger, Department of Psychology, University of Minnesota,
ORCiD: 0000-0001-9127-5509

Robert D. Latzman, Data Sciences Institute, Takeda;
ORCiD: 0000-0002-1175-8090

Elizabeth A. Martin, Department of Psychological Science, University of California, Irvine;
ORCiD 0000-0001-8893-1667

Kristin Naragon-Gainey, School of Psychological Science, University of Western Australia;
ORCiD: 0000-0002-2245-6880

Camilo J. Ruggero, Department of Psychology, University of North Texas,
ORCiD: 0000-0002-8482-6269

Irwin D. Waldman, Department of Psychology, Emory University,
ORCiD: 0000-0002-6862-1837

Cassandra Brandes, Department of Psychology, Northwestern University; Department of Psychiatry, University of Illinois at Chicago;
ORCiD: 0000-0002-8015-547X

Eiko I. Fried, Department of Clinical Psychology, Leiden University, The Netherlands

Vina M. Goghari, Department of Psychological Clinical Science, University of Toronto

Benjamin Hankin, Department of Psychology, University of Illinois Urbana Champaign.
ORCiD: 0000-0003-2446-0258

Sarah Sperry, University of Michigan, Department of Psychiatry,
ORCiD: 0000-0002-1111-7357

Kasey Stanton, University of Wyoming, Department of Psychology,
ORCiD: 0000-0002-4519-5214

Awais Aftab, Department of Psychiatry, Case Western Reserve University,
ORCiD: 0000-0002-3515-3993

Donald Lynam, Department of Psychological Sciences, Purdue University
ORCiD: 0000-0001-8306-498X

Michael Roche, Department of Psychology, West Chester University of Pennsylvania

Aidan G. C. Wright, Department of Psychology and the Eisenberg Family Depression Center, University of Michigan
ORCiD: 0000-0002-2369-0601

Correspondence to: Miriam K. Forbes, PhD
Australian Hearing Hub, Room 3.803
School of Psychological Sciences,
Macquarie University,
Sydney Australia 2109
Email: miri.forbes@mq.edu.au

Acknowledgements

Authors are listed in blocks that are intended to reflect their contributions to the manuscript and to developing the revisions process itself, although we note that these are crude distinctions, and the work has been a team effort spanning 5 years. Miriam K. Forbes and Aidan G. C. Wright led and oversaw the process of developing the revisions framework, as well as drafting and revising this manuscript. MKF also made the figures for the manuscript. Timothy Allen, David C. Cicero, Lee Anna Clark, Colin G. DeYoung, Nicholas Eaton, , Roman Kotov, Robert F. Krueger, Robert D. Latzman, Elizabeth A. Martin, Kristin Naragon-Gainey, Camilo J. Ruggero, and Irwin D. Waldman are listed in alphabetical order by last name and each 1) made major contributions to the development of the revisions process (i.e., in "mini workgroups" either testing out iterations of the revisions protocol, or working through problems identified in these tests), 2) drafted original text in the paper, and 3) gave feedback on multiple drafts of the paper. Whitney R. Ringwald met these criteria in addition to running the first trial run of the penultimate revisions protocol and providing administrative support preparing the paper for submission, so is moved up and out of alphabetical order in the first author block. Cassandra Brandes, Eiko I. Fried, Vina M. Goghari, Benjamin Hankin, Sarah Sperry, and Kasey Stanton are listed in alphabetical order by last name and each 1) made contributions to the development of the revisions process outside of the formal mini workgroups, 2) drafted original text in the paper, and 3) gave feedback on multiple drafts of the paper. Awais Aftab, Donald Lynam, and Michael Roche are listed in alphabetical order by last name, and each made other contributions to the revisions process and paper.

We also thank Benjamin Lahey, Leonard Simms, Matthew Sunderland, and Thomas Widiger, who all made formal contributions to the development of the revisions process prior to drafting the manuscript. Jennifer Tackett and Noah Venables also made contributions to the development of the revisions process in workgroup meetings as members of the HiTOP Revisions Workgroup. Thanks also to David Preece who provided feedback on the manuscript prior to resubmission.

Transparency and Openness: All study materials are available on the OSF page for the project: https://osf.io/8h7m6/

Abstract

Quantitative, empirical approaches to establishing the structure of psychopathology hold promise to improve on traditional psychiatric classification systems. The Hierarchical Taxonomy of Psychopathology (HiTOP) is a framework that summarizes the substantial and growing body of quantitative evidence on the structure of psychopathology. To achieve its aims, HiTOP must incorporate emerging research in a systematic, ongoing fashion. In this paper, we describe the historical context and grounding of the principles and procedures for revising the HiTOP framework. Informed by strengths and shortcomings of previous classification systems, the proposed revisions protocol is a formalized system focused around three pillars: 1) prioritizing systematic evaluation of quantitative evidence by a set of transparent criteria and processes, 2) balancing stability with flexibility, and 3) promoting inclusion over gatekeeping in all aspects of the process. We detail how the revisions protocol will be applied in practice, including the scientific and administrative aspects of the process. Additionally, we describe areas of the HiTOP structure that will be a focus of early revisions and outline challenges for the revisions protocol moving forward. The proposed revisions protocol is designed to ensure that the HiTOP framework reflects the current state of scientific knowledge on the structure of psychopathology and fulfils its potential to advance clinical research and practice.

General Scientific Summary: The Hierarchical Taxonomy of Psychopathology (HiTOP) aims to provide an empirically derived classification system for psychopathology. To achieve this aim, the HiTOP model needs to be able to evolve to reflect new and new perspectives on existing data. This paper explains the rationale for the principles and procedures for revising the model.

Keywords: classification; nosology; psychopathology; psychiatry; HiTOP

Principles and procedures for revising the Hierarchical Taxonomy of Psychopathology

The Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017; 2021) is intended to be an empirically derived structure of psychopathology. The existing HiTOP structure organizes features of psychopathology into a series of hierarchical dimensions ranging from narrow signs and symptoms to broad spectra according to their patterns of covariance (see Figure 1). It was established based on a comprehensive review of a substantial body of structural validity evidence (Kotov et al., 2017) and has since been supported by follow-up reviews (Kotov et al., 2020, 2021; Krueger et al., 2021; Watson et al., 2022) and a meta-analysis (Ringwald et al., 2022). However, to fulfil its goals, HiTOP cannot be static and must evolve based on new data, incorporating findings from emerging research in a systematic, ongoing fashion. After all, empirical work—especially in a science as complex as psychopathology—is an ongoing process rather than an endpoint.

In this paper, we describe the historical context and conceptual and empirical grounding of the principles and procedures developed by the HiTOP Revisions Workgroup for modifying the HiTOP model. We begin with a brief historical account of relevant approaches to establishing the structure of psychopathology and its classification, exploring the dual roles of classification in achieving clinical utility and scientific accuracy and how they map onto the validity of diagnostic constructs. We then briefly describe—in a necessarily selective review—how other nosological systems have approached revisions to existing models and developed guidelines to integrate new information. The organizing principles and core assumptions of the revisions process build on these rich histories of classification and related approaches to model revision. We go on to introduce the protocol for revising the HiTOP structure and outline how it will be applied in practice. Finally, we describe areas of the HiTOP model that will be a focus of early revisions and unresolved challenges for the proposed revisions protocol.

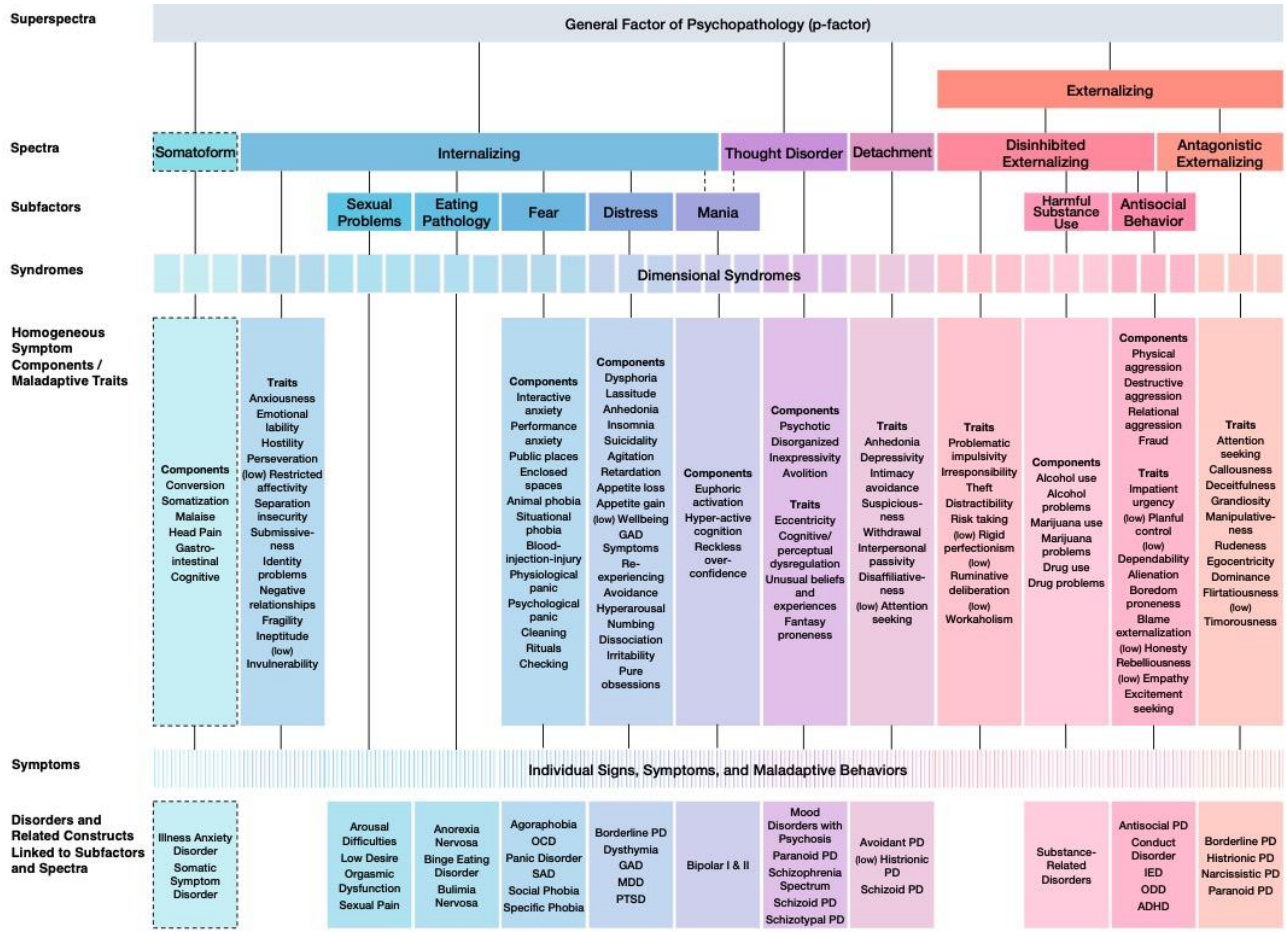

Figure 1. The official baseline HiTOP framework, as described in Kotov et al. (2017). The figure combines the information from Figures 2 and 3 in Kotov et al., as well as features described in the text. Dashed lines indicate dimensions included as provisional aspects of the framework. Notably, the "disorders and related constructs linked to subfactors and spectra" are not formal parts of the framework but were listed in Figure 2 in Kotov et al. "for convenience of communication" (p. 461) to identify the constructs that have been used in many studies of the higher order dimensions. ADHD = attention-deficit/hyperactivity disorder; GAD = generalized anxiety disorder; IED = intermittent explosive disorder; MDD = major depressive disorder; OCD = obsessive–compulsive disorder; ODD = oppositional defiant disorder; PD = personality disorder; PTSD = posttraumatic stress disorder; SAD = separation anxiety disorder.

A Brief Historical Account of Influential Approaches to Psychopathology Classification

Though the thread of scientific classification of psychopathology can be traced back to antiquity with Galen's four categories of temperament, multiple strands of contemporary descriptive psychopathology inform HiTOP's development and refinement (Williams & Simms, 2020). The most notable include psychiatric classification and the multivariate or quantitative approach to personality and psychopathology (Blashfield, 1984). Psychiatric classification's modern history dates back only to mid-19th century attempts by European epidemiologists and

statisticians to develop an internationally applicable classification of causes of death (for details, see Clark et al., 2017). After several revisions, the resulting "International List of Causes of Death" was eventually broadened into the *International Classification of Diseases, Injuries, and Causes of Death*—known as *ICD-6* (World Health Organization [WHO], 1949)—providing an integrated classification of morbidity and mortality to facilitate internationally consistent reporting of health information. Shortly after, the first version of the *Diagnostic and Statistical Manual of Mental Disorders (DSM;* American Psychiatric Association [APA], 1952) was published following several iterations of an effort to document diagnoses in large psychiatric hospitals. There was substantial collaboration between the WHO and the APA in developing the *ICD-8* (WHO, 1967) and *DSM-II* (APA, 1968), with the *ICD-8* introducing "the first predominantly symptom-based modern classification of mental disorders," which ultimately gave rise to the current descriptive, operationalist approach to psychiatric classification (Fulford & Sartorius, 2009, pp. 30).

The *DSM-III* (APA, 1980) fully embodied the transition to a descriptive, operational approach and was strongly influenced by the psychiatric school of thought from the Washington University in St. Louis School of Medicine, which had produced the Feighner Criteria (Feighner et al., 1972) and the Research Diagnostic Criteria (Spitzer & Robins, 1978). The *DSM-III* represented a marked philosophical shift in psychiatric nosology that ushered in the contemporary zeitgeist—rejecting psychoanalytic theory and adopting a medical model under the guise of an atheoretical diagnostic system (Aftab & Ryznar, 2021). Specifically, the *DSM-III* exemplified the full commitment to a philosophical model that typically has viewed psychiatric diagnoses as putative natural entities (e.g., Robins & Guzé, 1970). Although there was a strong emphasis within this system on subjecting constructs and criteria sets to rigorous empirical validation and iterative refinement, the core assumptions about diagnoses and disorders reflected those of the medical model.

Developing alongside these efforts at descriptive psychiatric classification, the quantitative empirical approach emerged from work to infer and describe the structures of a variety of domains

of human functioning (e.g., intelligence and personality; see Wright, 2017 for a review). The principal logic of this approach is that quantitative techniques (e.g., latent variable modelling or cluster analysis) can be enlisted to study patterns of co-occurrence or covariation in features of psychopathology to infer the underlying structure of mental disorder. Efforts to apply the quantitative empirical approach to psychopathology can be found throughout the 20th century, with some early notable examples being Moore's (1930, 1933), Lorr's (1964), and Achenbach's work (1966, 1978). Indeed, Achenbach's work resulted in a thorough model that has had an influential role in the conceptualization of developmental psychopathology (e.g., through the Achenbach System of Empirically Based Assessment across the lifespan; Achenbach, 2009). Similar statistical approaches have also been used for devising, testing, and refining psychopathological constructs with the goal of increasing construct validity, as described below (Clark & Watson, 1995, 2019; Loevinger, 1957).

The quantitative empirical approach to psychopathology arguably has had relatively little influence on formal psychiatric classification systems until very recently. We are now seeing the *ICD-11* integrating dimensions in the classification of mental disorders to some degree (see Rief et al., in prep), and although not fully replacing personality disorder categories, the *DSM-5* incorporated a hybrid Alternative Model of Personality Disorders. Looking back, in adult psychopathology, quantitative empirical efforts remained relatively isolated and uncoordinated through the 20th century, but quickly began accumulating early in the 21st. Important influences included advances in latent variable modelling techniques, cheap computing power, and large publicly funded and available data sets. Seminal publications on the latent structure of a subset of common mental disorders (i.e., unipolar mood, anxiety, substance use, and antisocial and disruptive behavior; e.g., Krueger, 1999; Lahey et al., 2004) led the way and were followed by the inclusion of psychotic/thought disorders (e.g., Kotov et al., 2011; Wright et al., 2013) and the personality disorders (e.g., Kotov et al., 2011; Markon, 2010; Wright & Simms, 2015), providing the outlines of a truly broad structure of adult psychopathology. These examples, and their evolution in further

studies, led to the establishment of the HiTOP model (see Figure 1) and consortium (Kotov et al., 2017), which coordinates systematic efforts for using quantitative empirical approaches to establish a valid structure of psychopathology for the improvement of scientific and clinical use.

Despite their differences, both traditions share a descriptive focus and are based primarily on observable or reportable signs and symptoms, and thus are quite similar in the breadth of psychopathology targeted for description and their focus on phenotypic characterization. However, traditional psychiatric classification systems have a long track record of development and revision, whereas the quantitative empirical approach has only recently aspired to be a comprehensive model. These aspirations raise the question of what the continued development and refinement of a quantitative empirical structural model should prioritize in revision efforts. To address this gap, the Revisions Workgroup of the HiTOP Consortium has worked on defining priorities and principles for revising the model, with a focus on fulfilling the multiple purposes of psychiatric classification systems.

The Purpose of Psychopathology Classification Systems

Blashfield and Draguns (1976) summarized five principal purposes of psychopathology classification: 1) to facilitate information organization and retrieval, 2) to assist with communication, 3) to provide a descriptive system, 4) to provide a predictive system, and 5) to be used in and to facilitate scientific theory.[1] These purposes can also be abstracted to the two more general goals of clinical utility and scientific accuracy (Kendell & Jablensky, 2003), which have common ground but have often been framed in opposition to each other (e.g., in the argument that even if dimensional systems are more accurate, they are not ready to be adopted clinically because of a lack of utility; Haeffel et al., 2021). To the extent that a nosology serves a practical purpose, it must have clinical utility—broadly construed as being helpful in organising clinical assessment, selecting treatment options, and communicating the nature of the problem(s) to patients and others

[1] To these five, Keeley et al. (2014) added a sixth of "socio-political functions." These may include both beneficent (e.g., promoting underserved populations), malevolent (e.g., marginalization), or unrelated goals (e.g., legitimizing institutions and retaining power).

involved in their health care. It should also be reliable and relatively easy to use. At the same time, scientific accuracy is an important foundation for clinical utility; our position is that the diagnosis and treatment decisions should follow from the evidence on which the classification system is based. Further, scientific accuracy is essential for the role classification systems play in the formulation and funding of research questions and study designs (Hyman, 2010).

Optimally, a good classification system can serve as an effective guide to both science and clinical practice by facilitating the translation of research findings to patient care and providing a bridge through common language. Further, new research findings can be incorporated to inform the refinement of nosology as well as the treatment selection and clinical decision-making it informs. To illustrate, we take an example from oncology, in which the early classifications focused on location in the body (e.g., lung, liver, brain) but are now being refined with molecular genetics (Louis et al., 2021). Improvements in cancer treatments are expected to follow revisions to cancer classification that have incorporated recent advances in molecular-genetic profiling (Carbone, 2020): With a shift in organization based on new evidence, cancer diagnoses and treatments will better align with specific mechanisms of tumor formation and functioning, rather than tissue type and location within the body. The hope is that HiTOP's reorganization of psychopathology will have a similar effect on our understanding of the mechanisms of psychopathology and support improved clinical care.

It is also important to distinguish between pragmatic decisions borne out of the necessity of taking clinical action versus decisions based on scientific accuracy. These are often confused in psychiatry and clinical psychology, where the distinction between the need for practical cut-offs or clinical shorthand and the scientifically accurate representation of phenomena are often blurred. Much psychopathology research treats *DSM* or *ICD* categorical diagnoses (e.g., major depressive disorder, schizophrenia, personality pathology) as *the* phenomenon of interest as opposed to one plausible and, importantly, fallible candidate operationalization (Fried, 2022; Wright & Ringwald,

2022).[2] To the extent that these diagnoses are sufficiently scientifically accurate constructs, this is acceptable; if not, the result is a reification of potentially problematic diagnostic criteria and thresholds that can hamper or misdirect psychopathology research. Unfortunately, without a foundation of scientific accuracy the clinical utility of many diagnoses is also compromised: Diagnostic categories have substantial symptom overlap (Borsboom et al., 2011; Forbes et al., 2023; Tio et al., 2016) and natural categories seem to be rare or non-existent (Haslam et al., 2012; 2020), contributing to low inter-rater reliability for many diagnoses (Markon et al., 2011; Regier et al. 2013). Further, the polythetic approach to diagnosis (i.e., some subset of features, but not all or any necessary feature, must be present) leads to groupings of patients that have highly heterogeneous symptom profiles (e.g., Fried & Nesse, 2015), obscuring patients' specific symptom presentations and reducing the clinical utility of diagnostic labels. Correspondingly, practitioners often ignore the formal nosology in practice, suggesting they find it to have limited utility. For example, in a global survey of 1,764 mental health professionals (predominantly psychiatrists) 50% reported that they often or routinely make initial diagnosis without referring to DSM/ICD diagnostic criteria (First et al., 2018). The HiTOP effort is based in part on the belief that the level of scientific accuracy sets the limit on clinical utility—a construct lacking in validity will ultimately serve as a suboptimal guide to practitioners—so we prioritize scientific accuracy with the expectation that improvements in clinical utility will follow.

Defining and Constructing Validity in Psychiatric Classification

Establishing the scientific accuracy of psychiatric classification systems is essentially the task of establishing their validity. Cronbach and Meehl (1955) introduced the concept of construct validity, which establishes the nature of a construct based on its nomological network—the pattern of relations among elements of the construct (e.g., symptoms assigned to the syndrome) and its

[2] This need not be the case, as the commonly used example of blood pressure illustrates: individual differences in blood pressure are understood to fall along a continuum, but diagnostic guidelines have been developed for treating high blood pressure that distinguish between basic construct definitions (e.g., blood pressure is a dimensional construct) and the clinically necessary threshold for treatment (e.g., treatment for high blood pressure is indicated by the categorical threshold of blood pressure >130/90).

links to other constructs (e.g., other syndromes, future outcomes, treatment response). However, the accumulating evidence on covariation among features of psychopathology suggests its scientific classification must go beyond an approach that establishes separate nomological networks for putatively independent constructs (Kotov et al., 2017). Our perspective is that a comprehensive classification system is integrative and includes the ensemble of how narrow constructs (e.g., individual features) associate to form higher order constructs, how they themselves associate to form broader ones (i.e., a hierarchy), and how all are distributed among individuals (e.g., dimensionally vs. categorically) and relate to other relevant and differential constructs. In other words, the scope of construct validity in the context of psychopathology classification is not the narrow validation of a single test, but the validation of a broader system.

Returning to Blashfield and Draguns' (1976) theory of psychiatric classification, the purposes of serving as both a descriptive and predictive system stand out as central considerations for construct validity. An effective descriptive system requires structural validity—including in how constructs are formed, how they relate to each other, and how they are distributed. By contrast, a predictive system requires not only structural validity, but also evidence of convergent, discriminant, concurrent, aetiological, and prospective (i.e., prognostic) validity through the patterns of association with constructs external to the descriptive system. The development and continued revision of formal psychiatric classification systems, such as the *ICD* and *DSM*, have mostly taken structural validity for granted and focused on other forms of validity—in large part due to various social and political pressures operating within and outside the profession of psychiatry (Blashfield et al., 2014). In the modern era (i.e., since *DSM-III*), prioritizing other forms of validity over structural validity has manifested as a concern with validating established diagnoses. In particular, the focus has been on validating diagnostic categories against external criteria first described by Robins and Guzé (1970), such as laboratory markers, disease trajectories over time, and family studies.

Historically, seeking to validate diagnostic categories in this way has been a major focus of psychiatric research (e.g., Kendler, 1980; 2013; Kendell, 1989; Andreasen, 1995). As awareness of the frequent co-occurrence of distinct diagnoses grew, so did recognition of the costs associated with not adequately establishing the structural validity of diagnostic constructs (Kendell & Jablensky, 2003; Zachar, 2015). In the commitment to seek out validity for psychopathology and legitimacy for the field, one of the most important questions was skipped over: Are the foundational constructs structurally valid? Specifically, are most psychiatric diagnostic categories discrete constructs with natural boundaries? Ultimately the evidence strongly suggests that the majority are not (for a review, see Haslam et al., 2020), but this realization has not yet had a major effect on the established approach to revising and updating the *DSM* despite comprehensive consideration of other sources of validity evidence. For example, in the revision process leading to *DSM-5*, the Robins and Guzé (1970) criteria were considerably revised and expanded to include a comprehensive set of ten indicators in three temporal groups—antecedent, concurrent, and predictive: familial aggregation and/or co-aggregation; socio-demographic and cultural factors; environmental risk factors; prior psychiatric history; cognitive, emotional, temperament, and personality correlates; biological markers; patterns of comorbidity; diagnostic stability; course of illness; and response to treatment (Andrews et al., 2009; Kendler 2013).

By contrast, the process for the initial construction and inclusion of diagnoses is unspecified in formal psychiatric classification systems and therefore unsystematic and quite heterogeneous. Accordingly, diagnoses have accumulated from a variety of sources. Some diagnoses were developed by observing individual patients (e.g., Alois Alzheimer describing illness in his patient Auguste Deter, which became known as Alzheimer's disease; Hippius & Neundörfer, 2003). Others were proposed based on clinical observation of a case series (e.g., Emil Kraepelin sorted patients into dementia praecox or manic-depressive groups to maximize the prognostic and pedagogic value of these diagnoses; Jablensky, 2007)—with this approach even resulting in duplicate constructs (e.g., Asperger's syndrome and Autistic Disorder; Baron-Cohen, 2015). Yet other diagnoses were

designed to capture a group already identified by other disciplines (e.g., attention deficit/hyperactivity disorder evolved from efforts in education research to describe students without cognitive impairment who nonetheless struggled with academic performance at school; Mayes & Rafalovich, 2007). A variety of other considerations have also motivated construction of new diagnoses (e.g., constructs important in psychoanalytic theory or lobbying by consumer groups; Blashfield et al., 2014). Although the *DSM* has certainly given some consideration to structural validity—typically in the narrow forms of the addition, deletion, or sub-typing of diagnostic categories—a wholesale consideration of the structural validity of the system, from how signs and symptoms reflect individual diagnoses and beyond, has not been undertaken despite the noted problems with the current system. Without close consideration of the structural validity of the constructs that organize the diagnostic categories, the focus in psychiatry on other forms of validity (e.g., etiology, prognosis, treatment response) of diagnoses has produced mixed results as evidenced, for example, by limited specificity in associations with external criteria and treatment response.

In contrast to the *DSM*, the quantitative empirical approach to psychopathology—and by extension the HiTOP consortium—has expressly prioritized structural validity as a foundation, but not at the exclusion of other forms of validity. Rather, the stated assumption is that structural validity will support success in establishing predictive utility by providing more coherent and reliable targets of inquiry. Of course, both structural and other forms of validity need to be considered in the development and revision of any psychiatric classification system that aims to fulfil the roles of clinical utility and scientific accuracy. The way that different aspects of validity are conceptualized, constructed, and prioritized in a classification system go on to inform the approaches to revision.

Approaches to Revision in Existing Models

With the goal of establishing HiTOP as an influential model of psychopathology comes the need for establishing processes and procedures to revise the model. In developing our own

processes, we first sought to learn how others have approached similar tasks, and to draw on the strengths of those approaches. For example, the process for revising the *DSM-5* is based on literature reviews to describe the evidence for new diagnoses or changes to existing diagnostic criteria to improve validity, reliability, utility, or to reduce deleterious consequences. Proposals for *DSM* revisions are to be organized around the set of validating criteria described above and must include summary tables of each criterion for which data exist; each row in these tables represents a study contributing evidence with columns summarizing the sample size, methods, and results along with a qualitative judgement of the overall methodological strength (rated 1-5) of each study. An additional optional table is recommended for proposals, rating the degree to which data from each criterion support the proposed change (rated 1-5). The HiTOP revision protocol draws on these features of the *DSM* process, though departing in noteworthy ways as well—for example, validity is not prioritised over improvements in reliability in HiTOP, as it is in the *DSM* revision process (APA, 2021; see criteria for Type 1A and Type 1B proposals). Further, the process for approving revisions in HiTOP deliberately minimises the role of social and political forces that influenced the *DSM-5* (Pilecki, Clegg, & McKay, 2011).

We also drew from the process for updating clinical practice guidelines, which has moved towards a comprehensive, systematic, and standardized approach in the last decade (Guyatt et al., 2011; Steinberg et al., 2011). The first step in this process is a systematic literature review, which is distinguished from a traditional narrative review by specifying the review protocol in advance and sometimes also involving a meta-analysis for statistical integration of findings. Guidelines for constructing these protocols have been established (Moher et al., 2009; Page et al., 2021). Next, evidence gathered in the reviews is rated to determine confidence in conclusions. Then, findings of reviews pertaining to different considerations involved—namely the benefits, harms, and costs—are integrated to produce a recommendation for clinical practice, along with a rating for the confidence in this recommendation. The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system was proposed as the process for rating systematic reviews and

resulting recommendations (Balshem et al., 2011; Guyatt et al., 2011). GRADE has become the standard in the development of practice guidelines and considers both the internal validity of individual studies and the overall quality of the body of evidence. In this context, internal validity refers to the degree to which the study was designed and conducted in such a way as to answer the question at hand adequately. Although developed in the context of practice guidelines, several core characteristics of the GRADE system—in particular the comprehensive, systematic, and standardised approach to rating the strength of evidence and level of confidence in a recommendation—are readily adapted to the process of revising a classification system for psychopathology, as described below.

In short, the revisions processes for the *DSM-5* and for updating clinical practice guidelines with GRADE were both influential in the design of the HiTOP revisions protocol, which adds a strong focus on structural validity. Whereas GRADE provides the template for developing procedures to conduct revisions hewing closely to the strength of the empirical evidence for or against a change, the *DSM-5*'s emphasis on the validity of each construct is important for maintaining consistency and relevance for psychiatric audiences that may ultimately choose to use the HiTOP framework. We turn now to introduce the HiTOP revisions protocol and its organizing principles.

Revising the HiTOP Framework

The goal of this revisions protocol is to ensure that the HiTOP framework reflects the state of current scientific knowledge on the structure of psychopathology—growing and changing as new evidence emerges—and to ensure that the extensions and modifications to the model are empirically based. The revisions protocol is thus designed as a formalized system to evaluate and incorporate new evidence into the model with three core organizing principles: 1) systematic evaluation of quantitative evidence by a set of transparent criteria and processes, 2) balancing stability with flexibility, and 3) promoting inclusion over gatekeeping, as is described below along with the known assumptions of the revision protocol.

The strategy for revising the HiTOP framework closely follows the standard analytic sequence for evaluation of nomological networks described above; that is, first establishing structural validity to ensure the foundation of a well-defined and reliable construct, then evaluating other forms of validity (e.g., convergent, discriminant, concurrent, aetiological, and prospective validity). Specifically, the strategy is first to investigate co-variation among individual signs, symptoms, traits, and maladaptive behaviors—or broader constructs at higher levels of the hierarchy—to identify coherent constructs that are distinct to varying degrees from other constructs. Relations among the resulting constructs are used to group narrow constructs into broader higher order dimensions (e.g., subfactors, spectra, or superspectra in Figure 1). Coherent constructs that emerge through the initial focus on reliable description of phenotypic manifestations of psychopathology (e.g., at the level of self- and other-reported symptoms, traits, and signs) are the focus in the initial development of the HiTOP framework; then we proceed to determine whether these hypothesized constructs are strengthened or challenged when evaluating other forms of validity. Indeed, structural analyses could parse psychopathology too finely in some cases, making distinctions that do not predict relevant external variables differently and that therefore are not informative for understanding and treating mental disorders. Also, existing structural data may have methodological confounds that distort the structure (e.g., two constructs loading together on a factor because of how they were assessed; see Podsakoff et al., 2003). External validation of the constructs may reveal such shortcomings and reorient structural research.

Organizing principles

The first organizing principle of both the construction and revision of the HiTOP structure is to follow quantitative evidence in developing the system, with a specific aim of curtailing decision-making based on special interests, tradition, or politics (Krueger et al., 2018). Naturally, evidence for or against a proposed revision to the framework will be a matter of degree, and evaluating this

evidence necessarily involves value judgements by the reviewers and committees described below.[3] However, our workgroup seeks to guard against biased decision-making to the extent possible by outlining criteria for reviewers and committees to evaluate evidence systematically, and establishing a transparent process by which decisions are made based on this evidence. The intended outcome is to minimize bias and maximize reliability in decision-making to enable different raters to arrive independently at the same conclusions based on the same evidence.

A second organizing principle reflects the need to balance stability with flexibility: The HiTOP framework must be stable enough to facilitate research and clinical application. For example, it is important to avoid the need for continual updating of assessment batteries by researchers and clinicians. At the same time, the framework needs to be flexible enough to incorporate emerging evidence expeditiously to avoid becoming a hindrance to scientific and clinical advances. To achieve this balance, the plan is to revise the HiTOP framework iteratively to reflect provisional or confirmed changes as they are approved. HiTOP measures are likely to inform and be informed by these changes to the model, and adjustments to both the model and measures will follow as indicated. Applying consistent standards of evidence to proposed changes will likewise balance the need for rigor with the ability to incorporate new constructs. Although initial revisions may be substantial, as discussed below, it is likely that many smaller iterative changes will also be indicated as components of the model are systematically evaluated.

The third organizing principle of the HiTOP revisions protocol includes promoting inclusion rather than gatekeeping, and this is embedded throughout the process as discussed further below. To further guard against bias embedded in expert evaluation and consensus, the revisions process also has rotating roles of authority (e.g., proposal coordinator, review panel members).

[3]Of note, the HiTOP Consortium overrepresents people from privileged groups with dominant voices in psychopathology research (e.g., non-Hispanic white, cisgender, non-disabled men working at research-intensive universities in the U.S.; see Rodriguez-Seijas et al., 2023). Homogeneity in reviewers' and committees' viewpoints is another potential source of bias in evaluating revisions proposals, so continuing to increase the diversity of expertise and perspectives in Consortium members is essential.

Known assumptions in the revisions protocol

The HiTOP framework's descriptive, data-driven classification system of psychopathology can create the erroneous impression of a purely atheoretical and objective system. It is important to highlight that the HiTOP framework and revisions protocol both carry assumptions that impose constraints on the resulting model. For example, the HiTOP framework assumes that psychopathology is best understood via a hierarchical structure (i.e., it can be described at various levels of generality/specificity that are nested within each other). The hierarchical nature of the HiTOP framework is a valuable heuristic for clinical practice and research—allowing the conceptual mapping of psychopathology at different levels of specificity or abstraction—but we cannot determine from model fit to data whether the data are truly hierarchical because there are many statistically equivalent or near-equivalent models that can be fit to the same covariance matrix (Greene et al., 2019; Markon, 2019; Mulaik & Quartetti, 1997; Yung et al., 1999). By contrast to these untestable assumptions, the dimensional nature of HiTOP constructs is amenable to direct empirical testing and reflects current research-based evidence that individual differences in nearly all domains of psychopathology are better represented as dimensions than as categories (e.g., Haslam et al., 2020); the HiTOP framework and revisions process can incorporate categorical constructs if the data so indicate.

The revisions protocol described below also weights evidence from latent variable models more heavily than alternative statistical approaches, assuming that latent variable models provide inherently stronger evidence than other approaches. The decision to do this was because the body of evidence underpinning the HiTOP framework is dominated by factor-analytic studies and is consistent with the dominant dimensional conceptualization of psychopathology constructs throughout the current framework. However, a reliance on factor analysis shapes the resulting framework—for example, maintaining the dominance of dimensions as the organising constructs in the framework, allowing constructs to cross-load across multiple levels of the hierarchical structure (see Clark & Watson, 2019), partitioning shared vs. unique variance among constructs in different

ways at each level of the hierarchy (cf. Van Bork et al., 2017), and largely relying on the idea that indicators may combine in a linear fashion rather than an interactive or causal fashion (cf. Borsboom, 2017); thus, limitations of factor analysis could beget limitations in the resulting model.

Notably, although the core features of the HiTOP framework carry known assumptions, even these assumptions are subject to revision based on emerging methodologies and data. Moreover, there may be "deep" assumptions that are so thoroughly embedded in our ways of conceptualizing psychopathology that we are unaware of them. Tests of the HiTOP framework parameterized as a statistical model (e.g., Ringwald et al., 2022)—rather than the schematic representation of the literature reviewed in Kotov et al. (2017)—and in the context of the full array of human variation may bring some of these assumptions to light and require fundamental revisions of the framework.

Protocol for Revising the HiTOP Framework

We move now to discuss the protocol itself. We first provide some brief background, then explain the scientific side of revisions—which entails two rubrics to score the strength of evidence from each study relevant to a proposed revision—and the administrative side of revisions (i.e., the process that each proposal must go through, from inception to evaluation and final recommendation). To aid readers in understanding the nature and roles of the individuals and groups mentioned throughout this section, we have included additional information in Table 1.

Table 1. *Explanations of the individuals and groups described in-text with a role in the revisions process.*

Individual or Group	Role
HiTOP Consortium	A formal association of approximately 200 academic researchers and clinicians working to advance the classification of psychopathology beyond traditional diagnostic systems to aid clinical practice and mental health research. Current members are listed at https://www.hitop-system.org/consortium-members
HiTOP Clinical Network	An informal association of clinicians and clinician-researchers dedicated to learning and disseminating information about the challenges and opportunities of using a model like HiTOP in practice (see https://www.hitop-system.org/the-clinical-network for more information)
HiTOP Trainee Listserv	A listserv where trainees interested in HiTOP can register to receive updates (see https://www.hitop-system.org/trainees for more information)
Revisions Workgroup	A workgroup within the HiTOP Consortium with a charter to (1) devise best practices for validating and revising the HiTOP structure; and (2) use these practices to guide ongoing revision of the working HiTOP model, as needed, as well as to collate topics for clarification and future research. A list of the members can be found at https://www.hitop-system.org/revisions-workgroup. MKF and AGCW (first and last authors, mentioned in-text) are the Chairs of the workgroup.
Executive Committee	The committee that oversees the HiTOP Consortium, comprising the three founders of the consortium (Roman Kotov, Robert F. Krueger, and David Watson) and the (co-)chairs of the ten workgroups. Current members (n = 20) are listed at https://www.hitop-system.org/consortium-members
Proposal Coordinator	A rotating position in the Revisions Workgroup, who is the nominated contact person for all proposals for an agreed period (e.g., 3-6 months) to facilitate the (optional) de-identification of the review process. The current Proposal Coordinator is listed at https://www.hitop-system.org/revisions-workgroup
Review Panel	A group of no less than three members of the HiTOP Consortium, Clinical Network, and/or Trainee Listserv (optimally, 5-7 volunteers). Members of past Review Panels can be seen at https://bit.ly/HiTOPRevisionsOutcomes
Proposer(s)	The individual or team submitting the proposal for change

Note. The organizational structure of HiTOP is evolving. Any changes in the organisational structure that affect the process described here (e.g., the formalization of student membership in the Consortium) will be documented on our OSF page (https://osf.io/8h7m6/).

The HiTOP revisions protocol was developed through an iterative process over 4 years. Mini workgroups, composed of members of the Revisions Workgroup, piloted each iteration of the proposed revisions processes. With each test, we found new challenges and weaknesses, applying what we learned to the next iteration. There were three main iterations with complete revision of the process proposed at each step: 1) implementing meta-analytic SEM with study-level moderators (cf. Jak & Cheung, 2020), 2) quantifying the strength of evidence *for each component* in the framework, and 3) quantifying the strength of evidence *from each study* relevant to the revision

being proposed. The current approved version of the revisions protocol is v3.20 (see https://osf.io/2g3sr), which incorporates multiple stages of input and feedback from members of the Revisions Workgroup, the HiTOP Executive Committee, and the HiTOP Consortium. MKF (first author) led the process of developing the revisions protocol together with AGCW (last author). All Revisions Workgroup and Executive Committee members approved the current version of the revisions protocol. In the future, any changes to the protocol will be documented on the OSF page (https://osf.io/8h7m6/), and the current version of the protocol will be updated and maintained there also.

The core features of the HiTOP revisions framework align with its organizing principles: Two scoring rubrics are used to evaluate systematically the strength of the quantitative evidence provided by each study submitted as part of each revision proposal (see below). Rating the degree of evidence for the revision permits a graded approach to revising the HiTOP framework that balances robustness and flexibility, and this process results in either confirmed, provisional, or no changes. Confirmed changes to the structure of the model currently require a systematic review to ensure robustness, whereas provisional changes require only a narrative review and sufficient supporting evidence—providing flexibility to incorporate emerging or changing bodies of evidence quickly. The standardized rating framework, as adapted from the GRADE criteria (Balshem et al., 2011; Guyatt et al., 2011) described above, provides well-defined criteria designed to ensure transparency and reliability of ratings within and between proposals. Finally, towards the goal of promoting inclusion rather than gatekeeping, the HiTOP revisions approach places minimal limits on what data can be used to support a proposal or on who can submit a proposal. Analyses of unpublished data and re-analysis of published data are encouraged, which allows groups to contribute even if they have not collected their own samples or published with their data. There are no restrictions on who can submit a proposal; at the same time, including one or more members of the HiTOP Consortium may aid in familiarity with the framework and revisions process. To reduce barriers to access, the initial letter of intent can include an open invitation to Consortium members

to join the proposal-writing team.[4] Echoing conventions from the scientific review and publishing process, revision proposals can also be masked prior to review to protect against bias based on characteristics of the research group (e.g., career stage, institutional affiliation, nationality, race, gender) instead of research quality, although strict masking of proposal authors' identities may not always be possible.

Each version of the HiTOP framework will be labelled with the date of revision. We are committed to making the most current version of the HiTOP framework publicly and freely accessible, alongside an archive of older versions. Currently, the plan to achieve this is to use the Open Science Framework (https://osf.io/gds3n) to document the evolution of the framework over time, together with strategies to disseminate news of revisions to the framework (e.g., updating the figures displayed on HiTOP webpages, and announcing the changes to the Consortium and on social media).

The Scientific Side of Revisions

Consistent with the focus of HiTOP on providing data-driven description of the structure of psychopathology, the core of the revisions protocol is a rubric designed to score the strength of the structural validity evidence conferred by each study relevant to a proposed revision (see Figure 2A). The aim of the scoring system, as mentioned above, is to weight and summarize the strength of the evidence provided by a study based on its relevance to the revision being proposed, its methodological approach, and its results.

[4]New Consortium members are welcome and encouraged (see https://www.hitop-system.org/get-involved-consortium).

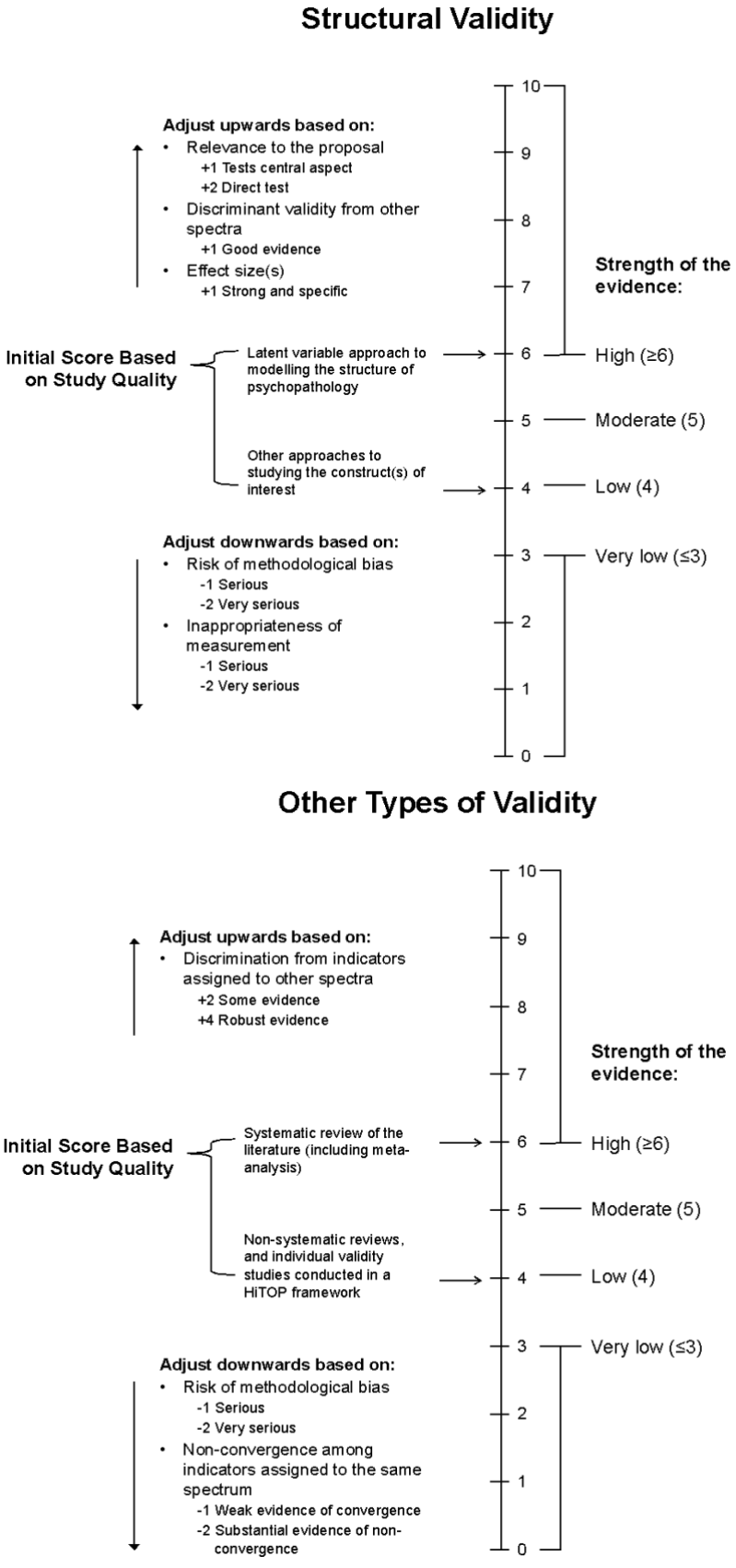

Figure 2. An overview of the approach to scoring the strength of evidence from each study evaluated. The initial score based on study quality is adjusted upwards and/or downwards based on study characteristics (adding or subtracting points), and the final score corresponds to an overall rating of the strength of evidence provided by the study for the revision proposal being assessed.

A study score is initially assigned based on study quality: Studies that adopt a latent-variable modelling approach (e.g., exploratory or confirmatory factor analysis) earning an initial grade of 6 (i.e., higher quality), whereas those using alternative designs (e.g., reporting the rate of overlap or correlations between constructs) receive an initial grade of 4 (i.e., lower quality). Scores are subsequently adjusted downward based on (i) risk of methodological biases, including any aspects of the study design that could distort or bias its conclusions (e.g., logical alternative models that were not tested, evidence of overfitting, small sample size); and (ii) inappropriate measurement, including the use of measures with poor reliability or inadequate breadth of measurement to capture the construct(s) of interest, which may include using categorical diagnoses rather than dimensional measures. Likewise, scores can be adjusted upward based on (iii) relevance to the proposal, including results that speak directly to the core question of a proposal (e.g., by measuring all relevant constructs in the proposal dimensionally or directly adjudicating between alternative model specifications); (iv) discriminant structural validity, reflecting the inclusion of sufficiently diverse dimensional markers of psychopathology to permit tests of convergence and divergence among multiple dimensions; and (v) effect size, including the strength, consistency, and precision of relevant statistical parameters. In the up- and down-weighting of studies' evidence, most criteria correspond to a 1-point adjustment; three fundamental criteria—risk of methodological bias, inadequacy of measurement, and relevance to the proposal—allow for up to a 2-point adjustment to give greater weight to particularly strong evidence in these domains. Scores are summed across all criteria to yield a final study score (see Figure 2A).

A similar rubric is used to weight the strength of other types of validity evidence conferred by each study relevant to a proposed revision (see Figure 2B). Because systematic data on patterns of relations with constructs beyond those in the HiTOP framework is not always available, evaluation of validity beyond structural validity is an optional but strongly encouraged step.[5] The

[5] Optional only if adding a construct to the model or moving a construct that has not been validated. If the position of a construct was supported by validity evidence in Kotov et al. (2017), the subsequent reviews of validity evidence for

literature is rated separately for each domain of evidence being reviewed in a proposal, drawing on the same criterion set used in developing *DSM-5*. These include: the magnitude of genetic and environmental influences from behavior genetic studies, molecular-genetic risks, specific environmental risks, cognitive-and-emotional-processing abnormalities, neural substrates, biomarkers, childhood-temperament antecedents, trajectory/illness course, and treatment response (Andrews et al., 2009). Other types of evidence can also be included with a rationale for their relevance to the revision proposal at hand.

As with structural validity evidence, a study score is initially assigned based on study quality, with systematic and meta-analytic reviews earning an initial grade of 6 (higher quality) and other studies (i.e., non-systematic reviews and individual validity studies conducted in a HiTOP framework) receiving an initial grade of 4 (lower quality). Although systematic and meta-analytic reviews have their own limitations, their breadth affords them an advantage over selective reviews and individual studies (Corker, 2020). Such high-quality data are becoming increasingly common, including genetic correlations among psychopathology constructs based on meta-analyses of single nucleotide polymorphisms, and correlations among neurobiological profiles of disorders based on meta-analyses of neuroimaging data (Kochunov et al., 2022; Opel et al., 2020; Waldman et al., 2020). However, particularly when systematic reviews are unavailable, it will also be necessary to consider non-systematic reviews or to aggregate findings of individual studies (e.g., Kotov et al., 2020; Krueger et al., 2021; Watson et al., 2022).

Scores are subsequently adjusted downward based on evidence of methodological weaknesses in the study being rated (e.g., *risk of method bias*, such as evidence of substantial publication bias, selective reporting, use of categorical diagnoses rather than dimensions, or a poorly designed validation study), or if the results do not show patterns of convergent validity in line with the proposal (e.g., *non-convergence*, such as inconsistent patterns of association for

---

HiTOP constructs (i.e., Kotov et al., 2020; Krueger et al., 2021; Watson et al., 2022), or in revisions to the model (https://osf.io/8h7m6/), then validity of the new position needs to be documented.

constructs within a single proposed dimension). Scores are also adjusted upward if there is evidence for *discriminant validity* in line with the proposal (e.g., patterns of association for constructs within a single proposed dimension are different from constructs in other locations of HiTOP). The scoring adjustments for risk of method bias and non-convergence can be one or two points, depending on the strength and consistency of the evidence, but the points for discriminant validity are doubled (either two or four points, depending on the strength and consistency of the evidence). This is because of rampant non-specificity evident in patterns of external validity in the literature to date (e.g., Conway et al., 2019), which makes discrimination among constructs' nomological networks particularly noteworthy validity evidence.

Notably, this process will result in separate scores for each domain of evidence related to the proposed revision (e.g., based on the strength of the behavior genetic vs. molecular genetic vs. biomarker evidence). These scores may well diverge, and there is currently no plan to weight some domains over others. The final rating of the strength of evidence for other types of validity (i.e., beyond structural validity) supporting the proposal is based on the robustness and convergence of results across studies and domains. Revisions that are indicated by structural research but clearly reduce the degree of other types of validity will not be incorporated; rather, the disconnect between these domains of evidence will be earmarked as high priority for further investigation.

The Administrative Side of Revisions

Prior to submitting a full proposal, researchers who plan to propose a change to the HiTOP framework (i.e., "proposers") submit a letter of intent to the active Proposal Coordinator (see Table 1). The letter of intent is a brief description of the proposed change and review strategy to be sent to all HiTOP Consortium members (see Table 1) for early feedback and suggestions for relevant research to include in the proposal. This step is intended to help "crowdsource" evidence to maximise the chance that all key studies are included from the outset, minimizing the need for revisions after the proposal is drafted, as well as to ensure that the proposers are working from the most current version of the HiTOP framework and revisions protocol.

Based on this feedback, the intended proposal may be withdrawn or a full proposal can be submitted in a standardized format (described at https://osf.io/2g3sr) to the Proposal Coordinator. The full proposal is sent to the corresponding authors of all studies cited in support of the proposal for a 4-week comment period to seek input on agreement/disagreement with the studies' evaluations, or on the proposal overall. This step aims to include perspectives from researchers outside of HiTOP and avoid creating an echo chamber. At the same time this feedback is sought, the proposal is sent to a Review Panel of no less than three members of the HiTOP Consortium, HiTOP Clinical Network, and/or HiTOP Trainee Listserv (optimally, 5-7 volunteers). If more than five people volunteer to participate on the Review Panel, panel members will be selected with a view to maximise diversity of perspectives and backgrounds represented on the panel. Each reviewer makes a recommendation (confirmed change, provisional change, or no change) based on the information in the proposal, noting if/where/why they disagree with ratings or inferences in the proposal, and sends their detailed review and recommendation to the Proposal Coordinator. The de-identified feedback from Consortium members on the letter of intent and from corresponding authors of studies in the proposal is then shared with the full Review Panel—along with the other reviewers' comments—and the Panel meets to discuss the revision proposal and feedback. Reviewers may make changes to their comments and/or scores at this stage. Next, the Review Panel makes a recommendation based on their scores (confirmed change with ≥ 75% supporting this recommendation, provisional change with ≥ 51% supporting a provisional or confirmed change, or no change if ≥ 50% of reviewers recommend no change). A summary of the de-identified reviewer ratings and reviews are sent with the panel recommendation to the proposers. If the recommendation differs from the proposal, proposers have 4 weeks to submit a written appeal to the Proposal Coordinator providing additional information or further explication of previously submitted information that they believe would change the end recommendation. If an appeal is submitted, the Review Panel will discuss the appeal and affirm or revise their recommendation.

The final approval process is designed as an iterative safety net to protect against any problems that arise in the review process, and against special interests resulting in the veto of a revision. The Proposal Coordinator sends the Review Panel's comments and recommendation together with the de-identified feedback from Consortium members to the Revisions Workgroup and Executive Committee (see Table 1), who have a 4-week comment period to raise questions or concerns to be addressed by the Review Panel. The Review Panel's responses to any issues raised is sent to the Executive Committee (currently $n = 20$) where final approval of the decision requires a simple majority vote (currently $n \geq 11$) via an online survey. If a majority vote *not* to approve the Review Panel's recommendation, the Executive Committee must provide a detailed rationale and specify a concrete solution(s) to their concerns with specific reference to the criteria outlined above for revisions. At this point, the proposal team may also be contacted by the Proposal Coordinator with a specific request for further details or evidence, if required. The back-and-forth between the Review Panel and Executive Committee will continue until consensus is reached, with the explicit focus being on the quality of the empirical evidence underlying the proposal.

When the HiTOP framework is formally revised, the final drafts of all relevant proposals together with their outcomes will be shared with Consortium members in the monthly email update, and with the scientific community at large (e.g., on social media and via the OSF page of the current model https://osf.io/8h7m6/). Proposers are also encouraged to publish the proposal in a peer-reviewed outlet, and any proposals that do not result in a revision are recommended for further research.

Priority Areas in Revising the HiTOP Framework

Although considerable progress has been made in understanding the empirical classification of common and uncommon forms of psychopathology, every level of the HiTOP framework (Figure 1) will likely require revisions to reflect new evidence since the publication of Kotov et al. (2017). For example, we expect the lower levels of the framework could be substantially revised when the HiTOP measure development project is complete (Simms et al., 2022). Given the dearth

of literature on this level—with very little evidence on homogeneous symptom components/maladaptive traits and empirical syndromes, particularly from studies specifically designed for this purpose—the data from the HiTOP measure will be pivotal in this regard. Whether the very detailed levels of the structure are too complex to identify a robust and reliable set of phenotypes is an important empirical question for future research.

Moving up the hierarchy, further investigation of the provisional somatoform spectrum is also required to clarify whether this dimension warrants inclusion among the core spectra in HiTOP; current evidence is mixed as to whether somatoform is nested within a broad internalizing dimension or forms a distinct higher-order spectrum (e.g., Forbes et al., 2017, 2021; Kotov et al., 2011b; Krueger et al., 2003; Markon, 2010; Sellbom et al., 2021; Simms et al., 2012). More research is needed that includes sufficient indicators of the somatoform dimension, broad coverage of other domains of psychopathology, and that tests a wide variety of potential models on both structural and external validity criteria.

At the highest levels of the hierarchy, two new *superspectra* were recently proposed—Emotional Dysfunction and Psychosis—alongside the Externalizing superspectrum (Kotov et al., 2020; Kotov et al., 2021; Krueger et al., 2021; Watson et al., 2022), and these are candidates for inclusion in an early revision of the HiTOP framework. These broad dimensions have evidence for both structural and external validity and could help to elucidate the upper levels of the framework.

Finally, although traditional diagnoses are not formal parts of the framework, most of the studies synthesized in Kotov et al. (2017) were anchored to *DSM* constructs. From this perspective, it is noteworthy that whole chapters of the *DSM* are not yet integrated into the HiTOP framework (e.g., paraphilic disorders, elimination disorders). Extending the breadth of the framework is a consortium priority. For example, some features of autism, ADHD, and other features traditionally described within the *DSM* Neurodevelopmental Disorders chapter—such as social communication and learning disorders—may form a distinct spectrum (Michelini et al., 2021).

Other needed revisions include clarifying the placement of symptoms used to diagnose *DSM-5* mania, which remain provisional aspects of the framework: Symptom-level analyses of mania criteria indicate that most dimensions align with thought disorder (Kotov et al., 2020), but the heterogeneity of mania criteria may require some symptoms to fall under other spectra (e.g., internalizing, externalizing, or as a unique mania-symptoms spectrum; Carpenter et al., 2009; Forbes et al., 2021; Stanton et al., 2019; Watson & Naragon-Gainey, 2014). Similarly, the placements of symptoms that are used to diagnose obsessive-compulsive and related disorders (OCRDs) and eating pathology have been a focus in several recent studies with implications for revisions and additions to the HiTOP framework (e.g., Cooper et al., 2021; Dunkley et al., 2020; Faure & Forbes, 2021; Marshall et al., 2020; Rossell et al., 2020). As research has begun to address disorder-level heterogeneity, it has become increasingly clear that symptom components from within one diagnosis can load on different HiTOP dimensions (e.g., negative schizophrenia symptoms loading on detachment, rather than thought disorder; Cicero et al., 2019; Kotov et al., 2022), so more fine-grained approaches to analysis may help to advance research in this area. A fundamental HiTOP aim is to move beyond traditional diagnostic categories to establish the empirical structure that emerges from quantitative analyses of comprehensive symptom-level and trait-level measurement of psychopathology.

At the time of writing, the first official revision to the framework has been made following formal approval of a proposal to change the name of the *substance abuse* construct to *harmful substance use* (see https://osf.io/8h7m6/ for the documentation on this proposal). Another proposal on the potential placement of paraphilias in the framework was submitted with a recommendation by the proposers to make no change, based on insufficient evidence following a systematic review. (Such “Investigator opt-out [no change]” proposals are publicly documented, with explanation, in a Google Sheet here: https://bit.ly/HiTOPRevisionsOutcomes.) Finally, a third letter of intent was submitted, but based on feedback from the broader consortium the authors chose not to proceed with their review and proposal.

Open Challenges

The original HiTOP framework (Kotov et al., 2017) operated on the premise that a single model could be appropriate for all people and contexts. However, this premise is likely false as the structure of psychopathology is not necessarily universal, in which case multiple models of psychopathology would be needed. This challenge is empirically tractable and will be tested as the framework is extended into understudied populations by the Developmental and Diversity, Equity, and Inclusion Workgroups. Much of the data leading to the original HiTOP framework came from relatively homogenous white and Western samples (with some exceptions; e.g., de Jonge et al., 2018; Ivanova et al., 2007, 2015, 2019; Krueger et al., 2003), so it is essential to test the measurement invariance of latent-variable models representing HiTOP constructs in samples of individuals from underrepresented groups, life-span samples, and with regional and international diversity. To avoid prioritizing the current HiTOP structure as a culturally universal norm from which underrepresented groups deviate, the best fitting structural model should be identified in each group before moving to testing for (in)variance between groups. To the extent that any non-invariance extends to configural non-invariance (i.e., indicators for constructs vary by group), we will need to do further research to understand why this might be the case (e.g., cultural or linguistic reasons) and may need to have multiple parallel frameworks. However, some research has begun to examine the measurement invariance of the hierarchical structure of psychopathology across various identities, including those defined by race, ethnicity, sexual orientation, and so on, finding promising results and highlighting the various considerations that make such investigations critical for mental health classification and disparities research (see Rodriguez-Seijas et al., 2023, for a review).

A related challenge will be the emphasis in the current revisions protocol described above on the psychometric properties of scales, sample sizes, and replications, which could be biased against phenomena such as cultural idioms of distress or structural determinants of health, and inversely associated with sample diversity or research in understudied populations. Interview and

self-report measures exist for some idioms of distress, but they have not been used as extensively as measures used in Western, educated, industrialized, rich and democratic samples (WEIRD samples; Kaiser et al., 2015). Thus, there currently are fewer psychometric data available to establish the reliability and validity of their scores. Similarly, existing samples with marginalized and/or historically underrepresented populations may be smaller and replications less likely to be conducted. One solution may be to weigh the results of these studies more heavily in evaluating the strength of evidence they provide for a revisions proposal (e.g., add points if the studies are conducted in diverse samples or understudied populations, as is currently done for other aspects of study quality), but this approach may have the downside of overweighting less precise parameter estimates, which could lead to incorrect interpretations. Thus, an effort to be inclusive could lead to harmful overgeneralizations about understudied groups based on too little data. The revisions protocol is in theory well positioned to incorporate previously unarticulated constructs of psychopathology that have particular salience for marginalised and/or historically underrepresented populations and to create a more inclusive and culturally informed model of psychopathology; however, we are still working on how best to achieve this in practice.

It is challenging to resolve these complex issues, so changes to both the HiTOP structure and the processes by which we revise the framework remain on the agenda for the Revisions Workgroup moving forward. In the meantime, it should be a priority to study large samples of understudied populations after determining that the focal measures have strong psychometric properties in the group of interest. This can be done through multiple methods, including both small- and large-scale collaborations (e.g., the Psychological Science Accelerator; Moshontz et al., 2018), particularly with researchers with cultural expertise, at less resourced institutions, and/or in understudied countries and languages, which will move the HiTOP enterprise into the complex realm of translations (e.g., Beck et al., 2003; Boehnke, 2022; Tan et al., 2020). These approaches also serve a separate function of increasing representation of perspectives that could be invited into the HiTOP Consortium. Increasing representation of researchers from historically excluded groups

decreases the likelihood of defaulting to majority-group assumptions about normality (e.g., cultural-neutrality or cultural-deficit models) and reification of a single perspective on what constitutes "good quality evidence" when evaluating new research (see Syed & Kathawalla, 2022 for a related discussion). Further, including formal representation and feedback from people with lived experience of mental illness will be essential before the framework is ready for large-scale implementation in practice (Jones et al., 2021).

Conclusion

We have taken the first steps, and it is now time to revise the HiTOP framework substantially to incorporate the evidence that has emerged in the past 5 years. These revisions will include expansions to the coverage of psychopathology domains and changes to the existing HiTOP structure. In making these revisions, it is essential to the aims of the whole endeavour that all are empirically based. Here we have presented the revisions protocol, which draws on a rich history across multiple fields. By focusing on structural validity as a first step before moving to other types of validity, we hope this systematic and transparent approach to evaluating evidence for changes will help the HiTOP framework to fulfil the dual purposes of a classification system that is useful for advancing both research and practice.

References

Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs: General and Applied, 80*(7), 1–37. https://doi.org/10.1037/h0093906

Achenbach, T. M. (1978). The Child Behavior Profile: I. Boys aged 6–11. *Journal of Consulting and Clinical Psychology, 46*(3), 478–488. https://doi.org/10.1037/0022-006X.46.3.478

Achenbach, T. M. (2009). *The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications.* Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.

Aftab, A., & Ryznar E. (2021). Conceptual and historical evolution of psychiatric nosology. *International Review of Psychiatry, 33*(5). 486-499. https://doi.org/10.1080/09540261.2020.1828306

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.* https://doi.org/10.1176/appi.books.9780890425596

American Psychiatric Association. (2021, February). *Guide to submitting proposals for changes to DSM-5*. https://psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/DSM5-Proposal-Submissions-General-Guidance.pdf

Andreasen, N. C. (1995). The validation of psychiatric diagnosis: new models and approaches. *American Journal of Psychiatry 152*, 161–162. doi: 10.1176/ajp.152.2.161

Andrews, G., Goldberg, D. P., Krueger, R. F., Carpenter, W. T., Hyman, S. E., Sachdev, P., & Pine, D. S. (2009). Exploring the feasibility of a meta-structure for DSM-V and ICD-11: could it improve utility and validity?: Paper 1 of 7 of the thematic section: 'A proposal for a meta-structure for DSM-V and ICD-11'. *Psychological Medicine, 39*(12), 1993-2000. https://doi.org/10.1037/pas0001085

Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., ... & Guyatt, G. H. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical*

Epidemiology, *64*(4), 401-406.

Baron-Cohen, S. (2015). Leo Kanner, Hans Asperger, and the discovery of autism. *The Lancet*, *386*(10001), 1329-1330.

Beck, C. T., Bernal, H., & Froman, R. D. (2003). Methods to document semantic equivalence of a translated scale. *Research in Nursing & Health*, 26(1), 64-73. https://doi.org/10.1002/nur.10066

Blashfield, R. K. (1984). *The Classification of Psychopathology*: *Neo-Kraepelinian and Quantitative Approaches*. Springer New York, NY. https://doi.org/10.1007/978-1-4613-2665-6

Blashfield, R. K., & Draguns, J. G. (1976). Toward a taxonomy of psychopathology: The purpose of psychiatric classification. *The British Journal of Psychiatry, 128,* 574–583. https://doi.org/10.1192/bjp.129.6.574

Blashfield, R. K., Keeley, J. W., Flanagan, E. H., & Miles, S. R. (2014). The cycle of classification: DSM-I through DSM-5. *Annual Review of Clinical Psychology*, *10*(1), 25-51.

Boehnke, K. (2022). Let's compare apples and oranges! A plea to demystify measurement equivalence. *American Psychologist*, 77(9), 1160-1168. https://doi.org/10.1037/amp0001080

Borsboom, D. (2017). A network theory of mental disorders. *World psychiatry*, *16*(1), 5-13.

Borsboom, D., Cramer, A. O., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PloS one*, *6*(11), e27407.

Carbone, A. (2020). Cancer classification at the crossroads. *Cancers*, *12*(4), 980.

Carpenter, W. T., Bustillo, J. R., Thaker, G. K., Van Os, J., Krueger, R. F., & Green, M. J. (2009). The psychoses: Cluster 3 of the proposed meta-structure for *DSM-V* and *ICD-11*. *Psychological Medicine, 39*, 2025–2042.

Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the

National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest, 18*(2), 72-145.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment,* 7(3), 309-319.

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment, 31*(12), 1412-1427.

Cicero, D. C., Jonas, K. G., Li, K., Perlman, G., & Kotov, R. (2019). Common taxonomy of traits and symptoms: linking schizophrenia symptoms, schizotypy, and normal personality. *Schizophrenia Bulletin*, 45(6), 1336-1348

Conway, C. C., Forbes, M. K., Forbush, K. T., Fried, E. I., Hallquist, M. N., Kotov, R., ... & Eaton, N. R. (2019). A hierarchical taxonomy of psychopathology can transform mental health research. *Perspectives on Psychological Science*, *14*(3), 419-436.

Cooper, S. E., Hunt, C., Stasik-O'Brien, S. M., Berg, H., Lissek, S., Watson, D., & Krueger, R. F. (2022). The placement of obsessive-compulsive symptoms within a five-factor model of maladaptive personality. *Assessment.*

Corker, K. S. (2020, January 2). Strengths and Weaknesses of Meta-Analyses. *Preprint on OSF*. https://doi.org/10.31234/osf.io/6gcnm

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

de Jonge, P., Wardenaar, K. J., Lim, C. C., Aguilar-Gaxiola, S., Alonso, J., Andrade, L. H., ... & Scott, K. (2018). The cross-national structure of mental disorders: results from the World Mental Health Surveys. *Psychological Medicine*, 48(12), 2073-2084

Dunkley, C. R., Svatko, Y., & Brotto, L. A. (2020). Eating disorders and sexual function reviewed: A transdiagnostic, dimensional perspective. *Current Sexual Health Reports, 12*, 1-14.

Faure, K., & Forbes, M. K. (2021). Clarifying the placement of obsessive-compulsive disorder in the empirical structure of psychopathology. *Journal of Psychopathology and Behavioral*

Assessment, 43, 671-685.

Feighner, J.P., Robins, E., Guzé, S.B., Woodruff, R.A., Winokur, G., Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry, 26*, 57–63. doi:10.1001/archpsyc.1972.01750190059011

First, M. B., Rebello, T. J., Keeley, J. W., Bhargava, R., Dai, Y., Kulygina, M., ... & Reed, G. M. (2018). Do mental health professionals use diagnostic classifications the way we think they do? A global survey. *World Psychiatry, 17*(2), 187-195. https://doi.org/10.1002/wps.20525

Forbes, M. K., Kotov, R., Ruggero, C. J., Watson, D., Zimmerman, M., & Krueger, R. F. (2017). Delineating the joint hierarchical structure of clinical and personality disorders in an outpatient psychiatric sample. *Comprehensive Psychiatry, 79*, 19-30.

Forbes, M. K., Neo, B., Nezami, O. M., Fried, E. I., Faure, K., Michelsen, B., … Dras, M. (2023, March 21). Elemental psychopathology: Distilling constituent symptoms and patterns of repetition in the diagnostic criteria of the DSM-5. *Preprint on OSF* https://doi.org/10.31234/osf.io/u56p2

Forbes, M. K., Sunderland, M., Rapee, R. M., Batterham, P. J., Calear, A. L., Carragher, N., Ruggero, C., Zimmerman, M., Baillie, A. J., Lynch, S. J., Mewton, L., Slade, T., & Krueger, R. F. (2021). A detailed hierarchical model of psychopathology: From individual symptoms up to the general factor of psychopathology. *Clinical Psychological Science*, 9, 139-168.

Fried, E. I. (2022). Studying Mental Health Problems as Systems, Not Syndromes. *Current Directions in Psychological Science*, 9.

Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR* D study. *Journal of Affective Disorders, 172*, 96-102.

Fulford, K. W. M., & Sartorius, N. (2009). The Secret History of ICD and the Hidden Future of DSM. In M. B. L. Bortolotti (Ed.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*.

Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., Waldman, I. D., Cicero, D. C., Conway, C. C., Docherty, A. R., Fried, E. I., Ivanova, M. Y., Jonas, K. G., Latzman, R. D., Patrick, C. J., Reininghaus, U., Tackett, J. L., Wright, A. G. C., & Kotov, R. (2019). Are Fit Indices Used to Test Psychopathology Structure Biased? A Simulation Study. *Journal of Abnormal Psychology*. https://doi.org/10.1037/abn0000434

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., ... & Schünemann, H. J. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, *64*(4), 383-394.

Haeffel, G. J., Jeronimus, B. F., Kaiser, B. N., Weaver, L. J., Soyster, P. D., Fisher, A. J., ... & Lu, W. (2022). Folk classification and factor rotations: Whales, sharks, and the problems with the Hierarchical Taxonomy of Psychopathology (HiTOP). *Clinical Psychological Science*, *10*(2), 259-278.

Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: A quantitative review of taxometric research. *Psychological Medicine, 42*, 903-920.

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

Hippius, H., & Neundörfer, G. (2003). The discovery of Alzheimer's disease. *Dialogues in Clinical Neuroscience.*

Hyman, S. E. (2010). The diagnosis of mental disorders: the problem of reification. *Annual Review of Clinical Psychology, 6*, 155-179.

Ivanova, M. Y., Achenbach, T. M., Dumenci, L., Rescorla, L. A., Almqvist, F., Weintraub, S., ... & Verhulst, F. C. (2007). Testing the 8-syndrome structure of the child behavior checklist in 30 societies. *Journal of Clinical Child and Adolescent Psychology*, *36*(3), 405-417.

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Turner, L. V., Árnadóttir, H. A., Au, A., ... & Zasępa, E. (2015). Syndromes of collateral-reported psychopathology for ages 18-59 in 18

societies. *International Journal of Clinical and Health Psychology*, 15(1), 18-28.

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Guo, J., Althoff, R. R., Kan, K. J., ... & Verhulst, F. C. (2019). Testing syndromes of psychopathology in parent and youth ratings across societies. *Journal of Clinical Child & Adolescent Psychology*, 48(4), 596-609.

Jablensky, A. (2007). Living in a Kraepelinian world: Kraepelin's impact on modern psychiatry. *History of Psychiatry, 18*(3), 381-388.

Jak, S., & Cheung, M. W. -L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods*, *25*(4), 430–455. https://doi.org/10.1037/met0000245

Jones, N., Atterbury, K., Byrne, L., Carras, M., Brown, M., & Phalen, P. (2021). Lived experience, research leadership, and the transformation of mental health services: Building a researcher pipeline. *Psychiatric Services, 72*(5), 591-593.

Kaiser, B. N., Haroz, E. E., Kohrt, B. A., Bolton, P. A., Bass, J. K., & Hinton, D. E. (2015). "Thinking too much": A systematic review of a common idiom of distress. *Social Science & Medicine*, *147*, 170-183.

Keeley, J. W., Morton, H. E., & Blashfield, R. K. (2014). Classification. In P. H. Blaney, R. F. Krueger & T. Millon (Eds.), *Oxford textbook of psychopathology* (3rd ed., pp. 42-70). New York, NY: Oxford University Press.

Kendell, R. E., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry, 160*, 4–12.

Kendell, R. E. (1989). Clinical validity. *Psychological Medicine, 19*, 45–55. Doi:10.1017/S0033291700011016

Kendler, K. S. (1980). The nosologic validity of paranoia (simple delusional disorder). *Archives of General Psychiatry, 37*, 699–706. Doi: 10.1001/archpsyc.1980.01780190097012

Kendler, K. S. (2013). A history of the DSM-5 Scientific Review Committee. *Psychological Medicine*, *43*(9), 1793-1800.

Kirk, S. A., & Kutchins, H. (1992). *The selling of DSM: The rhetoric of science in psychiatry.* Aldine de Gruyter.

Kochunov, P., Hong, L. E., Dennis, E. L., Morey, R. A., Tate, D. F., Wilde, E. A., ... & Jahanshad, N. (2022). ENIGMA-DTI: Translating reproducible white matter deficits into personalized vulnerability metrics in cross-diagnostic psychiatric research. *Human Brain Mapping*, *43*(1), 194-206.

Kotov, R., Chang, S. W., Fochtmann, L. J., Mojtabai, R., Carlson, G. A., Sedler, M. J., & Bromet, E. J. (2011). Schizophrenia in the internalizing-externalizing framework: a third dimension?. *Schizophrenia Bulletin*, *37*(6), 1168-1178

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., Jonas, K. G., Lian, W., Docherty, A. R., & Carpenter, W. T. (2022). Reconceptualizing schizophrenia in the hierarchical taxonomy of psychopathology (HiTOP). *Schizophrenia Research*, *242*, 73.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., . . . Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology, 126*(4), 454-477. doi: http://dx.doi.org/10.1037/abn0000258

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*, 83-108.

Kotov, R., Ruggero, C. J., Krueger, R. F., Watson, D., Yuan, Q., & Zimmerman, M. (2011b). New dimensions in the quantitative classification of mental illness. *Archives of General Psychiatry, 68*(10), 1003-1011.

Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry*, *56*(10), 921-926.

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*(3), 437–447. https://doi.org/10.1037/0021-843X.112.3.437

Krueger, R. F., Hobbs, K. A., Conway, C. C., Dick, D. M., Dretsch, M. N., Eaton, N. R., . . . & HiTOP Utility Workgroup (2021). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): II. Externalizing superspectrum. *World Psychiatry*, *20*(2), 171-193.

Krueger, R. F., Kotov, R., Watson, D., Forbes, M. K., Eaton, N. R., Ruggero, C. J., ... & Zimmermann, J. (2018). Progress in achieving quantitative classification of psychopathology. *World Psychiatry*, *17*(3), 282-293.

Lahey, B. B., Applegate, B., Waldman, I. D., Loft, J. D., Hankin, B. L., & Rick, J. (2004). The structure of child and adolescent psychopathology: generating new hypotheses. *Journal of abnormal psychology*, *113*(3), 358.

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

Lorr, M. (ed.) (1966) *Explorations in Typing Psychotics*, Oxford: Pergamon Press.

Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., ... & Ellison, D. W. (2021). The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-Oncology*, *23*(8), 1231-1251.

Markon, K. E. (2010). Modeling psychopathology structure: A symptom-level analysis of Axis I and II disorders. *Psychological Medicine*, *40*(2), 273-288.

Markon, K. E. (2019). Bifactor and Hierarchical Models: Specification, Inference, and Interpretation. *Annual Review of Clinical Psychology,* 15(1), 51–69.

https://doi.org/10.1146/annurev-clinpsy-050718-095522

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin,* 137(5), 856–879. https://doi.org/10.1037/a0023678

Marshall, E., Freeman, D., & Waite, F. (2020). The experience of body image concerns in patients with persecutory delusions: 'People don't want to sit next to me.' *Psychology and Psychotherapy: Theory, Research and Practice, 93,* 639–655.

Mayes, R., & Rafalovich, A. (2007). Suffer the restless children: the evolution of ADHD and paediatric stimulant use, 1900-1980. *History of Psychiatry, 18*(4), 435-457.

McIntyre, J. S., Yager, J., Everett, A., Galanter, C. A., Lyness, J. M., Nininger, J., Reus, V. I., & Vergare, M. (2021). The *DSM-5* Clinical and Public Health Committee (CPHC): operations, mechanics, controversies and recommendations. *Psychological Medicine, 51*(14), 2493–2500. https://doi.org/10.1017/

Michelini, G., Altschuler, M., & Wilson, S. (2021). *Where do neurodevelopmental problems belong in hierarchical models of psychopathology? Incorporating a new neurodevelopmental spectrum* [Manuscript in preparation].

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine,* 6(7), e1000097.

Moore, T.V. (1930.) The empirical determination of certain syndromes underlying praecox and manic-depressive psychoses. *American Journal of Psychiatry,* 86, 719–738

Moore, T.V. (1933). The essential psychoses and their fundamental syndromes. *Studies in Psychology and Psychiatry,* 3, 1–28

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., ... & Chartier, C. R. (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in*

Psychological Science, *1*(4), 501-515.

Mulaik, S. A., & Quartetti, D. A. (1997). First order or higher order general factor? *Structural Equation Modeling: A Multidisciplinary Journal*, 4(3), 193–211. https://doi.org/10.1080/10705519709540071

Opel, N., Goltermann, J., Hermesdorf, M., Berger, K., Baune, B. T., & Dannlowski, U. (2020). Cross-disorder analysis of brain structural abnormalities in six major psychiatric disorders: a secondary analysis of mega-and meta-analytical findings from the ENIGMA consortium. *Biological Psychiatry*, *88*(9), 678-686.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105906.

Pilecki, B. C., Clegg, J. W., & McKay, D. (2011). The influence of corporate and political interests on models of illness in the evolution of the DSM. *European Psychiatry*, *26*(3), 194-200.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879.

Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A., & Kupfer, D. J. (2013). DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses. *American Journal of Psychiatry*, *170*(1), 59-70.

Rief, W., Hofmann, S. G., Berg, M., Forbes, M. K., Pizzagalli, D. A., Zimmerman, J., Fried, E., & Reed, G. M. (in preparation). Should we change the way we classify psychopathology? A discussion paper.

Ringwald, W., Forbes, M., & Wright, A. (2022). Meta-analysis of structural evidence for the Hierarchical Taxonomy of Psychopathology (HiTOP) model. *Psychological Medicine*, 1-14. doi:10.1017/S0033291721001902

Robins, E., & Guzé, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: Its

application to schizophrenia. *American Journal of Psychiatry, 126*(7), 983-987. doi:10.1176/ajp.126.7.983.

Rodriguez-Seijas, C., Li, J. J., Baling, C., Brandes, C., Bernat, E., Boness, C., Forbes, M. K., … & Eaton, N. R. (2023). Diversity and the Hierarchical Taxonomy of Psychopathology (HiTOP). *Nature Reviews Psychology.*

Rossell, S. L., Labuschange, I., Castle, D. J., & Toh, W. L. (2020). Delusional themes in body dysmorphic disorder (BDD): Comparisons with psychotic disorders and non-clinical controls. *Psychiatry Research, 284*, 112694.

Sellbom, M., Kremyar, A. J., & Wygant, D. B. (2021). Mapping MMPI-3 scales onto the hierarchical taxonomy of psychopathology. *Psychological Assessment*, 33(12), 1153–1168

Simms, L. J., Prisciandaro, J. J., Krueger, R. F., & Goldberg, D. P. (2012). The structure of depression, anxiety and somatic symptoms in primary care. *Psychological Medicine, 42*(1), 15-28.

Simms, L. J., Wright, A. G., Cicero, D., Kotov, R., Mullins-Sweatt, S. N., Sellbom, M., ... & Zimmermann, J. (2022). Development of measures for the Hierarchical Taxonomy of Psychopathology (HiTOP): a collaborative scale development project. *Assessment*, *29*(1), 3-16.

Spitzer, R. L., Robins, E. (1978). Research diagnostic criteria: rationale and reliability *Archives of General Psychiatry, 35*(6), 773–782

Stanton, K., Khoo, S., Watson, D., Gruber, J., Zimmerman, M., & Weinstock, L. M. (2019). Unique and transdiagnostic symptoms of hypomania/mania and unipolar depression. *Clinical Psychological Science, 7*, 471-487.

Steinberg, E., Greenfield, S., Wolman, D. M., Mancher, M., & Graham, R. (Eds.). (2011). *Clinical practice guidelines we can trust*. National Academies Press.

Syed, M., & Kathawalla, U.-K. (2022). Cultural psychology, diversity, and representation in open science. In K. C. McLean (Ed.), *Cultural methods in psychology: Describing and*

transforming cultures (pp. 427–454). Oxford University Press. https://doi.org/10.1093/oso/9780190095949.003.0015

Tan, T. X., Yi, Z., Kim, E., Li, Z., & Cheng, K. (2020). Linguistic equivalence, construct validity, but lack measurement invariance: An illustration of challenges in cross-cultural research on adolescent adjustment. *Cross-Cultural Research: The Journal of Comparative Social Science*, 54(4), 323-345. https://doi.org/10.1177/1069397120914875

Tio, P., Epskamp, S., Noordhof, A., & Borsboom, D. (2016). Mapping the manuals of madness: comparing the *ICD-10* and *DSM-IV-TR* using a network approach. *International Journal of Methods in Psychiatric Research*, *25*(4), 267-276.

Van Bork, R., Wijsen, L. D., & Rhemtulla, M. (2017). Toward a Causal Interpretation of the Common Factor Model. *Disputatio: International Journal of Philosophy*, *9*(47).

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.

Watson, D., & Naragon-Gainey, K. (2014). Personality, emotions, and the emotional disorders. *Clinical Psychological Science, 2*, 422-422.

Watson, D., Levin-Aspenson, H. F., Waszczuk, M. A., Conway, C. C., Dalgleish, T., Dretsch, M. N., … & HiTOP Utility Workgroup (2022). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): III. Emotional dysfunction superspectrum. *World Psychiatry, 21*(1).

Williams, T. F., & Simms, L. J. (2020). Conceptual foundations of descriptive psychopathology and observational designs. In A.G.C. Wright & M.N. Hallquist (Eds). The Cambridge Handbook of Research Methods in Clinical Psychology. Cambridge University Press.

World Health Organization (1949). *International statistical classification of diseases, injuries, and causes of death, 6th revision*. Geneva, Switzerland: Author.

World Health Organization. (2022). *International Classification of Diseases for Mortality and Morbidity Statistics, 11th Edition*.

https://www.who.int/standards/classifications/classification-of-diseases

Wright, A.G.C. (2017). Factor analytic support for the five-factor model. In T.A. Widiger (Ed.) *Oxford handbook of the five-factor model* (pp. 217-242). Oxford, UK: Oxford University Press.

Wright, A. G. C., Krueger, R. F., Hobbs, M. J., Markon, K. E., Eaton, N. R., & Slade, T. (2013). The structure of psychopathology: Toward an expanded quantitative empirical model. *Journal of Abnormal Psychology, 122*(1), 281–294. https://doi.org/10.1037/a0030133

Wright, A. G. C., & Ringwald, W. R. (2022). Personality disorders are dead; long live the interpersonal disorders: Comment on Widiger and Hines (2022). *Personality Disorders: Theory, Research, and Treatment, 13*(4), 364–368. https://doi.org/10.1037/per0000552

Wright, A.G.C. & Simms, L.J. (2015). A metastructural model of mental disorders and pathological personality traits. *Psychological Medicine, 45*(11), 2309-2319.

Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, 64(2), 113–128. https://doi.org/10.1007/BF02294531

Zachar, P. (2015). Psychiatric disorders: natural kinds made by the world or practical kinds made by us? *World Psychiatry, 14*(3), 288–290. doi:10.1001/archpsyc.1972.01750190059011

Zachar, P., Krueger, R. F., & Kendler, K. S. (2016). Personality disorder in DSM-5: An oral history. *Psychological Medicine, 46*(1), 1–10. https://doi.org/10.1017/S0033291715001543