

ARTICLE

Psychiatry's New Validity Crisis: The Problem of Disparate Validation

Nicholas Zautra

Indiana University Bloomington, Cognitive Science Program, 1001 East 10th Street, Bloomington, IN 47405, USA.

Email: nzautra@iu.edu

(Received 12 June 2024; revised 30 September 2024; accepted 03 December 2024; first published online 10 December 2024)

Abstract

In response to the crisis in validity of the *Diagnostic and Statistical Manual of Mental Disorders*, psychiatry has seen a proliferation of alternative research frameworks for studying and classifying psychiatric disorders. In this paper, I argue that the existence of multiple frameworks in which each employs their own standards of validity is problematic methodologically speaking for trying to do any kind of unified validation work. Fundamental disagreements concerning the underlying phenomenon, sources of validating evidence, and the very nature of validity move each framework into an unrecognized plurality. The consequence for psychiatry is a new validity crisis.

1. Introduction

Over the past forty years, psychiatry has faced a “crisis in confidence” in the validity of its psychiatric classifications (Phillips 2013; Poland and Tekin 2017). Validity in this context, defined by Zachar (2012) as “big-V” validity, is understood as whether psychiatry’s diagnostic categories as featured in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) are either “valid” or “not valid” according to a single conception of validity—whether our psychiatric diagnoses can be judged to stand for real underlying clinical syndromes. In response, three alternative research frameworks, the Hierarchical Taxonomy of Psychopathology (HiTOP), the Network Approach to Psychopathology, and the Research Domain Criteria (RDoC), have emerged with the shared goal of studying and classifying psychiatric disorders in new ways. By approaching psychiatry’s validity problem in ways unbound by DSM, the hope is to bolster the validity of future psychiatric classifications. In addition, the DSM has recently been updated with a new continuous improvement revision model to address concerns regarding the validity of its diagnostic categories and contribute to a more evidence-based scientific nosology (American Psychiatric Association n.d.).

An as yet unexplored aspect of psychiatry's "validity crisis" is related to disagreements regarding *standards* of validity. Disagreements regarding standards of validity that amount to multiple distinct conceptions of validity point to a thornier methodological problem for psychiatry I term "the problem of disparate validation." This two-part problem can be summarized as follows: Scientific psychiatry aims at achieving empirically informed classifications that demonstrate validity, in that they correspond to real attributes of psychopathology. To achieve this, alternative frameworks are now approaching the conceptualization, testing, organizing, and validation of features of psychopathology by their own standards to inform more valid psychiatric classifications. The first problem is, given a classification system, by whose standard of validity should such a system be validated? Second, when we attempt to validate classifications informed by differing validity standards, will any such validation process be capable of assessing a unified fundamental conception of validity, or will each approach only be valid under its own conception?

In this paper, I assess the problem of disparate validation through faithful reconstructions of what I term the Holy Quadrinity of distinct conceptions of validity in psychiatry. By evaluating psychiatry's distinct validity conceptions, I argue that, despite the appearance of a shared goal of informing valid classifications, the existence of multiple frameworks in which each employs their own standards of validity is problematic methodologically speaking for trying to do any kind of unified validation work. At its core, differing standards that inspire fundamental disagreements concerning (i) the underlying phenomenon, (ii) the sources of validating evidence, and (iii) the very nature of validity and validation move each framework further into a state of an *unrecognized plurality*, where we have yet to fully realize to what extent these frameworks are not at all talking about the same thing when it comes to validity and as a result are presently engaged in very different projects with different aims. The consequence is a new validity crisis for psychiatry.

I conclude with a positive program in which I recommend how different frameworks with distinct validation procedures can achieve validity under their own specific conceptualization while also coming to inform another through a kind of *interactive pluralism*. To this end, I offer general recommendations as to how the frameworks may stabilize their own validity principles and procedures. Finally, given the inability to establish a unified conception of validity, I advocate for developing convergent standards of *utility*, which may help practically compare current and alternative frameworks.

2. Validity in psychiatry is *not* construct validity

Outside of psychiatry, validity is understood as a technical measurement concept and is "the most fundamental consideration" in developing and evaluating psychological measurement instruments for both practical and epistemic purposes (American Educational Research Association et al. 2014, 11; Cizek 2020). Modern validity theory, as outlined in the *Standards for Educational and Psychological Testing* (American Educational Research Association et al. 2014), maintains all validity evidence is integrated into a single, unified concept of validity referred to as *construct validity*. Construct validity falls within construct validity theory, defined by psychologist and construct validity expert Kathleen Slaney as the "general theoretical approach and

set of methods for judging whether empirical inferences and decisions made based on quantitative data are licensed by the most current theory regarding the construct purportedly measured by the test or assessment tool in question” (2017, 1).

Seeing how advances in validity theory in the *Standards* occur separate from psychiatry, it is important first to consider in what ways validity and validation in educational and psychological testing (i.e., construct validity) relate, if at all, to big-V validity. A further motivation for this inquiry is that when scholars in psychiatry and philosophy of science characterize psychiatry’s validity crisis, they tend to emphasize the validity the DSM’s diagnostic categories lack is “construct validity”:

Most DSM diagnostic categories do not have *construct validity*, that is, they do not “carve nature at the joints” by picking out just one kind of condition with a distinctive etiology. Rather, current categories are syndromes that encompass many different etiologies. (Wakefield 2013, 826)

There is scarce evidence that any DSM diagnostic categories—other than a small handful (viz., “schizophrenia,” “bipolar disorder,” “intellectual disability,” “neurocognitive disorders”)—possess construct validity. To have construct validity, a diagnostic category should accurately represent a construct as defined by theory. (Tsou 2021, 69)

There are several instances where construct validity is taken to represent this fundamental big-V validity. Specifically, there are attempts to apply concepts from validity in psychological measurement specifically to Robins and Guze’s (1970) method for achieving what they refer to as *diagnostic validity*—the empirical method adopted and updated alongside the DSM for establishing big-V validity based in evidence from clinical description, etiology, pathophysiology, prognosis, and treatment response. Psychologist Catherina Hartman and colleagues (2001, 818) claim “the hallmark of construct validity is external construct validity . . . through differential relations of current clinical concepts with aetiology, course, prognosis, or dysregulations in the neurobiological or cognitive system,” suggesting the five phases all contribute to a unified conception of construct validity. Aboraya et al. claim “Robins and Guze actually were the first to articulate the elements of construct validity in psychiatry” (2005, 50), and “construct validity, consisting of validity criteria, is the core of psychiatry” (2005, 55). Philosopher of science Kenneth Schaffner, who has provided very thoughtful work in the philosophy of psychiatry and validity in psychiatry, has also drawn a close connection between diagnostic validity and construct validity: “For our purposes, the notion of ‘diagnostic validity’ is of special importance. This concept comes from Robins and Guze’s classic and extraordinary influential 1970 article noted earlier. In a way, this article adapted the construct validity notion to psychiatric diagnosis by using the term ‘diagnostic validity’ (Robins and Guze 1970), though there is no reference to the term ‘construct validity’ nor to Cronbach and Meehl’s (1955) article in their 1970 paper” (Schaffner 2012, 169).

While these interpretations of relating construct validity as being consistent with or being the basis of validity in psychiatry may appear feasible, such interpretations are ultimately inaccurate. The DSM’s standards of validity have essentially remained consistent in establishing a type of validity distinct from construct validity. Two

examples demonstrate this distinction. First, psychiatrist John Livesley and personality psychologist Douglas Jackson first proposed constructing and evaluating the DSM's diagnostic categories (Livesley and Jackson 1992) wherein validating evidence from psychological measurement, such as content, criterion, predictive, convergent, and divergent sources of validity evidence would be the new required basis for establishing validity. Their proposal, published just after the release of DSM IV, was motivated by the fact that through the development of DSM IV, such a validation process had never been done before, and was intentionally presented in contrast to Robins and Guze's method. Second, Dr. Robert Kendell, one of the leading authorities on psychiatric classification in the DSM, clarified the distinction between validity in the DSM and validity concepts typically associated with construct validity that were ultimately never adopted:

Psychologists are accustomed to distinguishing several different kinds of validity—construct, concurrent, content, predictive, and so on. Although these are useful distinctions in many settings, in the context of clinical medicine statements about diagnostic validity are essentially statements about predictive power and hence practical utility. The more information a diagnosis provides about outcome and response to treatment—and thus about which treatments are appropriate—the higher its validity and the greater its utility. (Kendell 2002, 7).

The brief detour into the concepts of validity in educational and psychological testing and misinterpretations of construct validity onto diagnostic validity provides an important normative lesson for psychiatry: the concept of validity is complicated, be it for the experts in modern validity theory who contribute to the *Standards*, or for the psychiatrists and philosophers of science taking concepts and applying them to the DSM. It is very easy to engage in validation work without really having a thorough understanding of what validation amounts to and, as Slaney has observed, it is even easier to apply validity concepts and procedures in inconsistent and illogical ways (2017, 13–15). The fact there is little if any acknowledgment in psychiatry or the philosophy of science that construct validity is not the same conception of validity as diagnostic validity is an example of an *unrecognized plurality* in psychiatry. Given the complexities of the validation process as well as a tendency to put forth underspecified or inconsistent conceptions of validity, we have yet to realize the extent to which these disparate validation procedures support very distinct conceptions of validity. Now, as psychiatry faces not one but four distinct conceptions of validity, coming to a more precise understanding of psychiatry's different methods, procedures, and standards of validity is paramount.

3. The holy quadrinity of validity in psychiatry

The Holy Quadrinity of validity plays off the so-called trinitarian doctrine of criterion validity, content validity, and construct validity, and instead represents the four distinct conceptions of validity in psychiatry. Each conception is uniquely associated with one of psychiatry's four main scientific approaches. The Holy Quadrinity

includes *diagnostic validity*, associated with the DSM; *structure-first psychometric validity*, associated with HiTOP; *network psychometric validity*, associated with the Network Approach; and *etio-pathophysiological validity*, associated with RDoC.

3.1. Diagnostic validity

Validity in the DSM, referred to as diagnostic validity, was first articulated by Eli Robins (1921–1994) and Samuel Guze (1925–1996) in 1970 and subsequently updated alongside the various iterations of the DSM’s categorical classification system in the context of clinical medicine. Robins and Guze were motivated by an interest in achieving psychiatrist Emil Kraepelin’s (1856–1926) big idea: psychiatric disorders may be shown to be discrete disease entities that can be accurately identified through clinical observation of their signs and symptoms, direct observation of their pathological anatomy and underlying physiology, or through the study of their etiology. Their overall strategy was simple: (i) identify discrete and homogenous diagnostic groups based on what we know, being observable signs and symptoms, and (ii) study those diagnostic groupings to discover their underlying and corresponding pathological processes.

Diagnostic validity is the extent a DSM-based *diagnostic category* (e.g., Major Depressive Disorder), comprised of a set of operationalized *diagnostic criteria* being observable signs and symptoms considered infallible indicators of an underlying clinical syndrome, is supported by a specific set of *validators*, understood as acceptable sources of validating evidence for a diagnostic category. The central question is “whether we have any confidence in the validity of this syndrome based on the set of validators” (Kendler et al. 2009, 8). Diagnostic validity is based in evaluations made by expert-led committees from twenty-one disorder groupings of the “strength of evidence for each of the validators” and “the overall strength of evidence across all validators” (Kendler et al. 2009, 3).

Evaluations of the accumulation of evidence from the validators, i.e., the aggregation of the validators, contribute to a judgment regarding the diagnostic validity of a diagnostic category. An implicit hierarchy among the validators exists, which reflects Kraepelin’s original clinical method of prioritizing *predictive* sources of validating evidence, such as *course of illness* (i.e., evidence of a diagnostic category’s predicted duration) and *response to treatment*, over and above external sources of validating evidence such as *biological markers* (i.e., evidence describing the pathophysiological correlates)—although evidence from biological markers has recently increased in priority within the past 15 years. Table 1 lists the current set of validators for the DSM, of which those denoted by an asterisk (*) are deemed high priority.

Furthermore, whether a diagnostic category is judged to have sufficient diagnostic validity is only one of several considerations for the category’s placement in the DSM. A host of other considerations includes evidence the diagnostic category sufficiently meets criteria for a “Mental (Psychiatric) Diagnosis”; evidence of the category’s *reliability*, defined by DSM-5-TR as “the degree to which two clinicians could independently arrive at the same diagnosis for a given patient” (American Psychological Association 2022, 8); and *clinical utility*, defined as the ability to “help clinicians to determine prognosis, treatment plans, and potential treatment outcomes

Table 1. List of validators of DSM-5-TR

Antecedent validators
a. *Familial aggregation and/or co-aggregation (i.e., family, twin, or adoption studies)
b. Socio-demographic and cultural factors
c. Environmental risk factors
d. Prior psychiatric history
Concurrent validators
a. Cognitive, emotional, temperament, and personality correlates (unrelated to the diagnostic criteria)
b. *Biological markers, e.g., molecular genetics, neural substrates
c. Patterns of comorbidity
d. *Degree or nature of functional impairment
Predictive validators
a. *Diagnostic stability
b. *Course of illness
c. *Response to treatment

for their patients” (American Psychological Association 2022, 14); as well as many practical utility-related considerations such as the need for the category, ease of use, clinician acceptance, any potential harm caused by the category, and evidence of available treatments.

3.2. Structure-first psychometric validity

HiTOP’s specific conception of validity, which I interpret as structure-first psychometric validity, developed independently from diagnostic validity in the context of psychometrics, a quantitative scientific research tradition traced to the introduction of the common factor model of general intelligence by Charles Spearman in the early twentieth century. HiTOP’s psychometric approach rejects the DSM’s notion of a psychiatric disorder being the result of an underlying discrete clinical syndrome and instead views psychopathology as arising from coherent and distinct latent (unobserved) dimensions as represented by its HiTOP constructs that hold a shared causal influence on a set of indicator (observed) variables, being the symptom groupings of psychopathology. For example, under HiTOP, a clinician would not interpret the symptoms of depression as arising from a single underlying syndrome such as Major Depressive Disorder but would instead view depression symptoms as interrelated and overlapping with one another to varying degrees, such that there is ultimately a broader and coherent dimension, the *internalizing spectra*, which underlies all symptoms of depression and symptoms from related disorders (e.g., generalized anxiety).

HiTOP’s data-driven strategy, described as evidence-over-experts, uses a statistical technique called factor analysis to derive the most common hierarchically organized factors of psychopathology, i.e., HiTOP *spectra*, which it refers to as the structure of psychopathology, from psychological testing data (Kotov et al. 2017). HiTOP then organizes the empirically derived structure from broadest (e.g., the general factor of psychopathology or “p-factor”) to narrowest (e.g., homogeneous symptom components/maladaptive traits). HiTOP ultimately seeks to validate the entire

data-driven structure to inform a future transdiagnostic classification system that may replace the DSM (Forbes et al. 2024).

Structure-first psychometric validity falls within *psychometric validity*, broadly understood as the degree to which a test, being a response to a standardized situation devised to measure a psychological construct, has the desired psychometric features (e.g., various conceptions of validity, reliability, utility, etc.). Unlike diagnostic validity, structure-first psychometric validity is based in construct validity, the unified conception of validity that encompasses all psychometric validity within psychological testing. Structure-first psychometric validity is centered on supporting the development of psychiatric classifications that carry both *scientific accuracy*, understood as the degree HiTOP's constructs represent true features of psychopathology, and *clinical utility*, the degree classifications are considered practically and pragmatically clinically useful.

Structure-first psychometric validity is an ongoing three-stage validation process. The first and most crucial stage is the evaluation of structural validity (hence, structure-first), being the degree to which a particular construct accounts for the empirically observed *covariance* (i.e., the direct relationship) between different signs and symptoms of psychopathology. *Structural evidence* is based in exploratory and confirmatory factor-analytic research with a preference for continuous latent variable models. Such models produce factor loadings, i.e., standardized correlations between the original variables and an underlying factor (the construct), which are assessed based on the model's goodness of fit. The second stage is an evaluation of *external validity*, referring to the degree to which evidence for a certain HiTOP construct correlates with other (relevant) indicators of that construct. The third stage is an evaluation of *reliability*, the extent to which all items on a test measure the same construct, as well as clinical utility, and predictive utility, the degree to which a construct is helpful in differentially predicting outcomes of interest. Through structure-first psychometric validity, HiTOP aims to validate its constructs as well as the broader system of multiple, hierarchically organized constructs, i.e., the entire HiTOP hierarchical model itself.

3.3. Network psychometric validity

Network psychometric validity, also based in psychometric validity, operates within the network perspective, which represents a family of models and methodologies that draw on the network approach to psychopathology, network theory, complex systems theory, and applications in network science to model and study mental disorders as complex causal systems. For example, under the network approach, a depressive episode is hypothesized to arise from the causal interaction between symptoms such as depressed mood, anhedonia, and others (e.g., insomnia, fatigue). As a result, symptoms are not conceived as fallible indicators of some underlying common cause such as the DSM's Major Depressive Disorder or HiTOP's Internalizing Spectra. Instead, it is the mutual interaction between symptoms that constitutes depression itself.

Unlike the DSM or HiTOP, the network approach to psychopathology, which first gained traction from psychometrician Denny Borsboom (2008; 2017) and affiliated members of a psychometrics research group based in the Netherlands, is not affiliated

with any single governing body or research organization, nor does it (yet) directly inform a system of psychiatric classification. The aim of the network approach is to replace the data-driven latent variable models typical of the HiTOP approach, which the network approach claims do not support testable explanations, with theory-based network models of psychopathology. This “theory-first” approach means selecting network models based on theoretical reasons which, it is argued, provide the necessary rationale for developing and testing network theories to inform future psychiatric classifications.

A *network model* of mental disorders is a statistical model that represents features of a mental disorder as derived from a specific *network hypothesis*, being a testable and falsifiable hypothesis of how the components in a network influence each other over time. A *network* is a representation of the relationships (formally called edges) between constituent variables (formally called nodes) within a system. In psychopathology networks, the nodes represent various constituent elements of psychopathology (e.g., symptoms, biomarkers, cognitive processes). *Edges* represent conditional associations between nodes. Two or more networks may be connected by what has been referred to as *bridge symptoms* (or bridge nodes). The *external field* is an area outside of the network whose components (e.g., “stress”) may causally intervene on the nodes or edges inside the network. In a network model, the more thorough the connections of the nodes in the network, the more likely the network is to remain in a dysfunctional state even after removal of the original biopsychosocial variables (i.e., hysteresis), and may also reflect higher levels of severity in dysfunction.

Network psychometric validity can be summarized as being comprised of three stages. The first stage is the validation of the individual components in the network, based in the concept of *node validity*, a model-specific validation process (Bringmann et al. 2022). Node validity is a two-step validation procedure that involves (i) *node selection*, referring to the adequacy of selecting appropriate variables as nodes in a network model, and (ii) *node assessment*, referring to the quality of the operationalizations used for selected variables. The second stage is the validation of the dynamical relations between the components, i.e., the validation of the *network structure*, whereby the dynamic relation between specific nodes may be understood as a kind of useful construct to be validated. A third and final stage in network validation is to go beyond validating nodes and their relations, and toward deriving empirical implications from the model to test (and subsequently validate) a network hypothesis. At present, this stage is not discussed in terms of an explicit validation procedure, but rather in terms of validity-adjacent concepts such as *testability*, the ability to subject a hypothesis to appropriate empirical testing conditions, and *falsifiability*, the ability of a hypothesis to be rejected by empirical testing.

3.4. Etio-pathophysiological validity

Etio-pathophysiological validity is the conception of validity of RDoC, an alternative research framework of the National Institute of Mental Health (NIMH). Initially launched in 2009, in 2015 the RDoC program changed leadership following Thomas Insel’s departure as director, resulting in a significant change in research priorities—an RDoC 2.0—that would eventually lead to the retirement of the original RDoC matrix. RDoC represents an integrative approach to validation that attempts to bring

the syndromal model, psychometrics, dimensionality, and multicausality together with cognitive neuroscience in one unifying validity framework. RDoC's overall strategy is to group patients for clinical studies based on fundamental dimensions of behavior and neurobiological measures (genes, circuits, etc.). By adopting an approach that breaks up the DSM's diagnostic categories like HiTOP and the network approach, yet unlike those approaches fixes measuring the underlying biology of the phenomena as its starting point, RDoC aims to inform a more valid classification system in the future based firmly in biology.

The RDoC approach employs what I interpret as a dual-track model of validation. The first track, which I term *biology-first function validity*, is centered on the validation of the tools of the RDoC Framework, referred to as the "concepts for investigation," with a primary focus on RDoC constructs. These concepts include six major functional *domains*, which represent current understanding of the major systems of cognition, motivation, and social behavior, i.e., those systems which, when there is dysregulation and dysfunction within/across them, are thought to give rise to psychological and behavioral impairments. Each domain is accompanied by three to six *constructs*, i.e., concepts summarizing data about a specified psychological/biological dimension of behavior, recently defined as empirical functions. *Units of analysis* are the methods and instruments used to study the constructs from a normal to abnormal range of functioning.

The second track, which is intended to be informed and supported by the first, is the open-ended and under-specified validation of a future diagnostic classification system, which I term *biology-first syndromal validity*. Biology-first syndromal validity would be applied to a classification system that can accommodate RDoC's specific hypothesis concerning mental disorders as representing broad and biologically heterogeneous syndromes as opposed to discrete clinical syndromes and would require validation using etiology, pathophysiology, prognosis, and treatment response measures. A validation process on which biology-first syndromal validity may come to be modeled comes from an RDoC predecessor, the Bipolar and Schizophrenia Network for Intermediate Phenotypes (B-SNIP). B-SNIP's four-stage process for developing valid psychiatric classifications outlined by Keshavan et al. (2013) amounts to (i) agnostic deconstruction of disease dimensions, identifying disease markers and endophenotypes; (ii) applying such markers across translational domains from behaviors to molecules; (iii) re-clustering cross-cutting bio-behavioral data into *biotypes*, being transdiagnostic clusters defined by responses on measures across units of analyses using modern phenotypic and biometric approaches; and (iv) validating biotypes using etio-pathology, outcomes, and treatment-response measures.

The term biology-first applies to both tracks of validation and is in service of RDoC's primary aims: (i) develop an etiological and pathophysiological understanding of human systems of normal and abnormal functioning; and (ii) contribute to a future biologically based system of classification. In the first track, biology-first is evidenced by the specific criteria for RDoC constructs. For an RDoC construct to be initially selected and subsequently considered valid, it must include evidence that a neural circuit or biologically based system plays a role in implementing the function. In the second track, biology-first reflects the notion that a future classification system will be validating syndromes that have been shaped via an understanding of their biological basis. As mentioned, a leading frontrunner is the new classificatory concept

of biotypes, which RDoC considers to be “more biologically valid groupings than the diagnostic categories” (Cuthbert 2020, 84).

4. Similarities across the disparate conceptions of validity

Despite the four approaches having developed distinct conceptions of validity, several commonalities exist across them that have gone almost entirely unnoticed. The three most significant commonalities between the approaches are discussed below.

4.1. A return to the original validators of Robins and Guze (1970)

One of the most surprising commonalities is that, for those frameworks seeking to inform a future psychiatric classification system, i.e., HiTOP and RDoC, there is a return to some of the original validators of Robins and Guze (1970) such as etiology (i.e., family history), prognosis (i.e., the likely course or outcome of the illness), and treatment response. While the DSM, RDoC, and HiTOP have all emerged within the context of patient care and thus a focus on specific validators which support the clinical utility of psychiatric diagnoses may seem expected, a return to Robins and Guze is unforeseen for two reasons. First, the primary motivation of the alternative frameworks was to go *beyond* the DSM’s iterative approach to validation and toward a “paradigm shift model” (Kendler and First, 2010) where the original approach is discarded due to the belief that the limits within the original paradigm had been reached. Second, both HiTOP and RDoC have been so outwardly critical of the DSM’s standards for validity as well as utility that a return to what they deem problematic seems counterintuitive.

To clarify, HiTOP and RDoC do explore other validity avenues first. HiTOP first develops structural evidence of the HiTOP spectra, and then proceeds to validate HiTOP constructs against some of the original validators. For RDoC, following a focus on biology-first functional validity of its RDoC constructs, validation of a future RDoC-informed classification system via biology-first syndromal validity then depends on prognosis and treatment response. Ultimately, a return to some of the original validators makes HiTOP and RDoC much more in line with the DSM’s iterative approach in comparison to the paradigm shift models they have previously been characterized under.

4.2. Expert curation

A second unpredicted commonality is that each alternative framework employs expert curation, meaning decisions as to what is ultimately included in their model(s) or classification system(s) are based on compromises between experts. These experts, specifically MD and PhD key opinion leaders who are selected to an evaluation or oversight committee, are tasked with assessing and judging the validating evidence in relation to other various epistemic and non-epistemic aims of an approach, e.g., compromising between the truth of the classification and the degree to which it is considered to be useful in clinical practice. This is also surprising since the primary criticisms of the DSM are on the one hand that it is too expert-based and is thus overtly biased and subjective, and on the other hand that it fails to incorporate the “right” experts by not including patient perspectives in the revision process (e.g., see

Tekin 2022; Knox 2022). When alternative frameworks are promoted, the perception they put forward is that their “data-driven,” “theory-driven,” or “biologically driven” approach removes the overreliance on experts, permitting an empirically based process that better “discovers” its classifications in an “objective” and “scientific” manner.

To their credit, all the frameworks are trying to improve their approach to expert curation. The DSM seeks to systematize and standardize its expert-driven decisions in a way that permits empirical evidence to carry more weight. HiTOP distinguishes various features in its overarching model, such as the separation between Somatic and Internalizing Spectra, or the “p-factor” based on expert decisions as opposed to those features being empirically determined by factor analysis. At the same time, HiTOP implements an evidence-based GRADE rating system in their revision decisions to make the use of experts more objective. The Network Approach utilizes a partially expert-led process of node selection and node assessment, whereby the “clinical or theoretical hypothesis of a clinician . . . plays a role in the choice of the set of variables or nodes in the network” (Bringmann et al. 2022, 3). In turn, they maintain that rigorous empirical testing is the ultimate arbiter. RDoC developed the original RDoC matrix during a series of two-day workshops whereby experts hurriedly populated RDoC constructs into the matrix they took to be the most reasonable. RDoC still adopts specific tasks and experimental paradigms experts deem important, but their strategy now considers constructs in the updated RDoC Framework as more exemplars to be tested, thus reducing the strength of their claims regarding expert-selected constructs. Notably, the focus across frameworks on making expert decisions more “objective” advocates for what Gagné-Julien has criticized as the “ideal of value-free science” (2021, 9401) as well as a narrow conception of who counts as a so-called expert.

4.3. *Validity is broadly understood as that which is considered “good” or “desirable”*

Despite the elaborate presentation of validity as being empirically or scientifically based, accompanied by long lists of validators or intricate sequencing of how such and such evidence should be evaluated, a third commonality is that validity for each approach in the broadest conception boils down to that which is considered to be “good” or “desirable.” That validity in psychiatry may be interpreted so broadly is not inherently a criticism. In the 2010s, the great validity debate within educational and psychological testing centered specifically on how and whether to define validity in a narrow or very broad sense (Newton and Baird 2016). The question was whether validity should be defined in a narrow, traditional conception of determining whether a test is actually measuring the thing you want to measure (Markus 2015)—which is far more difficult to establish—versus a broad conception so that validity is more flexible and may come to mean anything to do with whether an assessment procedure is “good” or “bad” (Newton and Shaw 2015). How to define validity was essentially a debate concerning how the term validity should be used, while the more difficult philosophical questions as to whether validity should establish a link between the construct one is attempting to measure and the measuring instrument, or whether validity should be thought of as demonstrating the thing being measured exists, did not drive the debates.

Evidence to suggest that validity may be understood more broadly in psychiatry is that no current conception of validity across the four approaches engages in more difficult measurement questions in terms of what we are *really* doing when we're measuring and subsequently validating, or what is really required of psychological measurement so we may be confident the thing underlying the construct (e.g., a clinical syndrome, the general factor of psychopathology) exists. There is a heightened focus on conceptualizing the phenomenon, thinking about how to study it, deciding which types of evidence are important, and developing and testing specific hypotheses, but psychiatry has seldom engaged with the more challenging and rigorous philosophical problems of measurement.

5. Disparate standards of validity: Psychiatry's new validity crisis

While some unexpected and important similarities exist in validity among the various frameworks, the stark differences in validity reveal a deeper and more troubling divide for psychiatry than previously recognized. These discrepancies arise from fundamentally different interpretations of three critical aspects: the underlying phenomenon being studied, the accepted criteria for validating evidence, and the nature of validity itself.

First, each framework aims to improve the validity of psychiatric classifications, yet each framework operates under a vastly different understanding of what a psychiatric disorder amounts to, which leads to initial significant impasses. For instance, the HiTOP framework dismisses the DSM's categorical approach to psychiatric classification as lacking any real-world grounding of psychopathology as it truly exists, asserting that its diagnostic categories cause more harm than good. In contrast, the Network Approach strongly criticizes HiTOP's approach, which suggests its data-driven latent dimensional constructs of psychopathology, such as the "p-factor," exist in the brain, insisting that a conception of psychopathology as arising from the interaction between elements of a complex dynamical system is far more accurate. Meanwhile, RDoC argues that both HiTOP and the DSM overlook the biological realities underlying psychiatric disorders, positing that valid psychiatric classifications must first be rooted in their neurobiological mechanisms. The DSM views these alternative paradigm-shift approaches as misguided, countering with evidence supporting its iterative categorical model. This chasm in what constitutes the underlying phenomenon creates an initial battleground for establishing valid classifications in psychiatry whereby each framework perceives the others as flawed from the start.

Differing in their understanding of the phenomenon next influences what each framework considers acceptable sources of validating evidence. Each framework has meticulously defined its own criteria, encompassing relevant validity concepts, evaluation methods, and the relationship between validity, reliability, and utility. However, these criteria are shaped by their interpretations of the underlying phenomena as well as their disparate standards of validation, leading to a situation where each framework interprets and dismisses the validating evidence of the others as inadequate or irrelevant. To give an example, consider the disagreement between frameworks in evaluating the degree to which the DSM's diagnostic categories lack validity and why, based on *their own standards* for what counts as validating evidence. DSM proponents will argue that while poor performance on concurrent validators

(evidenced by a lack of biological markers) for DSM-based diagnostic categories exists, performance on predictive validators (e.g., differential response to treatment or diagnostic stability) justifies their being “valid.” In turn, HiTOP believes the DSM’s symptoms-first approach to be fundamentally misguided and lacking in structural validity. Since DSM categories do not reflect pure, dimensional constructs, not only do they not hold validity, but they should not be considered reliable. The Network Approach, in thinking of mental disorders as “systems, not syndromes,” means the DSM’s diagnostic categories cannot be tested in the way the Network Approach deems critical, and thus ultimately should not be interpreted as having validity. For RDoC, the DSM’s failure to locate biological underpinnings in its categories is the primary reason the former NIMH Director Thomas Insel stated the DSM has “0% validity” (Lynch 2018, 5). Thus, a lack of validity in this approach is not simply the result of a failure of performance on the validators but is due to disagreements on the interpretation and meaning of the validating evidence.

Most alarmingly, the nature of validity varies much more dramatically across the four conceptions than previously recognized, representing a deeper conceptual and methodological divide related to how each views validity, how each framework’s conception interacts with and relates to construct validity theory, and how each maintains distinct engagements with scientific realism.

For the DSM, diagnostic validity is an evaluation of predictive power and clinical utility. For HiTOP, structure-first psychometric validity is *only* one thing: scientific accuracy, which is overwhelmingly based in structural validity. For the Network Approach, network psychometric validity is theory testing. Lastly, RDoC’s etio-pathophysiological validity is the establishment of biological linkages.

In terms of their relation to construct validity theory, the DSM’s diagnostic validity, while not based in construct validity, maintains some general overlap with surface-level features, e.g., both view validity as a unitary concept supported by sources of validity evidence. HiTOP, being based in psychometric validity, at times appears as a mishmash of disparate conceptions of construct validity theory—a practice also common in experimental psychology that “carries nontrivial implications for both theory and practice” (Slaney 2017, 6). The network approach also draws on construct validity theory, but only enough to support appropriate conditions for testing network hypotheses. With RDoC, specific interpretations of convergent and divergent forms of construct validity are invoked that differ in emphasis and application.

While the frameworks and their disparate validity conceptions share aspirations for scientific realism, their underlying philosophical foundations include mixed and even contrasting positions. For example, the DSM’s diagnostic validity features a mix of operationalism, by which concepts are stipulated in terms of their operations to establish their existence, as well as scientific realism. HiTOP’s structure-first psychometric validity may (unintentionally) rely on what others have characterized as a positivist characterization of construct validity theory, whereby hypothetical constructs are conceived as being without reference or meaning and at best useful fictions (something the HiTOP framework would not agree with). At the same time, such appeals may also be interpreted as a “methodological move” for permitting the testing of hypotheses based in a form of scientific realism. The network approach’s first two stages of network psychometric validity maintain what is referred to as a

constructivist-realist view associated with validity theorist Samuel Messick (1989) whereby validity is not dependent on the reality of the construct but is instead a property of the inferences made. In contrast, the third stage of network theory testing within network psychometric validity maintains a realist interpretation of psychological attributes. Lastly, RDoC's biology-first function validity also maintains a constructivist-realist stance in relation to the validation of RDoC constructs, whereas biology-first syndromal validity holds scientific realist underpinnings.

From examining the key differences, we see how when one framework conceptualizes the phenomenon in one way and pursues a specific validation procedure, another framework may reject this in part or entirely, thinking of the underlying phenomenon in a different way and interpreting validating evidence through their standards, and thus going about on their own validation path. Similar validity concepts (e.g., "construct validity") and terms (e.g., "construct") are used in service of seemingly similar aims of establishing validity, but only presently within their own framing and thus only contributing to their own distinct conceptualization of validation. The current situation is such that each approach rejects the ideas of the others because their attitude is "I know what validity is, and you don't" and "my approach is the best and only way to do it."

Unfortunately, the approaches are currently in what I take to be a state of an *unrecognized plurality*, meaning that while each approach fully understands the other frameworks to be doing psychiatric research and classification in distinct ways, psychiatry has yet to fully recognize the implications of their distinct conceptions of validity in the form of three key methodological difficulties: difficulties in evaluating between frameworks, difficulties integrating and coordinating between frameworks, and difficulties in establishing a unified concept of validity.

5.1. Difficulties evaluating between frameworks

One of the most immediate implications is the difficulty in comparing frameworks in terms of which is more valid. Consider a recent debate between critics of HiTOP and HiTOP proponents that represents just a microcosm of the ongoing validity debates shaping validity standards. In their paper "Folk Classification and Factor Rotations: Whales, Sharks, and the Problems with the Hierarchical Taxonomy of Psychopathology (HiTOP)," psychologists Haefl et al. (2021) challenged the notion that HiTOP significantly improves upon the DSM. In critiquing HiTOP, they point to two major weaknesses that limit its potential toward achieving validity: its "data-driven" approach is atheoretical and, as a result, is unfalsifiable, meaning it is not suitable for theory-building. They further criticize HiTOP's simple-structure factor-analytic approach and its use of the degree of model fit as an indicator of validity (i.e., structural validity), claiming "... HiTOP's hierarchical approach is not valid" (2021, 262) nor does it hold the potential for achieving validity.

In their response, "Answering Questions about the Hierarchical Taxonomy of Psychopathology (HiTOP): Analogies to Whales and Sharks Miss the Boat," HiTOP proponents DeYoung et al. clarify HiTOP's validation procedure, arguing its atheoretical, data-driven approach "maximizes coherence of constructs and distinctiveness between them" (2021, 280) and is thus beneficial and valid, while also suggesting their approach is capable of hypothesis testing "according to their fit

to the data” (2021, 281). In response to the response, Haefel et al. assert DeYoung et al. (2021) fail to meet their initial criticisms, affirming “decisions to change or replace a classification system should be based on the results of scientific competition (e.g., tests of incremental) validity” (2022, 288), which they judge HiTOP to be incapable of producing.

Setting aside that neither group provides an account of validity, this version of a back and forth between frameworks is compromised since all parties are operating on entirely different standards as well as a fundamental conception for what counts as validity. As a result, we cannot presently adjudicate between approaches and conclude as Haefel et al. (2022) do that one approach is less valid than the other. Similar to how Zachar, Krueger, and Kendler describe the disparate approaches to the classification of personality disorders in DSM-5, they are “like two ships passing in the night” (2016, 5).

5.2. Difficulties coordinating and integrating between frameworks

A second implication is the prevention of coordinating and integrating between differing approaches. Despite each approach maintaining that their own framework is the best, there have been several proposals to suggest how certain approaches may be successfully linked. For example, Michelini et al.’s (2021) “Linking RDoC and HiTOP: A New Interface for Advancing Psychiatric Nosology and Neuroscience” suggests RDoC, with its biologically based focus, may provide helpful tools for “elucidating the underpinnings of the clinical problems in HiTOP” (2021, 4), whereas HiTOP may motivate RDoC studies “by providing psychometrically robust clinical targets” (2021, 4).

Any interface between HiTOP and RDoC assumes the potential for achieving *construct stabilization*. Construct stabilization, defined by Sullivan (2016a, 2016b), is the coordination across researchers situated in the same and different approaches to come to agreements regarding (i) how to generally define terms designating constructs, (ii) the best experimental paradigms for studying a given construct, and (iii) the conditions under which two experimental paradigms can be said to measure the same construct. Coordination and integration between frameworks are thought to be contingent on the ability to stabilize constructs.

The problem is, given distinct conceptions of validity, conditions (ii) and (iii) of construct stabilization cannot be met. Condition (ii) cannot be met because both HiTOP (structure-first) and RDoC (biology-first) maintain that their own approach informs the best experimental paradigms and, given these approaches are so disparate, they ultimately will not coalesce on specific paradigms. Condition (iii) cannot be met since differing standards for what is required for validation dictate diverging standards as to acceptable ways to measure HiTOP or RDoC constructs.

Perhaps the best argument against coordination and integration is that RDoC and HiTOP constructs do not even refer to the same thing. HiTOP constructs are understood as *pure* constructs, i.e., data-driven factors based on statistical output from factor analysis of psychological testing data. For RDoC, constructs are empirical functions selected via a process of expert curation and are thought to refer to specific biological and cognitive processes (e.g., Reward Learning). Thus, even if researchers were to come together to discuss what they mean by “construct,” without

dramatically overhauling their entire approach, construct stabilization across frameworks, and thus coordination or integration, isn't currently achievable.

5.3. Difficulties in establishing a unified concept of validity

A third implication is that the approaches will be unable to establish a unified concept of validity. Take the broad concept of depression for which researchers wish to develop valid psychiatric classifications. Ultimately, the presence of disparate conceptions of depression amounts to a troubling scenario for trying to do any kind of unified validation. The central issue is that the approaches don't agree on the phenomenon they're trying to make inferences about. If they don't agree on what the features or attributes are, progress in achieving a unified conception of validity for psychiatric classifications is incredibly difficult due to its preceding influence on each approach's standards of validation.

A similar problem arises in psychological testing, whereby if multiple testers are attempting to design their own inventory to measure "intelligence," if there is a significant enough difference in the understanding of what "intelligence" amounts to, then developing and comparing valid tests of "intelligence" becomes incredibly difficult. Those who view intelligence in one way will not just want but need to pursue one conception of validity with specific standards. In contrast, another approach, finding the original approach misguided will reject it completely and insist on pursuing validation in a different manner.

Thus, starting with disparate conceptions of psychiatric disorders as is the case across DSM, HiTOP, the Network Approach, and RDoC influences the development of divergent standards for validating psychiatric classifications. How each approach conceptualizes and represents the phenomenon necessitates how its researchers will study, evaluate, and make inferences concerning validating evidence, ultimately shaping any conception of validity achieved.

6. Resolving psychiatry's new validity crisis

Initially, there was hope that a plurality of approaches would lead us closer to achieving a unified conception of validity in psychiatry. However, to borrow an old metaphor from a conversation with validity theorist Gregory Cizek, you can try to mix oil and water, but ultimately, you don't have a solution. You can shake them up, and they'll sometimes appear to *sort of* come together. But upon closer examination, what you have is not truly a solution, but multiple sources that end up separating because they cannot be combined.

Such may be the case with the Holy Quadrinity of validity in psychiatry and the problem of disparate validation. The way each approach understands validity simply cannot be easily synthesized with the way another approach understands validity. Given each approach adopts its own conception, the "best approach" is not something evidence alone can tell us. We face the realization that (i) there does not exist a single validation approach by which frameworks may be validated against, and (ii) no single approach will be capable of achieving a unified conception of validity across frameworks, meaning each approach may be considered valid only under its own conception.

In addition, we see that the extent of a new validity crisis is likely the tip of the iceberg. Beyond validity, there is a much more extensive tangle of both constitutive and prescriptive standards within these frameworks, such as what counts as “progress,” what counts as “scientific,” what counts as an “explanation,” what counts as a “testable theory,” and finally, what counts as a “disorder” or “feature of psychopathology.” As a result, the problem is not only that disparate conceptions of validity result in disagreements as to what the current validity problem even is, but potential collaboration is further complicated by an extended list of additional inseparable standards. We can’t simply disentangle one set of standards, such as “validity” and “validation,” without impacting the others.

To move the field forward, I offer a few recommendations. First, we should abandon hope of developing a unified conception of validity. Psychiatry, instead, should embrace *interactive pluralism* (Van Bouwel 2014), which permits the plurality of distinct approaches with a plurality of distinct conceptions of validity to engage with one another, but without the presumption this will lead to a consensus or integration. This is separate from current attitudes in psychiatry advocating for (i) integrative pluralism, which suggests these disparate approaches may work together to inform valid psychiatric classifications such that they integrate their disparate findings concerning the phenomenon within an integrated conception of validity; or (ii) isolationist pluralism, which presupposes that interactions between disparate frameworks cannot be productive. Under interactive pluralism, each framework should be tasked with continuing with its own specific conception of validity while openly critiquing and challenging one another’s approaches in a productive way, promoting the rigorous testing and refinement of each framework’s conception of validity.

By embracing interactive pluralism, we can think of the distinct conceptions of validity as if they are different languages. While languages differ in many ways, people can still be taught and learn to speak more than one. To this end, I recommend each framework stabilize its own validation principles and procedures, which at present remain underspecified and inconsistent. Stabilizing should include, per Sullivan (2016b), defining the terms, the best experimental paradigms, and the conditions for measurement, through workshops and symposiums that bring together researchers from the various frameworks to engage in structured dialogues about their validity conceptions. Stabilizing should also expand beyond Sullivan’s recommendations and consider the scientific aims, standards for assuring consistency in application, and the epistemic and non-epistemic values that inform decision processes. Furthermore, there should be a willingness from each approach to engage with more fundamental questions of measurement, such as how validation methods may connect with the constructs they are trying to measure (Maul 2017).

Lastly, I recommend psychiatry strongly reconsider the high epistemic value its frameworks currently place on the concept of validity. Only once we’ve begun to stabilize validity within each framework—allowing psychiatry to get a hold on its disparate validation procedures—may validity have its special status returned.

As a provisional attempt to integrate across disparate concepts of validity, I recommend developing and utilizing standards of utility. Utility is collectively thought of as a “graded characteristic that is partly context specific” (Kendell and Jablensky 2003, 9) and more broadly applicable as the degree to which a psychiatric classification may predict “course, outcome, and likely response to available

treatments, even if their inner biological and psychological structure is not fully understood” (Jablensky 2016, 26). As it turns out, just as the concept of utility was all-important for diagnostic validity within the original Kraepelinian formulation and in that of the DSM, each of the other conceptions of validity across HiTOP, the Network Approach, and RDoC also align with the Kraepelinian tradition of treating utility as a primary validator. To this end, I recommend a moderate convergentist approach, which implies we may partially reduce utility to a shared conception that focuses on shared similarities between frameworks. Since most of the frameworks return to the original validators of Robins and Guze, which Solomon (2022) has characterized as *utilitators* for their ability to serve double duty as sources of evidence to establish both validity and utility, I suggest such utilitators as a starting point for developing *shared* utility standards based on compromises between frameworks.

An immediate challenge is, given the presence of disparate conceptions of utility, it is not immediately clear how the four frameworks would come to any consensus. However, given utility’s context-specific nature, there may be a greater opportunity for negotiation as to what is useful *within the context of different frameworks* in the context of patient care, in a sort of “good for you, not or me” approach. Thus, I recommend the development of a shared utility framework, which representatives from each approach collaboratively construct, that delineates shared criteria. Such criteria could be utilized for drawing immediate practical comparisons relevant for the clinic while psychiatry works to address its new validity crisis.

Acknowledgements. I would like to thank Peter Zachar and the anonymous referees for their detailed comments on the article. This article also benefited greatly from feedback from Kirk Ludwig, Dan Kennedy, Gary Ebbs, Jordi Cat, Evan Arnet, Siyu Yao, Dan Li, and participants of the 2024 Philosophy of Social Science Roundtable and of the 2024 Association for the Advancement of Philosophy and Psychiatry Meeting.

References

- Aboraya, Ahmed, Cheryl France, John Young, Kristina Curci, and James LePage. 2005. “The Validity of Psychiatric Diagnosis Revisited: The Clinician’s Guide to Improve the Validity of Psychiatric Diagnosis.” *Psychiatry (Edgmont)* 2 (9):48.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. *Standards for Educational and Psychological Testing*. Lanham, MD: American Educational Research Association.
- American Psychiatric Association. 2022. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TR*.
- American Psychiatric Association. n.d. “Submit Proposals for Making Changes to DSM-5-TR.” <https://www.psychiatry.org/psychiatrists/practice/dsm/submit-proposals>.
- Borsboom, Denny. 2008. “Psychometric Perspectives on Diagnostic Systems.” *Journal of Clinical Psychology* 64 (9):1089–108. <https://doi.org/10.1002/jclp.20503>.
- Borsboom, Denny. 2017. “A Network Theory of Mental Disorders.” *World Psychiatry* 16 (1):5–13. <https://doi.org/10.1002/wps.20375>.
- Bringmann, Laura F., Casper Albers, Claudi Bockting, Denny Borsboom, Eva Ceulemans, Angélique Cramer, Sacha Epskamp, Markus I. Eronen, Ellen Hamaker, Peter Kuppens, Wolfgang Lutz, Richard J. McNally, Peter Molenaar, Pia Tio, Manuel C. Voelke, and Marieke Wichers. 2022. “Psychopathological Networks: Theory, Methods and Practice.” *Behaviour Research and Therapy* 149: 104011. <https://doi.org/10.1016/j.brat.2021.104011>.
- Cizek, Gregory J. 2020. *Validity: An Integrated Approach to Test Score Meaning and Use*. London: Routledge.
- Cuthbert, Bruce N. 2020. “The Role of RDoC in Future Classification of Mental Disorders.” *Dialogues in Clinical Neuroscience* 22 (1):81–85. <https://doi.org/10.31887/dcns.2020.22.1/bcuthbert>.

- DeYoung, Colin G., Roman Kotov, Robert F. Krueger, David C. Cicero, Christopher C. Conway, Nicholas R. Eaton, Miriam K. Forbes, Michael N. Hallquist, Katherine G. Jonas, Robert D. Latzman, Craig Rodriguez-Seijas, Camilo J. Ruggero, Leonard J. Simms, Irwin D. Waldman, Monika A. Waszczuk, Thomas A. Widiger, and Aidan G. C. Wright. 2021. "Answering Questions About the Hierarchical Taxonomy of Psychopathology (HiTOP): Analogies to Whales and Sharks Miss the Boat." *Clinical Psychological Science* 10 (2): 279–84. <https://doi.org/10.1177/21677026211049390>.
- Forbes, Miriam K., Whitney R. Ringwald, Timothy Allen, David C. Cicero, Lee Anna Clark, Colin G. DeYoung, Nicholas Eaton, et al. 2024. "Principles and Procedures for Revising the Hierarchical Taxonomy of Psychopathology." *Journal of Psychopathology and Clinical Science* 133 (1): 4–19. <https://doi.org/10.1037/abn0000886>.
- Gagné-Julien, Anne-Marie. 2021. "Towards a Socially Constructed and Objective Concept of Mental Disorder." *Synthese* 198 (10):9401–26. <https://doi.org/10.1007/s11229-020-02647-7>.
- Haefel, Gerald J., Bertus F. Jeronimus, Bonnie N. Kaiser, Lesley Jo Weaver, Peter D. Soyster, Aaron J. Fisher, Ivan Vargas, Jason T. Goodson, and Wei Lu. 2021. "Folk Classification and Factor Rotations: Whales, Sharks, and the Problems With the Hierarchical Taxonomy of Psychopathology (HiTOP)." *Clinical Psychological Science* 10 (2):259–78. <https://doi.org/10.1177/21677026211002500>.
- Haefel, Gerald J., Bertus F. Jeronimus, Aaron J. Fisher, Bonnie N. Kaiser, Lesley Jo Weaver, Ivan Vargas, Jason T. Goodson, Peter D. Soyster, and Wei Lu. 2022. "The Hierarchical Taxonomy of Psychopathology (HiTOP) Is Not an Improvement Over the DSM." *Clinical Psychological Science* 10 (2):285–90. <https://doi.org/10.1177/21677026211068873>.
- Hartman, Catharina A., Joop Hox, Gideon J. Mellenbergh, Michael H. Boyle, David R. Offord, Yvonne Racine, Jane McNamee, Kenneth D. Gadow, Joyce Sprafkin, Kevin L. Kelly, Edith E. Nolan, Rosemary Tannock, Russell Schachar, Harry Schut, Ingrid Postma, Rob Drost, and Joseph A. Sergeant. 2001. "DSM-IV Internal Construct Validity: When a Taxonomy Meets Data." *Journal of Child Psychology and Psychiatry* 42 (6):817–36. <https://doi.org/10.1111/1469-7610.00778>.
- Jablensky, Assen. 2016. "Psychiatric Classifications: Validity and Utility." *World Psychiatry* 15 (1):26–31. <https://doi.org/10.1002/wps.20284>.
- Kendell, Robert E. 2002. "Five Criteria for an Improved Taxonomy of Mental Disorders." In *Defining Psychopathology in the 21st Century: DSM-V and Beyond*, edited by John E. Helzer and James J. Hudziak, 3–17. Washington, DC: American Psychiatric Publishing, Inc.
- Kendler, Kenneth S., and Michael B. First. 2010. "Alternative Futures for the DSM Revision Process: Iteration v. Paradigm Shift." *The British Journal of Psychiatry* 197 (4):263–265. <https://doi.org/10.1192/bjp.bp.109.076794>.
- Kendell, Robert, and Assen Jablensky. 2003. "Distinguishing Between the Validity and Utility of Psychiatric Diagnoses." *American Journal of Psychiatry* 160 (1):4–12. <https://doi.org/10.1176/appi.ajp.160.1.4>.
- Kendler, Kenneth, David Kupfer, William Narrow, Katharine Phillips, and Jan Fawcett. 2009. "Guidelines for Making Changes to DSM-V." Unpublished manuscript.
- Keshavan, Matcheri S., Brett A. Clementz, Godfrey D. Pearlson, John A. Sweeney, and Carol A. Tamminga. 2013. "Reimagining Psychoses: An Agnostic Approach to Diagnosis." *Schizophrenia Research* 146 (1–3):10–16. <https://doi.org/10.1016/j.schres.2013.02.022>.
- Knox, Bennett. 2022. "Exclusion of the Psychopathologized and Hermeneutical Ignorance Threaten Objectivity." *Philosophy, Psychiatry & Psychology* 29 (4):253–66. <https://doi.org/10.1353/ppp.2022.0044>.
- Kotov, Roman, Robert F. Krueger, David Watson, Thomas M. Achenbach, Robert R. Althoff, R. Michael Bagby, Timothy A. Brown, et al. 2017. "The Hierarchical Taxonomy of Psychopathology (HiTOP): A Dimensional Alternative to Traditional Nosologies." *Journal of Abnormal Psychology* 126 (4):454–77. <https://doi.org/10.1037/abn0000258>.
- Livesley, W. John, and Douglas N. Jackson. 1992. "Guidelines for Developing, Evaluating, and Revising the Classification of Personality Disorders." *The Journal of Nervous and Mental Disease* 180 (10):609–18. <https://doi.org/10.1097/00005053-199210000-00001>.
- Lynch, Terry. 2018. "The Validity of the DSM: An Overview." *The Irish Journal of Counselling and Psychotherapy* 18 (2):5–10.
- Markus, Keith A. 2015. "Alternative Vocabularies in the Test Validity Literature." *Assessment in Education: Principles, Policy, and Practice* 23 (2):252–67. <https://doi.org/10.1080/0969594x.2015.1060191>.

- Maul, Andrew. 2017. "Rethinking Traditional Methods of Survey Validation." *Measurement Interdisciplinary Research and Perspectives* 15 (2):51–69. <https://doi.org/10.1080/15366367.2017.1348108>.
- Messick, Samuel. 1989. "Validity." In *Educational Measurement*, edited by Robert L. Linn, 13–103. New York: Macmillan.
- Micheline, Giorgia, Isabella M. Palumbo, Colin G. DeYoung, Robert D. Latzman, and Roman Kotov. 2021. "Linking RDoC and HiTOP: A New Interface for Advancing Psychiatric Nosology and Neuroscience." *Clinical Psychology Review* 86: 102025. <https://doi.org/10.1016/j.cpr.2021.102025>.
- Newton, Paul E., and Jo-Anne Baird. 2016. "The Great Validity Debate." *Assessment in Education Principles Policy and Practice* 23 (2):173–77. <https://doi.org/10.1080/0969594x.2016.1172871>.
- Newton, Paul E., and Stuart D. Shaw. 2015. "Disagreement Over the Best Way to Use the Word 'Validity' and Options for Reaching Consensus." *Assessment in Education Principles Policy and Practice* 23 (2):178–97. <https://doi.org/10.1080/0969594x.2015.1037241>.
- Phillips, James. 2013. "The Conceptual Status of DSM-5 Diagnoses." In *Making the DSM-5: Concepts and Controversies*, edited by Joel Paris and James Phillips, 143–57. New York: Springer. https://doi.org/10.1007/978-1-4614-6504-1_10.
- Poland, Jeffrey, and Şerife Tekin, eds. 2017. *Extraordinary Science and Psychiatry: Responses to the Crisis in Mental Health Research*. Cambridge, MA: MIT Press.
- Robins, Eli, and Samuel B. Guze. 1970. "Establishment of Diagnostic Validity in Psychiatric Illness: Its Application to Schizophrenia." *American Journal of Psychiatry* 126 (7):983–87. <https://doi.org/10.1176/ajp.126.7.983>.
- Schaffner, Kenneth F. 2012. "A Philosophical Overview of the Problems of Validity for Psychiatric Disorders." In *Philosophical Issues in Psychiatry II: Nosology*, edited by Kenneth S. Kendler and Josef Parnes, 169–89. Oxford: Oxford University Press. <https://doi.org/10.1093/med/9780199642205.003.0026>.
- Slaney, Kathleen. 2017. *Validating Psychological Constructs: Historical, Philosophical, and Practical Dimensions*. London: Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-38523-9>.
- Solomon, Miriam. 2022. "On Validators for Psychiatric Categories." *Philosophy of Medicine* 3 (1):1–23. <https://doi.org/10.5195/pom.2022.74>.
- Sullivan, Jacqueline Anne. 2016a. "Construct Stabilization and the Unity of the Mind-Brain Sciences." *Philosophy of Science* 83 (5):662–73. <https://doi.org/10.1086/687853>.
- Sullivan, Jacqueline Anne. 2016b. "Stabilizing Constructs Through Collaboration Across Different Research Fields as a Way to Foster the Integrative Approach of the Research Domain Criteria (RDoC) Project." *Frontiers in Human Neuroscience* 10: 309. <https://doi.org/10.3389/fnhum.2016.00309>.
- Tekin, Şerife. 2022. "Participatory Interactive Objectivity in Psychiatry." *Philosophy of Science* 89 (5): 1166–75. <https://doi.org/10.1017/psa.2022.47>.
- Tsou, Jonathan Y. 2021. *Philosophy of Psychiatry*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108588485>.
- Van Bouwel, Jeroen. 2014. "Pluralists About Pluralism? Different Versions of Explanatory Pluralism in Psychiatry." In *New Directions in the Philosophy of Science*, edited by Maria Carla Galavotii, Dennis Dieks, W. J. Gonzalez, Stephan Hartmann, Thomas Uebel, and Marcel Weber, 105–19. Cham: Springer. https://doi.org/10.1007/978-3-319-04382-1_8.
- Wakefield, Jerome C. 2013. "The DSM-5 Debate Over the Bereavement Exclusion: Psychiatric Diagnosis and the Future of Empirically Supported Treatment." *Clinical Psychology Review* 33 (7):825–45. <https://doi.org/10.1016/j.cpr.2013.03.007>.
- Zachar, Peter. 2012. "Progress and the Calibration of Scientific Constructs: The Role of Comparative Validity." In *Philosophical Issues in Psychiatry II: Nosology*, edited by Kenneth S. Kendler and Josef Parnes, 21–34. Oxford: Oxford University Press. <https://doi.org/10.1093/med/9780199642205.003.0005>.
- Zachar, Peter, Robert F. Krueger, and Kenneth S. Kendler. 2016. "Personality Disorder in DSM-5: An Oral History." *Psychological Medicine* 46 (1):1–10. <https://doi.org/10.1017/s0033291715001543>.

Cite this article: Zautra, Nicholas. 2025. "Psychiatry's New Validity Crisis: The Problem of Disparate Validation." *Philosophy of Science* 92 (3):646–665. <https://doi.org/10.1017/psa.2024.71>