

DSM disorders and their criteria: how should they inter-relate?

K. S. Kendler*

Departments of Psychiatry, and Human and Molecular Genetics, Virginia Institute of Psychiatric and Behavioral Genetics, Medical College of Virginia/Virginia Commonwealth University, Richmond, VA, USA

While the changes in psychiatric diagnosis introduced by Diagnostic and Statistical Manual third edition (DSM-III) have had major benefits to the field of psychiatry, the reification of its diagnostic criteria and the widespread adoption of diagnostic literalism have been problematic. I argue that, at root, these developments can be best understood by contrasting two approaches to the relationship between DSM disorders and their criteria. In a *constitutive* relationship, criteria definitively define the disorder. Having a disorder is nothing more than meeting the criteria. In an *indexical* relationship, the criteria are fallible indices of a disorder understood as a hypothetical, tentative diagnostic construct. I trace the origins of the constitutive model to the philosophical theory of operationalism. I then examine a range of historical and empirical results that favor the indexical over the constitutive position including (i) evidence that individual criteria for DSM-III were selected from a broader pool of possible symptoms/signs, (ii) revisions of DSM have implicitly assumed an indexical criteria-disorder relationship, (iii) the indexical position allows DSM criteria to be wrong and misdiagnose patients while such a result is incoherent for a constitutive model, an implausible position, (iv) we assume an indexical criteria-scale relationships for many personality and symptom measures commonly used in psychiatric practice and research, and (v) empirical studies suggesting similar performance for DSM and non-DSM symptoms for major depression. I then review four reasons for the rise of the constitutive position: (i) the 'official' nature of the DSM criteria, (ii) the strong investment psychiatry has had in the DSM manual and its widespread use and success, (iii) lack of a clear pathophysiology for our disorders, and (iv) the absence of informative diagnostic signs of minimal clinical importance. I conclude that the constitutive position is premature and reflects a conceptual error. It assumes a definitiveness and a literalism about the nature of our criteria that is far beyond our current knowledge. The indexical position with its tentativeness and modesty accurately reflects the current state of our field.

Received 6 January 2017; Revised 17 February 2017; Accepted 28 February 2017; First published online 4 April 2017

Key words: Diagnostic and Statistical Manual, operationalism, philosophy, psychiatric diagnosis, psychometrics.

Since its beginning as a discrete field of medicine in the nineteenth century, psychiatry has been preoccupied by nosologic questions. Indeed, much of its history is taken up by debates about the nature and content of a wide range of proposed psychiatric diagnostic systems. The publication of Diagnostic and Statistical Manual third edition (DSM-III) in 1980 represented a watershed moment in that history (Decker, 2013). Its many benefits, especially improved standardization and reliability of the diagnostic process, have been widely documented. However, the immense popularity of the DSM-III and its subsequent revisions, along with its widespread use in other mental health fields and legal circles, has produced developments

unanticipated by its originators: reification and its closely inter-related diagnostic literalism (Hyman, 2010).

I here argue that the debate over reification of the DSM criteria can be best understood as arising from a disagreement about the nature of the relationship between the disorders contained in the DSM manual and the criteria proposed for them. Is the relationship an *indexical* one or a *constitutive* one? If the relationship is indexical, then the criteria index or reflect the disorder which is, in turn, is understood as a *hypothetical diagnostic construct*. If the relationship is constitutive, then the criteria simply are the disorder – they constitute it so that having a particular disorder is nothing more than meeting the relevant DSM criteria. That is, the criteria can be interpreted literally, as definitive and authoritative.

This essay has five parts. First, I present definitions of the related terms of reification and diagnostic literalism. Second, I review the historical underpinnings of the constitutive position in the philosophy of science

* Address for correspondence: K. S. Kendler, M.D., Departments of Psychiatry, and Human and Molecular Genetics, Virginia Institute of Psychiatric and Behavioral Genetics Medical College of Virginia/Virginia Commonwealth University, Box 980126, Richmond, VA 23298-0126, USA.

(Email: kenneth.kendler@vcuhealth.org)

theory of *operationalism*. I then clarify the important differences between operationalism as a philosophical position and the practical use of operationalized criteria in DSM-III and its immediate forebears [Feighner *et al.* (1972) and Research Diagnostic Criteria (RDC) (Spitzer *et al.* 1975)]. Third, I present historical and empirical data that help evaluate the indexical and constitutive positions for DSM-III. Fourth, I review the social and historical forces that have contributed to the constitutive position and the rise of reification and diagnostic literalism. Fifth, I conclude the essay by reviewing the psychometric measurement theories that underlie the indexical position and then review how the constitutive and indexical positions address the underlying nature of our DSM categories. I argue that the constitutive position – that DSM criteria constitute the disorders they were designed to assess – reflects a serious conceptual error. The field of psychiatry and those of other mental health disciplines that use the DSM manuals would be well advised to teach and implement the humbler and more accurate indexical view of the DSM criteria.

Reification and diagnostic literalism

Derived from the Latin, reification literally means ‘thing-making’ and a generic definition might be ‘taking something abstract, theoretical or tentative to be concrete, to be ‘real.’’ Here are three relevant quotes from famous philosophers:

Another great abuse of words is taking them for things. . . . How much names taken for things are apt to mislead the understanding. Locke [(1979) pp. 442–443].

The tendency has always been strong to believe that whatever received a name must be an entity or being, having an independent existence of its own. Mill (1843).

There is an error; but it is merely the accidental error of mistaking the abstract for the concrete. It is an example of what I will call the ‘Fallacy of Misplaced Concreteness.’ Whitehead (1925) p. 51.

Diagnostic literalism, as articulated by Peter Zachar, is a related concept, which he describes as ‘misplaced literalism about diagnostic constructs.’ [(Zachar, 2014) pp. 73–83]. In this context, literalism means the attribution of excess authority or definitiveness to the written word especially when the text derives from an expert or from some source considered to be ‘official.’

Operationalism

The assumptions of the constitutive view of DSM criteria can be traced back to operationalism, a theory of measurement in science closely associated with the American physicist and Nobelist, PW Bridgman

(1882–1961) (Chang, 2009). The theory is based on the reasonable intuition that a concept is not meaningful if we lack a way to measure it. Bridgman summarizes his position succinctly as follows, where ‘operations’ refers to methods of measurement

In general, we mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations [(Bridgman, 1927) p. 5].

In more common parlance, he is saying that a scientific variable – be it mass, temperature or (in our case) a psychiatric disorder – is nothing more than the methods by which we propose to measure it. Bridgman’s work was motivated by attempts to understand Einstein’s theory of special relativity and how it appeared to redefine critical constructs in physics such as time, distance and simultaneity. A range of objections was raised to Bridgman’s operationalism [see (Chang, 2009) for a succinct review], two of which are relevant here. First, critics argued that if this approach is fully adopted, the concept of validation of a measure is tautological. Revision would not make sense because by changing the measurement of a concept, you would change its meaning. It could no longer be the same concept. Second, others argued that there were quite useful scientific concepts that were not yet ready for operationalization. Concepts in science often begin as hypothetical constructs that cannot be fully specified. But, they can undergo an iterative process of improved understanding and better measurement that might, at some later date, permit a full specification with fully developed rules of measurement. This flexibility would be impossible within a strict theory of operationalism.

Outside of physics, operationalism was especially influential in behaviorist psychology, which sought to develop an objective science by eliminating mental elements in psychological constructs. For example, in rodent experiments, hunger could be defined as ‘time since last feeding.’ [(Chang, 2009) p. 14].

One of the major elements in the ‘DSM-III revolution’ was the adoption, following in the footsteps of the Feighner and RDC systems, of *operationalized* diagnostic criteria. Did that mean that the pioneers of these efforts, individuals such as Robins, Guze and Spitzer, were committed to the philosophical position of operationalism? The historical record here is clear. They were all in favor of the benefits of *practical* operationalism – giving psychiatrists rules by which to assemble symptoms and signs into diagnoses as a means of improving reliability. But these developments carried no commitments to a Bridgman-style *philosophical* operationalism. Both Rodrigo Muñoz – one of the original team who developed the Feighner Criteria – and Robert Spitzer, in personal communications, denied that the development of

the Feighner criteria, RDC and DSM-III were directly influenced by philosophical ideas about measurement, as articulated in particular by such figures as Bridgman and Hempel (1965). Muñoz recalled that Sam Guze – with his training in internal medicine – likely first got the idea about operationalized criteria from his familiarity with the operationalized criteria for rheumatologic disorders that were published and validated as far back as the 1940s (Jones, 1944; Ropes *et al.* 1957; Kendler *et al.* 2010).

Historical and empirical data relevant to the indexical and constitutive positions

Taking as an initial example, the origins of the DSM-III diagnostic criteria for major depression (MD) are relatively well understood [(Kendler *et al.* 2010) Table 1]. They derive, with minimal changes, from those proposed for the Research Diagnostic Criteria (Spitzer *et al.* 1975), which in turn were based, with modest modifications, on those included in the Feighner Criteria (Feighner *et al.* 1972). The Feighner criteria for MD were adapted from an earlier item set of Cassidy *et al.* (1957). Finally, Cassidy *et al.* cite as a key source a set of criteria for MD previously proposed by Stone & Burris (1950). Some of the differences across these criteria sets are illuminating. Cassidy *et al.* (1957) included slowed thinking, decreased libido and constipation, none of which were included in DSM-III. DSM-III added worthlessness, which was not present in the earlier diagnostic formulations, and added appetite/weight gain, not present in the Cassidy *et al.* (1957), or Stone & Burris (1950) criteria.

Examining textbook descriptions of the depressive syndrome from 1900 to 1960, a recent review provided a broader historical context within which to view the DSM criteria for MD (Kendler, 2016b). Of the 18 depressive symptoms and signs frequently noted by the textbook authors, ten were well covered by DSM-III MD criteria, two were only partly covered and six were entirely absent (Kendler, 2016b). For example, the historical experts noted that symptoms of anxiety occurred frequently in depression but were not included in the DSM-III criteria. In describing the common cognitive changes in depression, the textbooks noted a range of symptoms including hopelessness, pessimism and feelings of inadequacy, symptoms not well captured by the single relevant DSM-III criterion of guilt and worthlessness. A methodologically similar review was also recently completed for schizophrenia where 20 symptoms and signs were frequently reported by the historical authors, only eight of which were included in the DSM-III criteria (Kendler, 2016a). Common or conceptually important symptoms or signs for schizophrenia not present in DSM-III

included volitional changes, abnormal movements or posture, autism, negativism, and un-understandability.

These historical inquiries suggest the specific criteria selected for MD and schizophrenia for the DSM-III reflect one set from a broader number of possible criteria, which could have been chosen. The selection process for DSM involved precedent, clinical wisdom, research findings, and opinions of particular participants. Bob Spitzer's work on DSM-III was strongly influenced by his contact with Eli Robins and Sam Guze at Washington University, which arose as a result of the NIMH Collaborative Study of Depression (Kendler *et al.* 2010). The bereavement exclusion criterion for MD in DSM-III, which had no precedent in the Feighner or RDC criteria, was proposed by Dr Paula Clayton who served on the relevant DSM committee and had by then a long history of investigating the psychiatric sequelae of bereavement (P. Clayton, personal communication, 2016) (Clayton *et al.* 1968, 1974).

The constitutive view of DSM criteria would need to argue that there is a fundamental difference between the symptoms and signs selected for the DSM-III criteria and those historically important criteria that were excluded. Because the constitutive position defines disorders to be nothing more than their criteria, the excluded criteria would have to have a quite different status. By contrast, the indexical position would understand that diagnostic criteria are chosen from a larger possible set of relevant symptoms and signs. The goal of an indexical function is to maximize accuracy (good sensitivity and specificity) and utility. That there would be differing ideas about the optimal set of criteria would be expected.

In DSM-III-R, generalized anxiety disorder (GAD) included 18 symptomatic criteria of which six or more needed to be met for the diagnosis. With the goal of improving utility, in DSM-IV, the 18 criteria were reduced to six of which three or more needed to be endorsed (Abel & Borkovec, 1995). How would such a change be interpreted from constitutive and indexical positions? The constitutive position would conclude that this represented a major change in the concept of GAD given the number and nature of the symptoms required in the two DSM editions differed substantially. If the construct was temperature, this would represent making a new and different kind of thermometer. The indexical position, by contrast, would assume these two sets of items were both fallible indices of a single unmeasured construct. Such a position would be reinforced by empirical evidence, soon provided (Abel & Borkovec, 1995) that these two criteria sets, although objectively rather different, actually largely identified the same group of patients. As noted above, one of the objections to the strict operationalism of Bridgman was that it made revisions of

Table 1. Items to assess the personality trait of neuroticism from three short scales

Item number/ scale	Big five inventory (John <i>et al.</i> 1991)	EPQ short form (Eysenck <i>et al.</i> 1985)	Short IPIP-BFM (Skimina <i>et al.</i> 2014)
1	Is depressed, blue	Does your mood often go up and down?	Have frequent mood swings
2	Is relaxed, handles stress well ^a	Do you ever feel 'just miserable' for no reason?	Change my mood a lot
3	Can be tense	Are you an irritable person?	Worry about things
4	Worries a lot	Are your feelings easily hurt?	Often feel blue
5	Is emotionally stable, not easily upset ^a	Do you often feel 'fed-up'?	Seldom feel blue ^a
6	Is depressed, blue	Would you call yourself a nervous person?	Am easily disturbed
7	Can be moody	Are you a worrier?	Get irritated easily
8	Remains calm in tense situations ^a	Would you call yourself tense or 'highly strung'?	Get upset easily
9	Gets nervous easily	Do you worry too long after an embarrassing experience?	Get stressed out easily
10		Do you suffer from 'nerves'?	Am relaxed most of the time ^a
11		Do you often feel lonely?	
12		Are you often troubled about feelings of guilt?	

IPIP-BFM, International Personality Item Pool-Big Five Markers.

^a Reverse coded.

the measurement constructs difficult if not nonsensical. If a scientific construct, be it a psychiatric disorder or a physical property like weight, is defined by a certain set of operations, when would it ever make sense to change those operations? The revision processes for the DSM manuals have implicitly assumed an indexical and not constitutive view of the relationship of DSM disorders to their criteria.

How do the indexical and constitutive positions respond to diagnostic problems that might arise from the application of DSM criteria? Assume you see a quite depressed patient. However, a careful examination reveals that her symptoms and signs do not meet DSM criteria. She describes prominent hopelessness but not worthlessness, guilt or suicidal ideation. Her active libido has disappeared but her appetite is OK and she has not lost weight. Her interests have not diminished but her sad mood is quite unresponsive to her usual pleasures in life and caused a deep sense of derealization. From the constitutive approach, if you have applied the criteria correctly, you would need to conclude that she does not have MD. From an indexical perspective, however, you could argue she really has MD but the criteria have malfunctioned. This situation would be analogous to the teacher who worked all year with a promising student and found her to be quite bright and insightful. However, to her surprise, a standard IQ test put the student in the normal range and the student was well and alert when the test was given. A constitutive view of intelligence

(it just is what the IQ test says) would say the teacher is wrong. The indexical position would be that it is perfectly plausible that the test result is wrong and the teacher has a more valid sense of the student's intelligence than the test does.

We can further clarify the indexical and constitutive positions for DSM-III by examining two quantitative traits often used in psychiatric research: neuroticism and depressive symptoms. Table 1 depicts the individual items used in three short, publically available neuroticism scales. Item content shows substantial resemblance. Terms like irritable, nervous, worried, blue, upset, and tense occur in the items from all three scales. The constitutive position would be that these scales define three different personality constructs. You could take your pick but in each case, having high neuroticism is defined by endorsing a large proportion of items on that scale. The indexical position would posit the existence of a hypothetical personality construct called neuroticism that these three scales each, in different and likely imperfect ways, attempt to assess.

Fried (2016) recently examined the item content of seven popular depressive symptom rating scales. These scales contain 52 different symptoms. The mean overlap among all scales was rather low (0.36) and 40% of symptoms appeared in only a single scale. Here are a few items that appeared in all scales: sad mood and decreased appetite; in three scales: crying and hypersomnia; and in only one scale: talking less and self-dislike. Do we wish, as would be required from a constitutive perspective, to

assume seven different manifest constructs of depressive symptoms each of which is definitively defined by its specific items? Or is it more plausible to suggest one or a small number of latent constructs of depressive symptoms which each of these seven scales are, with their imperfections, trying to index?

Finally, two empirical studies relevant to our discussion have recently compared the function of DSM and non-DSM criteria for MD. Both performed network analyses of large samples of depressed patients meeting DSM-IV criteria and both utilized a concept of 'centrality', which reflects the degree of inter-connection between individual items in the network and all the other criteria. Fried et al examined 28 depressive symptoms assessed by the Inventory of Depressive Symptomatology (IDS) (Rush *et al.* 1996) in 3463 depressed outpatients from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study (Fava *et al.* 2003). Of the 28 IDS items, 15 derived from DSM MD criteria and 13 did not. The two sets of criteria were intermingled within a single network and the centrality/inter-connectedness of the DSM and non-DSM items did not differ significantly. A second study applied the same methodology to all nine of the DSM-IV depressive criteria and 11 other depressive symptoms chosen for their research value. They were all assessed at personal interview in 6008 depressed Han Chinese women, ascertained in psychiatric treatment facilities (CONVERGE consortium, 2015) (Kendler *et al.*, unpublished results). Again, the DSM and non-DSM depressive symptoms were intermingled within a single network with no obvious qualitative differences between them. Furthermore, on average, the DSM criteria were no more central to the network than the non-DSM symptoms. Interestingly, some of the items noted by earlier textbook authors (e.g. hopelessness) or included in the Feighner and earlier depressive criteria (low libido), were more central than the related DSM criteria (e.g. worthlessness and changes in appetite, respectively). While we should not over-interpret these results as they utilized only one statistical method, they support the historical inquires that the DSM criteria for MD were indeed selected from a broad possible set of relevant symptoms and signs. While sensible and clinically useful, the criteria selected for DSM-III did not appear to strikingly different from or out-perform other possible depressive symptoms. These findings are consistent with the indexical but not the constitutive view of DSM criteria for MD. They do not support the position that the DSM criteria should be considered literally, as being definitive and authoritative.

The social and historical forces that have contributed to the rise of reification

The impetus toward reification of DSM criteria and the associated diagnostic literalism, driven by a constitutive

view of the role of the DSM criteria, was not a view promulgated by the creators of the DSM-III and its successive editions. It is inconsistent with the historical record, the approach taken to DSM revisions and considerations of how we understand items in related constructs like personality and symptom assessment. How then has it arisen? I suggest four possible reasons.

First, while not mandated by law, DSM has become the 'official' nosology for psychiatry and all mental health related fields in the USA. It is also widely used in other parts of the world. While we have many versions of personality and intelligence scales, aside from the ICD-10 manual, which is not a serious competitor in the USA, DSM is the only show in town. Mental health providers use the diagnoses and the criteria for billing for our services. Psychiatric residents, and psychology and social work interns are taught the criteria and can be expected to be tested on them. The criteria are often cited in legal contexts. Research projects to be funded by the NIH and published in high-impact journals need to use the DSM diagnoses. Each time a new DSM edition is published, structured interviews, which are used extensively in clinical and epidemiological studies, are updated to reflect the newest criteria. All of these social processes tend to give great weight to the DSM criteria – to give them an aura of authority and definitiveness. I suggest that the mental health field has confused a social process – whereby DSM criteria have become 'official' – with an ontological and scientific one – that the DSM criteria definitively describe and define the nature of psychiatric disorders.

Second, the development of DSM-III was part of a social-historical movement within American psychiatry with the goal of legitimizing our discipline and helping it to return to the medical fold by rejecting the anti-diagnostic views of psychoanalysis. It has succeeded in these goals. Next to the Merck manual, DSM is the best-selling book in medicine with DSM-IV selling over a million copies (Anonymous, 2013). Our field benefits when non-psychiatric mental health professionals, patients, and lawyers consider the DSM criteria to be authoritative and to describe definitively psychiatric illness. So, we as a field may be guilty of encouraging excessive respect for our own creation. It is understandable we would want to feel proud of DSM and join in the chorus of praises. Who would want to be a killjoy and protest against this reification?

Third, because we do not have for most DSM disorders a well understood pathophysiology or definitive laboratory tests, it is easier to 'elevate' our symptomatic criteria to constitutive status. Given that we have a clear understanding of the etiology of myocardial infarction and have a definitive diagnostic procedure (coronary angiography), it is easy to see why the

criteria of typical chest pain, EKG changes and troponin increases, are indexical and not constitutive. We cannot do this for our DSM criteria.

Fourth, medicine is full of signs of illness that are obviously indexical because they reflect a key pathophysiological process but do not reflect major symptoms. Consider Kayser–Fleischer rings in Wilson’s disease that are present in 95% of patients with neurological symptoms (Wilson Disease Association, 2016). This sign reflects the underlying metabolic abnormality of excess copper but has no causal relationship with the key hepatic and neurological symptoms. It is easy to see that this sensitive sign is an index of illness. We do not have signs like this in our DSM manual. The criteria for depression or schizophrenia all describe important and often disabling symptoms and signs that can be easily confused for a constitutive description of the illness.

Psychometric theory and the indexical and constitutive positions

Taking MD as an example, one important advantage of the constitutive position is that it gives a lucid and unambiguous answer to the question: ‘What is the DSM disorder of MD?’ The response is ‘It is meeting the specified criteria.’ The indexical position has a vaguer and less satisfactory response: ‘MD is a hypothetical diagnostic construct indexed by the specified criteria.’ We say this because we do not yet know enough to operationally define true MD. It is a work in progress.

While the constitutive position argues that MD is just the DSM criteria, the indexical position says that it is *more than just the criteria*. For the indexical position, the concept of MD has surplus meaning above and beyond the specific DSM criteria.

What is this surplus meaning? Let me illustrate this by a thought experiment. Imagine a 100 years from now, with a combination of imaging and molecular tools, we find the definitive biological explanation for MD – a true gold standard. But measuring it is expensive and involves subject risk so we develop, in DSM-15, a set of clinical criteria that can identify with good sensitivity and specificity that pathophysiological state. We can then adopt an indexical perspective for our DSM-15 clinical criteria for MD. But we have no uncertainty about what the core of the surplus meaning is. It is the ability to identify that pathophysiological state plus all the subsequent information we obtain from that about etiological processes, treatment response and expected course of illness.

In the present, however, we do not have a gold standard to definitively characterize the surplus meaning for the MD criteria. We are back to the vaguer

terms such as ‘hypothetical diagnostic construct’ or ‘tentative clinical entity.’ But we should not sell ourselves short. Clinicians know a lot about what depressed patients are like. We know, among other things, the natural history of the disorder, its typical age at onset, its genetics, the role of particular personal and social stressors, the expected patterns of comorbidity and its response to certain pharmacological agents, various hormonal challenge tests, and forms of psychotherapy. There is a lot we know as a field about MD, which together forms the surplus meaning for our DSM criteria. The great psychometrician Paul Meehl referred to this process as construct validation and argued that this validation occurs via a nomological network consisting of all the things we know about the construct to be measured. He writes: ‘Learning more about a theoretical construct (here DSM categories) is a matter of elaborating the nomological network in which it occurs.’ [(Meehl, 2006) p. 19] This is, of course, another way to express the concept of validation of our diagnostic categories we have used since the days of Robins & Guze (1970). Meehl summarizes the process as follows: ‘the best construct is the one around which we can build the greatest number of inferences, in the most direct fashion’ [(Meehl, 2006) p. 16]. That is, we create the best diagnostic constructs by selecting criteria that produce a diagnosis, which is well validated using our tried and true traditional methods.

The constitutive model locks in our definitions. It is premature and reflects a definitiveness about the underlying nature of our disorders, which is far in advance of our current knowledge. Its inflexibility would impede our progress as we iterate from data to diagnostic revision and back again. The constitutive position over-promises what DSM can now actually deliver. The entire DSM enterprise is a rough approximation of a reality of psychiatric illness, the deep nature of which remains largely hidden to us. This is exactly what the indexical model assumes. We have fallible measures of working hypothetical diagnostic constructs. The recent DSM editions reflect solid and reasonable first approximations to a definitive psychiatric nosology. There is a lot about them of which we can be proud. But suggesting we have developed constitutive and authoritative descriptions of all major psychiatric diagnoses should not be one that we claim because it is not true.

Acknowledgements

None.

Declaration of Interests

None.

References

- Abel JL, Borkovec TD** (1995). Generalizability of DSM-III-R generalized anxiety disorders to proposed DSM-IV criteria and cross-validation of proposed changes. *Journal of Anxiety Disorders* **9**, 303–315.
- Anonymous** (2013). DSM-5: By the Book. *The Economist*, May 18, 2013.
- Bridgman PW** (1927). *The Logic of Modern Physics*. Macmillan: New York.
- Cassidy WL, Flanagan NB, Spellman M, Cohen ME** (1957). Clinical observations in manic-depressive disease: a quantitative study of one hundred manic-depressive patients and fifty medically sick controls. *Journal of the American Medical Association* **164**, 1535–1546.
- Chang H** (2009). Operationalism. In *Stanford Encyclopedia of Philosophy* (ed. E Zalta (Principal ed.), U Nodelman (Senior ed.), C Allen (Assoc. ed.), P Oppenheimer, J Alama, E Pease, M Magnani, H Kim (All Assit. eds)), pp. 1–43. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University: Stanford, CA.
- Clayton P, Desmarais L, Winokur G** (1968). A study of normal bereavement. *American Journal of Psychiatry* **125**, 168–178.
- Clayton PJ, Herjanic M, Murphy GE, Woodruff R** (1974). Mourning and depression – their similarities and differences. *Canadian Psychiatric Association Journal* **19**, 309–312.
- CONVERGE Consortium** (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588–591.
- Decker HS** (2013). *The Making of DSM-III: A Diagnostic Manual's Conquest of American Psychiatry*, 1st edn. Oxford University Press: Oxford, UK.
- Eysenck SBG, Eysenck HJ, Barrett P** (1985). A revised version of the psychoticism scale. *Personality and Individual Differences* **6**, 21–29.
- Fava M, Rush AJ, Trivedi MH, Nierenberg AA, Thase ME, Sackeim HA, Quitkin FM, Wisniewski S, Lavori PW, Rosenbaum JF, Kupfer DJ** (2003). Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatry Clinics of North America* **26**, 457–494.
- Feighner JP, Robins E, Guze SB, Woodruff Jr. RA, Winokur G, Munoz R** (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry* **26**, 57–63.
- Fried EI** (2016). The 52 symptoms of major depression: lack of content overlap among seven common depression scales. *Journal of Affective Disorders* **208**, 191–197.
- Hempel CG** (1965). Fundamentals of taxonomy. In *Aspects of Scientific Explanation: And Other Essays in the Philosophy of Science* (ed. CG Hempel), pp. 137–154. The Free Press: New York, Collier-Macmillan Limited: London.
- Hyman SE** (2010). The diagnosis of mental disorders: the problem of reification. *Annual Review of Clinical Psychology* **6**, 155–179.
- John OP, Donahue EM, Kentle RL** (1991). *The Big Five Inventory-Versions 4a and 54*. University of California Berkeley, Institute of Personality and Social Research: Berkeley, CA.
- Jones TD** (1944). The diagnosis of rheumatic fever. *Journal of the American Medical Association* **126**, 481–484.
- Kendler KS** (2016a). Phenomenology of schizophrenia and the representativeness of modern diagnostic criteria. *JAMA Psychiatry* **73**, 1082–1092.
- Kendler KS** (2016b). The phenomenology of major depression and the representativeness and nature of DSM criteria. *American Journal of Psychiatry* **173**, 771–780.
- Kendler KS, Munoz RA, Murphy G** (2010). The development of the Feighner criteria: an historical perspective. *American Journal of Psychiatry* **167**, 134–142.
- Locke J** (1979). *An Essay Concerning Human Understanding*. Oxford University Press: London.
- Meehl PE** (2006). Construct validity in psychological tests. In *A Paul Meehl Reader: Essays on the Practice of Scientific Psychology* (ed. NG Waller, LJ Yonce, WM Grove, D Faust and MF Lezenweger), pp. 9–30. Lawrence Erlbaum Associates, Inc.: Mahwah, NJ.
- Mill JS** (1843). *A System of Logic, Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*, vol. 2. John W. Parker: London.
- Robins E, Guze SB** (1970). Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *American Journal of Psychiatry* **126**, 983–987.
- Ropes MW, Bennett GA, Cobb S, Jacox R, Jessar RA** (1957). Proposed diagnostic criteria for rheumatoid arthritis: report of a study conducted by a committee of the American rheumatism association. *Annals of the Rheumatic Diseases* **16**, 118–125.
- Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH** (1996). The inventory of depressive symptomatology (IDS): psychometric properties. *Psychological Medicine* **26**, 477–486.
- Skimina E, Strus W, Topolewska A, Rowinski T, Ciecuch J** (2014). The short IPIP-BFM-20 questionnaire for measuring the Big Five. *Roczniki Psychologiczne* **17**, 385–402.
- Spitzer RL, Endicott J, Robins E** (1975). *Research Diagnostic Criteria for a Selected Group of Functional Disorders*, 2 edn. New York Psychiatric Institute: New York.
- Stone TT, BURRIS BC** (1950). Melancholia; clinical study of 50 selected cases. *Journal of the American Medical Association* **142**, 165–168.
- Whitehead AN** (1925). *Science and the Modern World*. Free Press (Simon & Schuster): New York, NY.
- Wilson Disease Association** (2016). *Kayser-Fleischer Rings*. The Wilson Disease Association (WDA): Milwaukee, WI. 11-4-2016. Online Source.
- Zachar P** (2014). *A Metaphysics of Psychopathology*, 1 edn. Massachusetts Institute of Technology: Cambridge, MA.