

Capturing minds: Towards a methods critique of questionnaire-based mental health surveys

Theory & Psychology
1–23

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09593543221104734

journals.sagepub.com/home/tap**Ger Wackers^{id}** and **Marthe Schille-Rognmo**

UiT The Arctic University of Norway

Abstract

Mental health surveys of general populations use psychometric instruments derived from psychiatric symptom checklists and assessment scales. Mental health surveys of this type have become so ubiquitous and influential that the psychometric methods that are at the heart of them seem to be beyond reproach. Are these the right tools to do the job of capturing the minds of general populations? This article pursues a critical assessment of psychometric instruments embedded in mental health surveys through a historical reconstruction of the major epistemic shifts in the investigative practices through which these psychometric instruments developed. The reconstruction traces a strong influence of physics and physicists' notion of fundamental measurement of quantities on psychologists' attempts to measure mental phenomena. Surveys employing these instruments inherit unresolved methodological issues from their psychophysical predecessors: problems of causal inference from mathematical abstractions (correlations) and reification of mental entities from theoretical concepts.

Keywords

causal inference and reification, infrastructural inversion, material–discursive reconstruction, methods critique, psychometric instruments

Supported by the emergence of national burden of disease programs, expanding into mental and behavioural disorders, an increased focus on health promotion and disease prevention has engendered a proliferation of population health mapping activities.

Corresponding author:

Ger Wackers, Department of Health and Care Sciences, Faculty of Health Sciences, UiT The Arctic University of Norway, P.O. Box 6050 Langnes, Tromsø, 9037, Norway.

Email: ger.wackers@uit.no

Among them are questionnaire-based mental health surveys, applied to general populations, that aim to capture psychological states of mind or phenotypic expressions of mental disease entities. Frequently, these mental health surveys of general populations return prevalence numbers for depressive symptoms or anxiety, for example, that are substantially higher than prevalence numbers derived from diagnostic and treatment data produced by mental health services. These numbers, as are all estimates in the social epidemiology of mental health and disorder, are the result of specific *practices of calculation*. From a governance perspective, the higher prevalence numbers of general population health surveys suggest the existence of undiscovered suffering and unmet needs for mental health care that can no longer be neglected.

The digitalization of earlier pen-and-paper surveys has provided these mental health surveys with a data infrastructure that allows for: (a) larger numbers of participants per survey; (b) less time for the participants to complete the questionnaire; (c) shorter time intervals between surveys; and (d) shorter loop times¹ between data production, publication, and discussion of results in the public domain, as well as the initiation of public health interventions based on these survey results.

Participation in the surveys is voluntary and the symptoms of mental distress are said to be self-reported. However, participants are not free to choose either the items to which they respond or the format of their response. The items typically contain terms indicative of some form of deficit, derived as they are from psychiatric classifications of disease (ICD, DSM). Response formats are restricted to digitized,² Likert-type numerical scales.

Surveys of this type have become so ubiquitous and influential that the psychometric methods at their hearts no longer attract critical scrutiny. Yet, in the practice of these surveys there is a claim that mental states and traits can be expressed in numbers, at the aggregate population level even in a single prevalence number.

In this article, we aim to develop an epistemic-methods critique of the psychometric instruments embedded in questionnaire-based mental health surveys by retracing and reconstructing their development. How can mental health be *measured and mapped* through a limited set of itemized questions and a response format limited to digitized, numerical rating scales? How can *mental health* be mapped with questions derived from *psychiatric symptom checklists*? How do various variables correlate with each other and with the concealed mental objects they purportedly are attributes of? What do the numbers represent? What is lost in this pursuit of “strong calculations” that seek “strong results” presentable as a single number? What is the foundation for their claim to objectivity and their usefulness in mental health governance? Are these the right tools for the job of capturing of minds in a general population?

The practice of questionnaire-based mental health surveys is framed by specific ontological and epistemological assumptions and positions. An epistemic-methods critique should address these framing assumptions, not reproduce them. A critical inquiry must *shift the frame* in order to achieve critical momentum. Before embarking on the reconstruction of psychometric instruments employed in mental health surveys we will therefore, in the next section, make explicit our own theoretical and methodological position underlying the arguments developed throughout the article.

Theory and method

Theoretical position: World-making agencies of observation

Obviously, mental health surveying is a measurement-based *knowledge producing practice*. Our goal is to examine ontological and epistemological presuppositions underlying and framing psychometric instruments. We distinguish between a *representational* and a *performative* perspective on knowledge. In the former view, scientific knowledge represents a preexistent world objectively, preserving an a priori, ontological distinction between reality and the knowing subject. The object to be measured occupies a position of externality with regard to both the measuring equipment and the agent or scientist who deploys this equipment. On this view, object and measuring equipment are conceived of as separable. The representational view tries to establish how the world is, and how knowledge mirrors or corresponds with the world.

In the latter *performative* view, that we adopt in this study, this reality-representation scheme is inverted. Knowledge-producing practices, also called *epistemic practices* (Knorr Cetina, 1999), perform the world through the knowledge they produce. A performative view emphasizes all the practical work that is required to make, stabilize, reproduce, and transform the world through epistemic, investigative practices. Knowledge is, on this performative view, not about the object-in-the-world, but the epistemic object is the reified object as produced in and through knowledge practices. The scientist-experts, understood as *epistemic subjects*, as procurers of knowledge are enfolded, socially, physically, and cognitively, through training and socialization, in what Karin Knorr Cetina (1999) calls *epistemic cultures*. Where the representational view of knowledge places the subject opposite the world, the performative view places more emphasis on the agency of research collectives in producing the world. Agencies of observation are part of, not separate from, material–discursive practices that extend into and are entangled with other culturally specific legal, political, and economic formations.

Inspired by Niels Bohr's early 20th-century work in quantum mechanics, Karen Barad (2007) insists that what is to be measured is fundamentally *indeterminate*. It only becomes determinate through the material and discursive practices of the agencies of observation. It is these practices that constitute the epistemic object as a phenomenon in the world. The implication of this position is that changes in the material conditions and procedures of agencies of observation will rework the phenomenon. As a result, the epistemic object produced will also have changed. Barad urges us, as a methodological imperative, to not “presume that an object has determinate boundaries and properties in the absence of their specification through the larger material arrangement” (p. 160). Apparent boundaries do not reside in the nature or essence of things that exist as objects in the world, but are produced, performed, and preserved.

The methodological challenge in this article is to show how shifts in material–discursive practices in psychometrics have both enabled and constrained the epistemic objects produced by them, thus generating a space in which questionnaire-based mental health surveys can exist as ubiquitous and uncontested mental health survey tools.

Method: Infrastructural inversion and material–discursive reconstruction

This article pursues a critical appraisal of psychometric methods deployed in mental health surveys through what Geoffrey Bowker and Susan Leigh Star have called an *infrastructural inversion* (as cited in Bowker et al., 2015, p. 477). As a method, infrastructural inversion brings to the surface for critical scrutiny what usually remains hidden, sunk as it is into other structures, social arrangements, and technology. Much of what we will bring to the surface for critical scrutiny will be of a conceptual nature, consisting of onto-epistemological assumptions about the measurability of mental phenomena. We will focus on a description and analysis of shifting views on the relationship between the concealed objects of the mind and the observable, phenotypic attributes being measured and mapped. We have selected, and present in some detail, the work of psychophysicists like Charles Spearman, Louis Leon Thurstone, and Stanley Smith Stevens, because in their work, we can trace the major shifts that shaped the space in which questionnaire-based mental health surveys exist.

The technical psychometric literature is of a bewildering statistical and mathematical sophistication. The esoteric, technical uses of statistical concepts and obfuscating mathematical notation has created a veritable sociocognitive barrier to intelligibility and direct critical engagement. Access to assumptions embedded in psychometric methods is possible though, by retracing their origins, in a historical, diachronic material–discursive reconstruction. The word *material* signifies a sensitivity to the changes in concrete, physical aspects of investigative practices. The word *discursive* signifies a sensitivity to conceptual shifts and arguments. The dash indicates that these two aspects are inseparable.

Material–discursive reconstruction, or *reconstruction* for short, as we understand and use the term here, makes use of the past to better understand the present, but it has no ambition or claim to present a full history. A timeline roughly supports the diachronic narrative. Like cinema, to use Hacking's metaphor (2002, p. 6), it cuts between close shots and distant perspective, connecting and juxtaposing points or events that are disparate in time and space. Reconstruction does not produce tales of linear progress towards the present. There is no straight line that can be fitted to the diachronic, historical data. Reconstruction is sensitive to historical patterns of diversification as well as patterns of conceptual blending or fusion. Reconstruction also aims to explicate patterns of deletion or displacement, of what is lost and excluded through the emergence of particular material–discursive formations like mental health surveys. Hence, in this article, material–discursive reconstruction is a manner of explicating and critically evaluating the deeper generative processes and conditions of possibility that gave rise to and allow for particular knowledges about mental attributes to be presented numerically with the authority of objective, pure science. However, according to sociologist and historian of scientific knowledge Stephen Shapin (2010), knowledge is not and never has been pure. Claims of independence and impartiality, of objectivity and validity, Shapin argues, support an emerging science's credibility, authority, and usefulness in practices of control and governance. This holds true, not in the least, for the science of correlative psychology that emerged with Spearman's work in the first decades of the 20th century. To appreciate the shift in investigative practice accomplished by Spearman, a note on the German school of experimental psychology of the last quarter of the 19th century is in order.

Spearman's correlative psychology

The German school of experimental psychology

During the second half of the 19th century, an experimental psychology diverged from the philosophy of mind. German researchers like Gustav Fechner, Wilhelm Wundt, and Hermann Ebbinghaus applied the experimental methods of classical, Newtonian physics and of 19th-century physiology to the study of the mind and soul (Danziger, 1991, 1997; Hacking, 1995b). Wundt still published his work as *philosophical investigations*, whereas Fechner published his work under the heading of a new *psychophysics*. Instead of philosophical reflection, which is still the tool of the trade in the philosophy of mind today, this German school adopted experimentation and measurement as the primary methods of investigation in the study of mental phenomena. The concepts of cause–effect and stimulus–response relationships blended (Fauconnier & Turner, 2002; Turner, 2014) into a new logic for the investigation of mental faculties. Through experimental stimulation, Wundt and other experimental psychologists endeavoured to study the antecedent causes of mental phenomena. Mental phenomena were conceived as effects in consciousness (acts of sensory perception, discrimination, and recall). Typical for the German school's investigative practice was the combination of experimentation with a first-person perspective on the knowability of mental phenomena. These experimentally elicited effects in consciousness were observable, hence knowable, through introspection, especially for the trained minds of the psychologist-researchers (Danziger, 1991). This introspective investigative practice emphasized and recognized the uniquely individual and subjective nature of mental phenomena.

Spearman's trust in the mathematics of associations

Working in England in the first decades of the 20th century, in an effort to remedy what he called the dismal state of experimental psychology, Charles Spearman (1904a, 1904b) initiated a program of correlative psychology as the foundation for a new scientific psychology. In 1904, Spearman (1904b) boldly claimed to have objectively determined and measured an innate and heritable entity called *general intelligence*. Note that Spearman worked under the influence of the eugenics of his place and time, of which Francis Galton and Karl Pearson were prominent proponents.³

Spearman's ambition was to go beyond the German school's description of "bare isolated occurrences" observed introspectively by individuals. His aim was to produce scientific knowledge "that deals with uniformities," with objectifiable relations that hold across populations (Spearman, 1904a, p. 72). To do so, Spearman employed mathematical–statistical tools developed by Francis Galton and Karl Pearson for the study of heredity in populations. In Spearman's program of correlative psychology, the mental phenomena that were to constitute the subject matter of scientific psychology (intelligence, personality, attitude) were conceptualized as objects-with-attributes that could be located in the minds of every individual in a population. Furthermore, Spearman assumed mental entities to be quantitative in nature. This was a crucial prerequisite with regard to their purported measurability. Spearman was not alone in that assumption. The quantitative nature of mental phenomena has been assumed by experimental psychologists since the days of Fechner, but has, according to Michell (1999), never been proven.

Spearman recognized that the essential, complex nature of the mental object in itself was not amenable to direct measurement; it remained concealed. The object's phenotypic attributes, though, were observable. Data derived from measurements of purported attributes would have to serve as a proxy for indeterminate, immeasurable, inaccessible, and private objects in the mind.

Spearman followed Francis Galton and Karl Pearson in claiming that neither insight into the nature of the object, nor the causal architecture underlying the pattern of observable, phenotypic attributes, were necessary for a scientific psychology to proceed. This claim was and is the hallmark of any correlative science. Hypothetical causal mechanisms may be proposed, even argued to be plausible, but they are not necessary for the knowledge to be valid. Spearman (1904b) strove for a "precise *quantitative expression* [emphasis added] derived impartially from the entire available data," aiming for "a more complete acquaintance . . . concerning objective *relations* [emphasis added]" (p. 225). In other words, valid, objective knowledge could be obtained through the measurement of the strength of association (correlation) between multiple observable, phenotypic attributes. The measurements of attributes of objects-of-the-mind could, furthermore, be correlated with other features that could be assumed to stand in a causal relationship to the mental object, as either cause or effect: gender, age, race, ethnicity, class, socioeconomic status, level of parental education, and so forth.⁴ Hence, there are two relationships here: one between attributes and an internal mental object and one between the object so measured and external factors.

Spearman adopted the experimental stimulus–response methods of the German school. The tools of this trade consisted in the precise measurement of just-noticeable differences (JNDs) in the perception of weights, of grey-scales and of the pitch of sounds.⁵ In an attempt to measure the children's intelligence, as well as the distribution of natural and innate intelligence in the population, he applied these methods to school children from different schools in different socioeconomic settings.⁶ By rejecting the method of introspection, Spearman's program of correlative psychology effectively silenced the participants who were the object of investigation. It prevented them from speaking their own minds. He also deleted the second-person perspective of a human interlocutor in conversation or interview. For Spearman, human observers were sources of error and he tried to eliminate them altogether. He placed his trust in the "strong calculations" (Winther Jørgensen, 2015) that mathematical statistics could provide:

The whole of our experimentally gained figures must without any selective treatment *simply of themselves issue into one plain numerical value* [emphasis added] (varying conveniently from 1 for perfect correspondence down to 0 for perfect absence of correspondence). (Spearman, 1904b, p. 225)

The italicized part of the quote is an instance of what Gigerenzer et al. (1989) have called "the dream of the mechanization of knowledge" (p. 210), attainable through the application of mathematical and statistical tools that will allow the scientist to arrive at new knowledge in the service of objectivity by eliminating all personal judgement. In Spearman's view, mathematical formulae and statistical manipulations allowed strong correlations to reveal themselves.

Of course, the strong correlations did not simply issue forth of themselves. Page after page, Spearman (1904b) described the work he performed to boost the initially partial and low correlations towards higher values in amalgamated series and higher order correlations. The stronger correlations boosted the certainty of Spearman's beliefs. Negating and disowning the agency of his own manipulations of data, Spearman relocated the agency of his own work to a set of mathematical formulae. These formulae simultaneously secured, Spearman argued, the researcher's impartiality, an impartiality that issued forth of the methods used.

This claim to impartiality is a discursive–rhetorical strategy that, if we accept it, conceals the work going into the production of data, as well as the unarticulated assumptions embedded in the measurement techniques through which they are produced and processed. Spearman's claim to impartiality performs what Barad (2007) calls an *agen-tial cut*, that is, it discursively separates the numerical value of the correlation that issued of itself from the data from his a priori, eugenic assumptions about the innateness and heritability of intelligence (Spearman, 1904b, p. 225).

The rhetorical power of this cutting operation should not be underestimated: it produced simultaneously a claim to the objective, prediscursive, and premeasurement existence of the object measured, to the measurability of the mental entity under investigation, to the representational objectivity of the knowledge produced, and to the independence of the data-producing and data-processing agencies of observation. That is, no doubt, an impressive achievement. If we chose to accept it.

Spearman's error: Reverse inference and reification

Epistemic objects live by the grace of the specificity of the practices that produced them (Barad, 2007; Knorr Cetina, 1999). A correlative science like Spearman's correlative psychology faces crucial epistemic challenges. One is the problem of reverse inference and subsequent reification. That is, inferring the existence of a concealed entity from measurement of its purported phenotypic attributes.

Spearman adopted the tools of mathematical statistics from Karl Pearson. However, Spearman did not comply with Pearson's (1911/2007) philosophy of science. Pearson warned against the ascription of ontological status (reification) to the mental constructs of the scientist. Such an ascription would amount, in Pearson's words, to illogical inference. The scientist's mental constructs should not, in Pearson's view, be reified and projected into the world.

Despite Spearman's (1904b) assertion that he postponed for later "a discussion as to the psychical nature" (p. 284) of the correlations obtained, he addressed it already in the same 1904 article. Apparently, Spearman was unable to resist the lure of reification, of ascription of ontological status, against which Pearson warned.

Using mathematical techniques for principle component or factor analysis, derived from the mathematical discipline of linear algebra, Spearman artificially unified, reified, and naturalized a set of mathematical abstractions (correlations) into an independently existing heritable entity. This epistemic object became known as Spearman's *g*.

In the timeline of the investigative work process, the fabricated epistemic object, Spearman's *g*, came after the measurements on which it was based. As Spearman's

confidence peaked, the object flipped back in time, from a thing that emerged after the measurements had taken place, to some-thing (natural and innate) that had been there all along.

Spearman was convinced that he had “objectively determined and measured” the natural innate faculty that he set out to find, namely general intelligence. In terms of our performative view on knowledge production, he failed to see that the epistemic object or mental phenomenon he had produced was a performative reality-effect of the emergent agency of observation, characterized by an epistemic culture influenced by the eugenics of the time, of which he himself was an instance. Instead of the conceptual invention that it was, Spearman could now, against Pearson’s advice, present general intelligence as a representation of something natural, innate, and heritable. Spearman’s *g* did not survive, however. It disappeared again. Its demise was due to changes in the methods of factor analysis (Gould, 1981/1996). Some of these changes were brought about by the American psychophysicist Louis Leon Thurstone.

Thurstone’s measurement of attitudes and opinions

Working as a psychophysicist in Chicago in 1928, Thurstone boldly claimed that “Attitudes can be measured.” Whereas sociologists conceived of attitudes as the subjective side of culture, “for psychologists, attitudes were strictly *individual* attributes where individuals were understood as separate entities and not as the parts of a social or cultural collectivity” (Danziger, 1997, p. 144). Attitudes were taken to be actually existing states inside individuals. Once formed, they were carried around by the individual on a more or less long-term basis (Allport, 1935).

Like Spearman, Thurstone recognized that the complexity inherent in the notion of attitudes was far beyond the reach of direct measurement. To work around that problem, Thurstone adopted, like Spearman, the strategy afforded by the conceptual blending of the concepts of cause–effect, stimulus–response, and object–attribute relationships. Thurstone added independent–dependent variables to the blend. In the blend, the independent variable occupies the place of cause in the cause–effect relationship. Since mental objects are concealed and unobservable, in psychometrics, the independent cause-variable came to be called a *latent variable* (Borsboom, 2005).

Thurstone introduced a distinction between *attitude* and *opinion*. Attitudes were, according to Thurstone (1928), “the sum total of a [person’s] inclinations and feelings, prejudice or bias, preconceived notions, ideas, fears, threats, and convictions about any specified topic” (p. 531). He defined opinion as “a verbal expression of attitude” (p. 532). Thurstone then operationalized the measurement of attitudes by proposing to “use opinions as the means for measuring attitudes” (p. 532). Thurstone called these opinions attitude *variables*.

To construct a psychometric instrument that would measure attitudes, Thurstone replaced physical stimuli of weights, grey-scales, and pitch of sounds with verbal statements expressing opinions.⁷ Subscribing to the physicists’ classical notion of measurement of quantities, Thurstone argued that an attitude variable should be describable in such a way “that one can speak of it in terms of ‘more or less,’” because “the very idea of measurement implies a linear continuum of some sort such as length, price, volume,

weight, age” (Thurstone, 1928, p. 534).⁸ For the measuring instrument to be valid, it had, furthermore, to be independent of and external to the quantities measured. In his explication of requirements of validity that would apply for psychometric instruments, Thurstone (1928) used the deceptively simple example of measuring the length of familiar objects with a ruler:

A measuring instrument must not be seriously affected in its measuring function by the object of measurement. To the extent that its measuring function is so affected, the validity of the instrument is impaired or limited. If a yardstick measured differently because of the fact that it was a rug, a picture, or a piece of paper that was being measured, then to that extent the trustworthiness of that yardstick as a measuring device would be impaired. Within the range of objects for which the measuring instrument is intended, its function must be *independent of the object of measurement* [emphasis added]. (p. 547)

Thurstone’s example was an instance of the measurement of an attribute (length) that had an additive and scalable structure (the linear continuum) in the classical sense of a quantity. Any quantity that exhibits this empirical structure of additivity and scalability can be expressed as a ratio between the whole and an arbitrary segment of that quantity that serves as a conventional unit. It is the empirical structure of quantities (additivity and scalability) that supports ratios and ratio scales. That is how the conventional ruler works, whether the arbitrary segment of the whole is a centimetre or an inch. In this classical view on the relationship between measurement and quantities, natural numbers exist in the world. They are part of the furniture of the world (Michell, 1999, p. 25). As physical properties of the world, numbers may be discovered through proper measurement. In 1920, physicist and philosopher of physics Norman Robert Campbell (1928) axiomatized this form of measurement as *fundamental measurement*.

Thurstone performed a final bootstrapping operation that performed Barad’s (2007) *agential cut*, separating the object to be measured (attitudes) from the material conditions of the production of the instrument. Thurstone (1928) claimed that the statistical procedures applied in the construction of the scale warranted

the assumption that the scale values of statements are independent of the attitude distribution of the readers who sort the statements. . . . If the assumption is correct, then the scale is an instrument independent of the attitude which it is itself intended to measure. (p. 548)

The word “independent,” and its implied externality, as used by Thurstone, supported the idea that not only the measuring equipment, but the whole of scientific psychology as an agency of observation was independent of, separable from, and external to the mental entities it measures and maps.

In Thurstone’s view, all forms of qualitative, psychological complexity could be reduced to independently defined, linear, scalable, and numerically expressible variables. The linear scales of agreeability with verbally expressed opinions act as quantifiers. They exhibit agency in the sense that they turn something indeterminate that does not have quantitative structure (in the classical sense) into something determinate that is numerically expressed as if it is a quantitative attribute. As a result, psychometric instruments that use these linear scales imply that the observed measurement is supported by

an independently existing, although concealed, real-world structure, a latent variable, that acts as the sole cause of the measurement results (effects) and correlations obtained.

Thurstone constructed a template for questionnaire-based psychometric instruments that could serve as the tools of the trade of a scientific psychology.⁹

Operationism

Thurstone's work was consistent with the methodological principles that Stanley Smith Stevens some years later would articulate under the heading of operationism. The advent of relativity theory and quantum mechanics in the early 20th century created trouble for classical Newtonian physics, especially for Newtonian concepts like absolute space and absolute time (see Pearson, 1911/2007). To expunge metaphysics from physical theory, the Harvard-based physicist Percy Bridgman argued in 1927 for an *operational analysis* of theoretical concepts. Key to Bridgman's approach was that only concepts that could be defined in terms of the empirical operations that were employed to determine or measure them were to be retained. Length, for example, would thus be operationally defined by the operation of moving a ruler repeatedly along the object to be measured. Newtonian concepts like "absolute space" and "absolute time," that could not be so defined, would have to be abandoned.

Under the influence of the logical-positivist philosophy of science of the Vienna Circle, Harvard psychologists Edwin Boring and Stanley Smith Stevens adopted this notion of *operational definition of concepts*. Especially the work and publications of Stanley Smith Stevens (1935a, 1935b, 1942) turned operationism into a founding principle of a new scientific, psychophysical psychology (Miller, 1974).

Operationism helped Stevens (1960) circumvent two long-standing problems of measurement in psychophysics: the privateness of subjective mental sensations and the role of mathematics.

Then, too, there was the issue concerning the privacy of sensation, which was regarded as a nonphysical mental affair, inaccessible to objective methods. Under the modern view of things, in the study of sensation there need be no question of penetrating privacy, because the sensation that science deals with is a type of human reaction that lends itself to public scrutiny. (p. 27)

Where Thurstone maintained a distinction between object and attributes, between attitudes and opinions, Stevens collapsed object and attributes into one. "What is here meant by sensation," Stevens wrote in 1959, "is a construct, a conception built upon the objective operations of stimulation and reaction." Stevens explained: "We study the responses of organisms, not some nonphysical mental stuff that by definition defies objective test" (p. 612).

Concerning the role of mathematics, Stevens argued that in the early 20th century, scientists became aware that numbers were not physical properties of the world but a game of signs and rules that could be divorced from the world and then pinned arbitrarily to things. Defining measurement as "the assignment of numerals to objects and events according to rules," Stevens (1946, p. 677) extended the province of measurement in psychology to include a variety of scales used in experimental psychology. These scales

were constructed on the assumption of “a certain isomorphism between what we can do with *aspects of objects* [emphasis added] and the properties of numeral series” (p. 677). Like Thurstone before him, Stanley Smith Stevens contributed substantially to the digitization of measurement in psychology.

Stevens’ operationism invites us to accept the collapse of the indeterminate and concealed object of mental phenomena into its observable phenotypic attributes. It invites us to accept that measurement in psychology is possible by creatively pinning numbers on purported attributes through digitized scales. None of these invitations is it necessary to accept at face value. On the contrary, behaviourist operationism has not resolved, but worked around the inevitable question of reverse inference where measurement is concerned. What is it that the measured attributes are signs of? What is it that has been measured and mapped? These questions become especially acute when the entities to be measured imply personal deficits and mental disorders.

The confluence of psychometry with psychiatry

Thurstone’s explorations into psychopathology

In the preceding sections, we have been occupied with an infrastructural inversion (Leigh Star as cited in Bowker et al., 2015, p. 477) and critical assessment of the framing assumptions underlying psychometric methods that were developed with the ambition to objectively determine and measure psychological entities. However, our concerns with questionnaire-based mental health surveys include the use of terms derived from a psychiatric vocabulary to measure and map mental states of mind in a general population. Where, when, and how was the link between psychometrics and psychiatry established?

Thurstone entered the domain of psychopathology when he, in 1934, in “The Vectors of Mind,” published “A Factor Study of the Insanities.” In this study, Thurstone derived the items for the construction and initial validation of a psychometric instrument from the professional corpus of psychiatric terms for the description and categorization of psychiatric symptoms. Thurstone (1934) “used a very elaborate set of data which Dr. Thomas Verner Moore of Washington D.C. collected” (p. 18). Moore, a practising psychiatrist, worked with an inventory of 48 psychiatric symptoms, featuring, among others, “alcoholism of parents”; “anxious, bizarre delusions”; “homicidal”; “insane relatives”; “absence of insight”; “suicidal”; “tantrums”; and “voices, speaking to.” Moore recorded for each of several hundred patients a rating or test measure for each of these items:

With these records it was possible to ascertain to what extent any two symptoms tend to coexist in the same patient . . . The multiple factor method was then applied to the table of . . . coefficients and we found that five factors are sufficient to account for the correlations, with residuals small enough so that they can be ignored. (Thurstone, 1934, pp. 18–19)

Thurstone (1934) found “twenty-six symptoms which are more or less related and for which the factorial clusters of symptoms could be profitably investigated” (p. 19). Thurstone provided a table listing five clusters of psychiatric symptoms. He was

cautiously optimistic about having pointed psychiatry in the right, rational direction with his explorative multifactor analysis. He claimed that his “results indicate that by the multiple factor methods it should be possible to arrive at a rational classification of the insanities and of personality types” (pp. 20–21).¹⁰

Thurstone established a bridge between his new scientific psychometry, using verbal statements and digitized scales, and the symptom checklists that were used in clinical psychiatric practice. Having been established, the link subsequently allowed for the flow of a psychiatric vocabulary into psychometric instruments. Questionnaires for the measurement of attitudes fused and hybridized with psychiatric symptom checklists into psychiatric assessment scales. Subsequently, these hybrid psychometric instruments diverged and were adapted to two different regulatory regimes: one for the approval of psychoactive drugs, the other for public mental health governance.

Psychopharmacology and regulatory requirements for clinical effect measurement

After the Second World War, a number of events and processes set the stage for further developments. In 1948, psychiatric sequelae of the war prompted the inclusion of a chapter on mental disorders in the 6th version of the *International Classification of Diseases* (WHO, 1948). ICD-6 was the first to be issued under the aegis of the World Health Organization (WHO) since its establishment in 1948. In 1952, discontented with ICD-6's classification of mental and behavioural disorders, the American Psychiatric Association published the first version of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-1) (APA, 1952).¹¹

That year, 1952, also witnessed the introduction of the first antipsychotic drug, chlorpromazine. “Its discovery,” Healy (1997) wrote, “was the critical event in the foundation of psychopharmacology” (p. 43). Chlorpromazine was quickly followed by the first antidepressant drugs.

These psychoactive substances had profound implications for the theoretical understanding of the nature and cause of mental disorders. The monoamine theory of depression followed soon after the bio- and neurochemical elaboration of the molecular mechanisms that were proposed to explain the antidepressive effects observed in clinical trials: antidepressants worked through the modulation of monoaminergic neurotransmission at a synaptic level (López-Muñoz & Alamo, 2009). This theory provided a causal-etiologi- cal explanation and definition of depression. It turned the search for the causal infrastructure generating depressive symptoms firmly to the interior of the patients, and to their brains. Depression became a brain disease. This view supported, according to Nikolas Rose (2019), the belief in an “epistemologically misleading biological universality” (p. 148) underlying symptoms of mental distress, implying, as a logical entailment, that the “key direction of causation is from brain processes to mental life and behaviours” (p. 113). This understanding developed into the “medical model” underlying current international classifications of mental and behavioural disease, DSM and ICD: disease processes internal to the patients cause the observable symptoms that constitute the syndromic description of the mental disorder listed in the “manuals.”

The symptoms (dependent variables) correlated statistically, because, as parallel effects, they had a common cause (the independent, latent variable).

The introduction of psychoactive substances held profound promise for the treatment possibilities they implied. These possibilities triggered the rapid involvement of a range of pharmaceutical companies (Healy, 1997). Regulatory frameworks for the approval of new drugs were already under development in both the United States and Europe, requiring not only evidence for the effectiveness of new psychoactive drugs with regard to the symptoms they were intended to reduce, but also with regard to side effects and the absence of harmful effects. A need emerged for clinical effect measures in clinical trials, the construction of dose-response curves and the development of guidelines for the new drugs' use in practical treatment.

The comprehensive symptom checklists that were used in practical psychiatric care to assess changes in the overall clinical condition of in-house patients were not suitable to serve as instruments for the measurement of effects in clinical trials. They had to be adapted and trimmed. The argument went roughly as follows. A positive correlation between symptoms implied a common factor and a kind of redundancy in the data. This redundancy limited the amount of useful information that any one item, or additional item, could yield. The goal was to reduce the number of items and select those items that were independent of each other and that, furthermore, could be spaced along a linear, unidimensional scale.

The development of short psychiatric assessment scales for effect measurement in clinical trials was supported and justified by the work of mathematicians Duncan Luce and John Tukey (1964). Remaining true to the physicists' ideal of fundamental measurement, Luce and Tukey took fundamental measurement into domains where the objects or attributes to be measured did not possess the properties of additivity and scalability of physical quantities like length or weight. That is, into all of psychology, the behavioural, educational, and social sciences. Luce and Tukey's 1964 paper was published in the first issue of the first volume of the new *Journal of Mathematical Psychology*.

Additivity of symptoms, disability, and suffering in linear assessment scales

One psychometric instrument combines two linear, digitized scales. At item level, there is Likert's digitized scale of agreeability with statements (items) about symptoms. Across the range of items there is a linear scale of severity. By preserving the additivity of the linear scale across the set of items, the argument goes, the numerical scores per item can be added into a meaningful total score of suffering for the clinically manifest mental disorder.

In a lecture given in Copenhagen in 1977, David Hamilton, the designer and developer of the Hamilton Depression and Hamilton Anxiety Scales, explained the clinician's take on psychometric assessment scales. Hamilton claimed a general acceptance, in all branches of medicine, that the more symptoms a patient experiences, the more ill they are. Consistent with a claim to the preserved additivity of items across the instrument, symptoms, and the suffering resulting from them, could be added. "The doctor goes

through a list of symptoms and checks how many are shown by the patient. The total checked is a measure of the severity of the illness” (Hamilton, as cited in Bech, 2012, p. 118).

When we add scores we are not so much adding scores on depression, loss of weight or loss of libido, as *adding up* [emphasis added] measures of disability. It is disability which is common to all the symptoms and so a total score *represents* [emphasis added], in a way, the suffering of the patient. (Hamilton, as cited in Bech, 2012, p. 119)

Psychiatric rating scales—Observer and self-report scales

Observer scales, that is, rating scales scored by the observing clinician, were, in Hamilton’s (as cited in Bech, 2012) view, “no more than a particular way of recording a clinical judgment” (p. 117). The validity of their use remained firmly anchored in the psychiatrists’ clinical practice and experience:

The observer scale when used by an experienced clinician can record very small and delicate changes, which are difficult for the inexperienced person and especially for the patient, to recognize. However, they do take a long time, even half an hour’s interview is, in my opinion, not really enough. (Hamilton, as cited in Bech, 2012, p. 119)

Due to the effort involved, observer-scored scales constitute high transaction cost methods, that is, high costs involved in the production of data. However, the observer scale had the advantage that it could include items soliciting information which the patient, by definition, could not give, such as loss of insight or delusions. Self-assessment scales could not include such items but had the advantage that they were easy to use repeatedly (Hamilton, as cited in Bech, 2012, p. 119). A patient could take them home and score them on a daily or hourly basis, greatly increasing the intensity of data production and tracking changes in much more detail over shorter periods of time in noninstitutional settings.

Psychometrics of mental health surveys of general populations

The hybrid instruments that blended questionnaires for attitude measurement with psychiatric symptom checklists were also adapted into another regulatory environment, namely that of public mental health governance, and therewith into the branch of psychiatric epidemiology underlying public mental health policies.

Since its initiation in 1990, the Global Burden of Disease project has fostered an increased focus on health promotion and disease prevention. This has been accompanied by a call for ways in which to monitor populations for mental health risk factors and early signs of disease. Consequently, another major shift in the development of psychological measurement took place when psychiatric symptom checklists and assessment scales, like the Hopkins Symptom Checklist, found their way into public health surveys that were intended to provide the knowledge base for public health policies and interventions.

Hopkins symptom checklist

The Hopkins symptom checklist (HSCL) was rooted in the Cornell medical index (Wider, 1948) and was further expanded by investigators at the Johns Hopkins University in the 1950s, the decade that ushered in the age of psychoactive drugs (Parloff et al., 1954). Since the 1960s, the development of the HSCL was supported by grants from the psychopharmacology research branch of the United States' National Institute of Mental Health (Lipman et al., 1979). Considered to be "sophisticated inventories of established reliability and validity" (Uhlenhuth, 1975), various versions of the HSCL (comprising 10, 35, 58, 72, or 90 items) were used to measure and assess the effects of various treatment modalities. The HSCL was developed primarily as a general improvement measure for research in psychotherapy (Derogatis et al., 1974). As such, its validity was based on a population in a clinical setting found to have some form of psychological problem that was deemed to warrant treatment with psychoactive drugs.

From an observer-scored symptom checklist, the HSCL was adapted into self-report symptom inventories (Derogatis et al., 1974). The shift from observer-scored to subject-scored checklists afforded their transfer from clinical populations to general population surveys. The transfer introduced new methodological challenges, though, one being the problematic of false positives that plays a crucial role in surveys of general populations (Cooper, 2013). The identification of false positives and false negatives requires a second test with a different specificity/sensitivity profile. Without such a second test, mental health surveys in general populations take on the shape of "single-shot" surveys.¹²

The transfer of the psychometric instruments from the clinic to the general population was possible because self-report inventories eliminated the costs of interviewing—at least half an hour per participant in Hamilton's experience. The self-report questionnaires reduced the transaction costs of data production, allowing for the coverage of larger populations beyond selected, representative samples (Schille-Rognmo, 2017). The price to be paid was the loss of the anchoring of the survey's validity in the psychiatrists' clinical experience. Self-report surveys of mental states depend on the participants' own ability to differentiate between and name subtle differences in the experience of emotions, on what Barrett (2018) calls the participant's *emotional granularity*.

Subsequent developments and uses of the HSCL reproduced and normalized the deletion of the first- and second-person perspective from questionnaire-based epidemiological studies of mental phenomena, encouraging and legitimizing the inference and reification of mental disorders from data gathered in single-shot surveys.

Alternative ontologies for mental health and disorders?

What does the methods critique developed here entail? When these are not the right tools for the job, that is, to capture minds in general populations, should public health investigators stop using questionnaire-based mental health surveys?

One sociological answer is that we do not expect that public health investigators, with what Wiebe Bijker (1997) called a "high degree of inclusion" in their epistemic culture, will do away with the knowledge-production tools on which their field rests. There is a

strong recursive, mutually stabilizing relationship between public health policy as a practice of governance and the knowledge produced to scaffold it.

Are there alternative understandings of mental health and disorder? This question requires an epistemic answer. Recognizing that we are trapped in tight conceptual shackles, throwing off these conceptual blinders does not result in a clear and unhampered view on what mental disorders really are. Changing the way we understand and perform mental health and disorder is not easy. Alternative understandings will only be able to live by the grace of recovered or new investigative practices that support them. Needless to say, alternative understandings of conceptualizing mental health and disorders is not entailed in the methods critique developed here. Yet, the combination of a low degree of inclusion in the field and the recognition of the critique makes it easier to see and appreciate alternative conceptualizations under development.

By way of example, we will briefly point to one that originates within the field of psychometric research itself. It starts from a recognition of the impossibility of inferring a hidden mental disease from observable symptoms. According to Borsboom (2017), “we cannot find central disease mechanisms for mental disorders because no such mechanisms exist” (p. 5).

Borsboom (2008, 2017), Cramer et al. (2016), McNally et al. (2014), and others (Fried et al., 2017) propose a radically different conceptualization of mental disorders. They turn to the new physics and new, nonlinear mathematics of complex adaptive systems. In their network models, they no longer try to construct inferential connections that reach below the surface of the individual’s symptomatic behaviour but ascribe causal agency to the symptoms themselves. Mental disorders arise from the interaction between symptoms in a network:

Instead of being *effects of a common cause*, psychiatric symptoms have been argued to cause each other. . . . Symptoms may form feedback loops that lead the person to spiral down into the state of prolonged symptom activation that we phenomenologically recognize as mental disorder. (Borsboom, 2017, pp. 5–6)

Mental disorders, their genesis, and the course that they take, can be thought of in terms of trajectories, tipping points, and attractors in an abstract mental state space. A whole new set of concepts comes into play. In their mathematical models, these investigators have demonstrated *hysteresis*. In its most general formulation, hysteresis is the dependence of a system on its history. Hysteresis is common in biological systems (Noori, 2014). Cramer et al. (2016) found it in their model of major depression. Here, it had to do with the threshold for tipping into another stable part of the mental state space. *Connection strengths* between the causally interacting symptoms, that are tweaked in mathematical models and simulations, are theoretically imagined to influence the speed and dynamics of initial symptom activation through the network. *Bridge symptoms* shared by multiple symptom networks allow for the spreading of activation from one network or cluster to another. In a network approach, bridge symptoms explain on the one hand the often-observed comorbidity of mental disorders (Fried et al., 2017, p. 2), and on the other the fact that research efforts have failed to find “zones of rarity” between mental disease categories (Cooper, 2013). *Critical slowing down* is investigated as a

predictive marker for approaching a tipping point (van de Leemput et al., 2014). Critical slowing down refers to the increase in the time it takes for a complex adaptive system to return to its equilibrium state after a perturbation. In mental health care, the phenomenon is of interest regarding predicting or preventing the onset of or relapse into, for example, a depressed state.

This network approach to mental disorders is emergent. The material investigative practices associated with it are under development. Conceiving of mental health, distress, and disorders as trajectories through a mental state space, with threshold phenomena and tipping points between more or less stable attractors, construes minds as uniquely individual and historical entities. The shape and height of the thresholds come into focus as a target for the building of psychological robustness and resilience. The dependency of the thresholds on the history of the system (hysteresis) suggests that they are built from a range of meaning-generating, social, and cultural developmental resources during a person's life history.

It is not a question about the reality of experiences of mental distress. These abound and are inevitable responses to the perturbations and challenges of life. The key question is about the interactions between the way in which people are described, classified, and named by experts and institutions on the one hand and the people so classified on the other. The classic philosophical formulation of this problematic is Ian Hacking's *looping effects* (Hacking, 1995a, 2007; Haslam, 2016). Alternative ways of understanding and performing mental health and disorder warrant investigating as resources for a much-needed pushback against what Nick Haslam (2016) has called *concept creep* associated with the psychiatrization of society: rising rates of mental illness, increasing rates of mental health service utilization, and evidence of over-diagnosis, over-treatment, and over-prescription (Haslam et al., 2021).

Conclusion

Wherever we went for our material–discursive reconstruction and infrastructural inversion of psychometric instruments, we met psychophysicists engaged in attempts to construct a foundation from which psychology could become scientific. These psychophysicists demonstrated a strong commitment to a particular view of science. Only knowledge that derives from observation through measurement and quantification is science proper. Having been adapted to allow for measurement in the psychological and social realm, physics' theory and practice of “fundamental measurement of extensive quantities,” including its associated linear algebra-derived mathematics, has been psychometrics' gold standard. These are features of an epistemic culture that current mental health surveys inherited from their psychophysical predecessors. These methods resulted from an ambition to develop psychology as a correlative science proper, based on an adapted form of fundamental measurement.

The mathematization of mind in psychometrics and the digitalization of data infrastructures have contributed to a concealment of the shifts and displacements that were constitutive for the conditions of possibility for mental health surveys, for the ways in which these are culturally intelligible and seemingly irreproachable, and hence, for their role as knowledge bank for knowledge-based mental public health policies.

These developments have come at a price, though. The subject that can freely speak their mind has been silenced and replaced by forced choice methods regarding both items and response formats. The day-to-day anchoring of the validity of psychometric instruments in the clinical experience of psychiatrists has been lost and replaced with indirect chains of validation against other psychometric instruments derived from fluid¹³ ICD or DSM disease categories. The transfer of psychometric instruments from the psychiatric clinic to general populations introduced unresolved probabilistic problems concerning false positives and false negatives. These are issues that cannot be resolved in single-shot surveys that do not break population aggregates down to the level of individuals. The use of adjectives derived from psychiatric classification systems for the kind of distress experienced contributes to the cultural scaffolding of the regulatory ideal of an autonomous, self-mastering human subject. This use of psychiatric terms renders those who “self-report” mental distress (through forced choice methods) as human subjects with a deficit.

The authority of objectivist science, with which mental health surveys can publish their results, further scaffolds people’s mental ill-health as an important object for public health governance and public health interventions. Mental health surveys prime the public debate semantically and semiotically through their use of psychiatric adjectives and they provide numerical anchors for the seriousness of the problem. However, the historical processes that have given rise to the space in which mental health surveys can exist displaced or stand in the way of alternative understandings of mental distress that could serve as cultural resources for people’s self-understanding, as alternative descriptions under which one could live.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Ger Wackers  <https://orcid.org/0000-0002-8467-6029>

Notes

1. Loop time is the time it takes for a recursive loop or cycle to be completed. The term suggests an iterative process in which the intervention influences the production of data in the next survey.
2. *Digitizing*, as a verb and operation, is a form of *discretization* (parsing into discrete units) of a continuous process. A scale used to measure a continuous process is *digitized* when it offers response alternatives in the form of discrete units, like fingers or digits, often with numbers attached to them. *Digitized* is therefore not synonymous with *digitalized*, understood as computerized or computer-assisted, with computers operating a binary code.
3. For a historical and sociological account of statistics in Britain around the turn of the 19th to the 20th century, see Donald Mackenzie (1981). For a historical and sociological account of the rise of psychology in turn of the century Britain, see Rose (1985). On Pearson, see Norton (1978). On Spearman, see Norton (1979).

4. Each of these terms is historically and discursively constituted and contested. Yet, they are often rendered unproblematic through specific *operationalizations* that set them up as independent or dependent variables in multiple regression analyses.
5. The tests always involved a series of comparisons between two physical stimuli selected from a continuous series. Compare two balls with slightly different weights: which one is the heavier one, or are they the same weight? Grey-scales consist of black dots on white paper but with different densities: which one is the greyer one, or are they the same? The same was done for “pitch of sound,” that is, for differences in tone, not loudness. The smallest difference in weight, greyness, or pitch that the participant could distinguish as being different was the “just-noticeable difference (JND).” Differences smaller than the JND would be perceived as the same by the participant.
6. Spearman found that ability to discriminate “pitch of sound” could serve as a proxy for intelligence.
7. The notion of JNDs was retained in Thurstone’s (1927) “Law of Comparative Judgement.”
8. Judgments of “more or less” are ordinal, in the sense that they allow for a rank ordering of the judgments in relation to each other. Thurstone assumes here that the degrees of difference in “more or less” judgments are quantitative and therefore measurable. The assumption that an attribute is quantitative from the premise that it is ordinal constitutes, according to Michell (2006, 2009), psychometricians’ fallacy that has occupied a central place in the paradigm of psychometrics.
9. Questionnaires had a long history in opinion polling and marketing research before Thurstone adopted and adapted them as a foundational building block for a scientific psychology. That is, for a psychology striving to become a science proper, after the model of physics, through experiment, measurement, and quantification.
10. Symptom checklists had a long history in psychiatry before Thomas Verner More made his checklist data available to Thurstone. In the last quarter of the 19th century, the German psychiatrist Emil Kraepelin used a list of checklist symptoms, printed on diagnostic cards, to systematically document symptoms presented by the patients he saw in a mental hospital.
11. For a history of the evolution of the classification of psychiatric disorders, see Suris et al. (2016).
12. In what we here call a “single-shot survey,” there is only one “measurement act” through which data are produced. There is no repetition of the measurement to increase precision of measurement or discover incidental errors. Neither is there a follow up with a different kind of instrument or assessment device (e.g., an interview) in a multimodal assessment setup.
13. “Fluid” refers to the well-documented “changeability” of psychiatric disease categories over time. Not only has the number of possible psychiatric diagnoses multiplied over time, their contents, criteria, and causes have also shifted over time. The differences between DSM and ICD have not been resolved. Some mental disorders, like “multiple personality disorder” (Hacking, 1995b) or “hysteric and epileptic fugue” (Hacking, 1998), have enjoyed a short-lived, transient existence; they have come and gone. Sets of correlating symptoms that were once thought to constitute an autonomous mental disorder vanished into many other places where they were reordered into new disease categories.

References

- Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *A handbook of social psychology* (pp. 798–844). Russell and Russell.
- American Psychiatric Association. (1952). *Diagnostic and statistical manual. Mental disorders (DSM-1)*.

- Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.
- Barrett, L. F. (2018). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt/Mariner Books.
- Bech, P. (2012). *Clinical psychometrics*. John Wiley & Sons.
- Bijker, W. E. (1997). *Of bicycles, bakelites and bulbs: Toward a theory of sociotechnical change*. The MIT Press.
- Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge University Press.
- Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology, 64*(9), 1089–1108. <https://doi.org/10.1002/jclp.20503>
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry, 16*(1), 5–13. <https://doi.org/10.1002/wps.20375>
- Bowker, G. C., Timmermans, S., Clarke, A. E., & Balka, E. (Eds.). (2015). *Boundary objects and beyond: Working with Leigh Star*. The MIT Press.
- Bridgman, P. W. (1927). *The logic of modern physics*. Macmillan.
- Campbell, N. R. (1928). *An account of the principles of measurement and calculation*. Longmans, Green, and Co.
- Cooper, R. V. (2013). Avoiding false positives: Zones of rarity, the threshold problem, and the clinical significance criterion. *Canadian Journal of Psychiatry, 58*(11), 606–611. <https://doi.org/10.1177/070674371305801105>
- Cramer, A. O. J., Van Borkulo, C. D., Giltay, E. J., Van der Maas, H. L. J., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLOS ONE, 11*(12), Article e0167490. <https://doi.org/10.1371/journal.pone.0167490>
- Danziger, K. (1991). *Constructing the subject*. Cambridge University Press.
- Danziger, K. (1997). *Naming the mind: How psychology found its language*. SAGE.
- Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins symptom checklist (HSCL): A self-report symptom inventory. *Behavioral Science, 19*(1), 1–15. <https://doi.org/10.1002/BS.3830190102>
- Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.
- Fried, E. I., van Borkulo, C. D., Cramer, A. O. J., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: A review of recent insights. *Social Psychiatry and Psychiatric Epidemiology, 52*, 1–10. <https://doi.org/10.1007/s00127-016-1319-z>
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life*. Cambridge University Press.
- Gould, S. J. (1996). *The mismeasure of man* (2nd rev. ed.). W. W. Norton & Company. (Original work published 1981)
- Hacking, I. (1995a). The looping effects of human kinds. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: An interdisciplinary approach* (pp. 351–383). Oxford University Press.
- Hacking, I. (1995b). *Rewriting the soul*. Princeton University Press.
- Hacking, I. (1998). *Mad travelers: Reflections on the reality of transient mental illnesses*. Harvard University Press.
- Hacking, I. (2002). *Historical ontology*. Harvard University Press.
- Hacking, I. (2007). Kinds of people: Moving targets (British Academy Lecture, 2006 Lectures). *Proceedings of the British Academy, 151*, 285–318. <https://www.thebritishacademy.ac.uk/documents/2043/pba151p285.pdf>

- Haslam, N. (2016). Looping effects and the expanding concept of mental disorder. *Journal of Psychopathology*, 22, 4–9. https://www.jpsychopathol.it/wp-content/uploads/2016/02/02_Art_INTRO_Haslan1.pdf
- Haslam, N., Jesse, S. Y., & De Deyne, S. (2021). Concept creep and psychiatrization. *Frontiers in Sociology*, 6, Article 806147. <https://doi.org/10.3389/fsoc.2021.806147>
- Healy, D. (1997). *The antidepressant era*. Harvard University Press.
- Knorr Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.
- Lipman, R. S., Covi, L., & Shapiro, A. K. (1979). The Hopkins symptom checklist (HSCL): Factors derived from the HSCL 90. *Journal of Affective Disorders*, 1(1), 9–24. [https://doi.org/10.1016/0165-0327\(79\)90021-1](https://doi.org/10.1016/0165-0327(79)90021-1)
- López-Muñoz, F., & Alamo, C. (2009). Monoaminergic neurotransmission: The history of the discovery of antidepressants from 1950s until today. *Current Pharmaceutical Design*, 15(14), 1563–1586. <https://doi.org/10.2174/138161209788168001>
- Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new scale type of fundamental measurement. *Journal of Mathematical Psychology*, 1(1), 1–27. [https://doi.org/10.1016/0022-2496\(64\)90015-X](https://doi.org/10.1016/0022-2496(64)90015-X)
- Mackenzie, D. A. (1981). *Statistics in Britain 1865–1930. The social construction of scientific knowledge*. Edinburgh University Press.
- McNally, R. J., Robinaugh, D. J., Wu, G. W. Y., Wang, L., Deserno, M. K., & Borsboom, D. (2014). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science*, 3(6), 836–849. <https://doi.org/10.1177/2167702614553230>
- Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge University Press.
- Michell, J. (2006). Psychophysics, intensive magnitudes, and the psychometricians' fallacy. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 37(3), 414–432. <https://doi.org/10.1016/j.shpsc.2006.06.011>
- Michell, J. (2009). The psychometricians' fallacy: Too clever by half? *British Journal of Mathematical and Statistical Psychology*, 62(Pt 1), 41–55. <https://doi.org/10.1348/000711007X243582>
- Miller, G. A. (1974). Stanley Smith Stevens: 1906–1973. *The American Journal of Psychology*, 87(1/2), 279–288. <https://www.jstor.org/stable/1422022>
- Noori, H. R. (2014). *Hysteresis phenomena in biology*. Springer.
- Norton, B. J. (1978). Karl Pearson and statistics: The social origins of scientific innovation. *Social Studies of Science*, 8(1), 3–34. <https://doi.org/10.1177/030631277800800101>
- Norton, B. (1979). Charles Spearman and the general factor in intelligence: Genesis and interpretation in the light of sociopersonal considerations. *Journal of the History of the Behavioral Sciences*, 15(2), 142–154. [https://doi.org/10.1002/1520-6696\(197904\)15:2<142::AID-JHBS2300150206>3.0.CO;2-X](https://doi.org/10.1002/1520-6696(197904)15:2<142::AID-JHBS2300150206>3.0.CO;2-X)
- Parloff, M. B., Kelman, H. C., & Frank, J. D. (1954). Comfort, effectiveness and self-awareness as criteria of improvement in psychotherapy. *American Journal of Psychiatry*, 111(5), 343–352. <https://doi.org/10.1176/ajp.111.5.343>
- Pearson, K. (2007). *The grammar of science* (3rd ed.). Cosimo Classics. (Original work published 1911)
- Rose, N. (1985). *The psychological complex: Psychology, politics and society in England 1869–1939*. Routledge & Kegan Paul.
- Rose, N. (2019). *Our psychiatric future*. Polity Press.

- Schille-Rognmo, M. (2017). *Ungdata, mental health and gender differences: A study of gendered mental health re-enactments in Ungdata's dLTC youth surveys* [Master's thesis, UiT The Arctic University of Norway]. <https://hdl.handle.net/10037/11215>
- Shapin, S. (2010). *Never pure: Historical studies of science as if it was produced by people with bodies, situated in time, space, culture, and society, and struggling for credibility and authority*. Johns Hopkins University Press.
- Spearman, C. (1904a). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72–101. <https://doi.org/10.2307/1422689>
- Spearman, C. (1904b). "General intelligence," objectively determined and measured. *American Journal of Psychology*, *15*(2), 201–293. <https://doi.org/10.2307/1412107>
- Stevens, S. S. (1935a). The operational basis of psychology. *American Journal of Psychology*, *47*(2), 323–330. <https://doi.org/10.2307/1415841>
- Stevens, S. S. (1935b). The operational definition of psychological terms. *Psychological Review*, *42*(6), 517–527. <https://doi.org/10.1037/h0056973>
- Stevens, S. S. (1942). Operationism. In D. D. Runes (Ed.), *The dictionary of philosophy* (pp. 219–220). Philosophical Library.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science, New Series*, *103*(2684), 677–680. <https://doi.org/10.1126/science.103.2684.677>
- Stevens, S. S. (1959). The quantification of sensation. *Daedalus*, *88*(4), 606–621. <https://www.jstor.org/stable/20026531>
- Stevens, S. S. (1960). On the new psychophysics. *Scandinavian Journal of Psychology*, *1*(1), 27–35. <https://doi.org/10.1111/j.1467-9450.1960.tb01278.x>
- Suris, A., Holliday, R., & North, C. S. (2016). The evolution of the classification of psychiatric disorders. *Behavioral Sciences*, *6*(1), Article 5. <https://doi.org/10.3390/bs6010005>
- Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review*, *34*(4), 273–286. <https://doi.org/10.1037/h0070288>
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*(4), 529–554. <https://doi.org/10.1086/214483>
- Thurstone, L. L. (1934). The vectors of mind. *The Psychological Review*, *41*(1), 1–32. <https://doi.org/10.1037/h0075959>
- Turner, M. (2014). *The origin of ideas: Blending, creativity and the human spark*. Oxford University Press.
- Uhlenhuth, E. H. (1975). Evaluation of behavior change and therapeutic effectiveness. In D. X. Freedman & J. E. Dyrud (Eds.), *American handbook of psychiatry* (pp. 938–954). Basic Books.
- van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., van Nes, E. H., Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K. S., van der Maas, H. L. J., Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *PNAS*, *111*(1), 87–92. <https://doi.org/10.1073/pnas.1312114110>
- World Health Organization. (1948). *International classification of diseases* (Rev. 6).
- Wider, A. (1948). *Cornell index and manual*. Psychological Corporation.
- Winther Jørgensen, M. (2015). Patient-centred decision making? Biocitizens between evidence-based medicine and self-determination. *Evidence & Policy*, *11*(3), 311–329. <https://doi.org/10.1332/174426415X14381755121530>

Author biographies

Ger Wackers earned a medical degree and subsequently a PhD in science and technology studies from Maastricht University in the Netherlands. He currently holds a position as associate professor in the Department of Health and Care Sciences of the Faculty of Health Sciences at UiT The Arctic University of Norway. Working at the intersection of STS and health sciences, he has previously addressed issues in end-of-life-care at home. His current research focuses on the study of technologies and practices that aim to capture people's minds. Recent publications include "Into the Wild Country: Epistemic Issues in Professional Guidelines for Palliative Sedation in End-of-Life Care," in *Evidence & Policy* (2015) and "Making a Place for Dying at Home: Liminality, Territoriality and Care at the End of Life," in B. Pasveer, O. Synnes and I. Moser (Eds.), *Ways of Home Making in Care for Later Life* (Palgrave Macmillan, 2020).

Marthe Schille-Rognmo is a PhD student and lecturer at the Department of Health and Care Sciences of the Faculty of Health Sciences at UiT The Arctic University of Norway. Her research interests focus on the use of diagnostic terminology and psychometrics in population health studies, with a specific emphasis on digitalized questionnaire-based surveys targeting youngsters.