Uher, J. (2021b). Quantitative psychology under scrutiny: Measurement requires not result-dependent but traceable data generation. *Personality and Individual Differences, 170,* 110205. https://doi.org/10.1016/j.paid.2020.110205

Original article

Quantitative psychology under scrutiny: Measurement requires not result-dependent but traceable data generation

Jana Uher *

[1] *School of Human Sciences, University of Greenwich*
[2] *London School of Economics*

* Correspondence:
University of Greenwich
School of Human Sciences
Old Royal Naval College, Park Row
London SE10 9LS, United Kingdom
Telephone: +44(0)20-8331 9654
E-mail: mail@janauher.com

Abstract

Various lines of critique of quantitative psychology, well-established and new, are used to trace along the field's typical steps of research a complex network of misconceptions and fallacies codified in psychological jargon. The article explores what constructs actually are, why they are needed in psychology, fallacies and challenges in construct research, and the crucial role of language. It shows how common misconceptions of language and concepts mislead psychologists to conflate phenomena, qualities, quantities and constructs with one another and with their semiotic encodings in terms, variables and scores. The article clarifies the conceptual relations between nomological networks, representation theorems and psychometric modelling. It reveals conflations of disparate notions of causality and unobservability, and erroneous equations of nomological networks with semantic networks, description with explanation, and measurement theories with explanatory theories. Instead of establishing causal measurand-result relations, common practices match data generation to the results rather than the phenomena and properties studied. Mathematical meaning for scores is often created from differences between individuals and between different phenomena and properties, which constitute mere conceptual entities and cannot reflect magnitudes attributable to individuals. This entails biased inferences on the actual study phenomena and shows that replicability problems may be even larger than assumed.

Keywords
Measurement; Assessment; Psychometrics; Replicability; Test interpretation; Test scores; Validity; Instrument development

Highlights
- Psychological measurement jargon codifies numerous fallacies about key concepts.
- Nomological networks, representation theorems and psychometric modelling are linked.
- Data generation is often aligned to results not to causal measurand-result connections.
- Meaning of 'scores' is derived from differential analyses rather than the measurands.
- Replicability problems in psychology are therefore even larger than currently assumed.

1. Introduction

Quantitative psychology has always met with criticism (Michell, 1999; Trendler, 2009, 2013). Replication and validity crises are indicators of persisting conceptual problems that current initiatives (e.g., robust statistics, open science) cannot solve (Maraun et al., 2009; Smedslund, 2016). There is an increasing hunger for understanding what is wrong with psychological measurement. The list of problems is long. Much discussed are the (alleged) lack of substantive theories about psychology's study phenomena to inform measurement theories (Haig & Borsboom, 2008; Kyngdon, 2011), implicit but untested assumptions of quantitative properties in psychical phenomena (Michell, 2008), lack of representation theorems (Kyngdon, 2008; Uher, 2018a), erroneous analogies with physical measurement (Trendler, 2019; Uher, 2020a), and a thick web of fallacies surrounding constructs, language and concepts (Maraun & Gabriel, 2013; Slaney & Garcia, 2015).

Still, most psychologists ignore the growing criticism of their measurement fundamentals—will it just pass or is quantitative psychology in real danger of collapsing like a house of cards? Many colleagues may, indeed, be still unaware of these critiques because mainstream journals filter what is deemed acceptable so that critical papers appear in multi-disciplinary and methodology journals not commonly read by psychologists. But even if known, they are often considered irrelevant given many quantitative psychologists' reluctance to explore their own philosophy of science and, without pertinent knowledge, the points made may not be readily understood.

Most confusion, however, is caused and maintained by psychological jargon because it codifies numerous conceptual fallacies underlying psychological measurement that are tightly interrelated with one another and with common misconceptions of language and concepts. This complicates their recognition in one's own thinking and writing so that, when using common jargon, even psychologists who critically discuss fallacies and misconceptions and their detrimental impact often still rely implicitly on them in their own works. Understanding these fallacies requires critical analyses of the current psychological measurement terminology. This is a challenging task, prone to misunderstandings, because everyone intuitively relies on their habitual language and alternative terminologies are not yet established.

Some supposed radical reconceptualisation is needed (Barrett, 2008). But what is needed is not radical. Needed is a shift in thinking toward self-reflection and preparedness to critically analyse one's own language, implicit concepts and presumptions—core skills every psychologist should have. Progress also requires embracing fundamental criticism as an ordinary element of science that is essential for stimulating new developments rather than a threat to be silenced—core beliefs every scientist should have. Real progress also requires not ever more complex statistics (Barrett, 2003; Lamiell, 2019) but much more conceptual thinking about psychology's study phenomena (Fahrenberg, 2019; Valsiner, 2017).

This article applies a transdisciplinary paradigm for research on individuals in which various lines of substantive theory from different disciplines, historical and new, have been integrated into coherent frameworks, among them lines needed to address key problems in psychological measurement and to reflect on scientists' own role in research processes. For simplicity, a few relevant concepts are briefly introduced where needed (see footnotes for definitions, more information and references), the paradigm's overall structure will be outlined later (3.3). The aim is to make the central lines of criticism of psychological measurement—and further ones not yet well considered—accessible to mainstream psychologists in a holistic way by explaining them along typical research processes, focussing on personality and individual differences. *Jargon codifying fallacies and erroneous conflations is italicised* and explained subsequently, though not always immediately because necessary KEY CONCEPTS, CAPITALISED below, must first be introduced.

2. Under scrutiny: Key steps in quantitative psychology

Psychologists seek to establish causal explanatory theory for phenomena observable in individuals. This journal's readers aim to explore *structures of personality differences, underlying processes causing their emergence,* and practical applications. Key steps of research typically involve *1)* defining study constructs and terms, *2)* operationalising constructs in item variables and *3)* developing *multi-stage rating scales as measuring instruments*. For data collection, *4)* persons are asked to complete these scales. Psychologists then *5) recode the answer categories into numbers* to analyse *6)* reliability and *7)* validity. They use *8)* psychometric modelling of latent variables to create *scores on the construct level* before *9)* interpreting their findings. These steps are now scrutinised, highlighting conceptual fallacies where these may occur.

2.1 Defining constructs and terms

The APA Dictionary of Psychology[1] defines 'extraversion' as an element of personality models "characterized by an orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience. Extraversion *is* a broad personality *trait* and, like introversion, *exists on a continuum of attitudes and behaviors*. *Extraverts are* relatively outgoing, gregarious, sociable, and openly expressive". This definition mentions *traits* and various psychical and behavioural phenomena—core phenomena studied in psychology. How are these defined?

2.1.1 Substantive theories about study phenomena: Why should they be considered?

Although considered well-established, theoretical definitions about *traits*, psyche and *behaviour* are seldom discussed and surprisingly vague (Zagaria et al., 2020). "The profound lack of descriptive theories of *human behaviour* is arguably the single most important problem facing the scientific measurement of *psychological systems*" (Kyngdon, 2011, p. 493). Psychologists focus instead on understanding the constituent properties of quantity and on *identifying in their measurements at least interval-scaled units* (e.g., using psychometric modelling). However, providing compelling evidence for quantitative structures in *psychological systems* presupposes substantive theory; "but theories in the relevant content areas, such as those concerned with … personality, … are not yet able to connect the hypothesized additive structure of *latent attributes* to *identifiable features of test items*" (Michell, 2008, p. 21). Relevant concepts exist—but are still largely ignored.

2.1.2 Psyche and behaviour: Why should they be differentiated?

Psychologists often speak of 'covert' and 'overt' *behaviours*, implying psyche and behaviour would be phenomena of the same kind—likely misguided by the immaterial, transient and processual nature of both. In the paradigm applied here (3.3.), PSYCHE is defined as the "entirety of the phenomena of the immediate[2] experiential reality both conscious and non-conscious of living organisms" (Uher, 2015c, p. 431), highlighting their "non-physical" nature, exclusive accessibility to the individual itself and inaccessibility by others (Uher, 2016a). BEHAVIOURS are defined as the "external changes or activities of living organisms that are functionally mediated by other external phenomena in the present moment", highlighting their physical nature and public accessibility (Uher, 2016b, p. 490).

[1] https://dictionary.apa.org/extraversion
[2] Immediacy here indicates absence of phenomena mediating their perception.

Many psychical and behavioural phenomena are transient and processual. Of PROCESSUAL ENTITIES, only a part exists at any moment; their determination thus requires knowledge of previous occurrences obtained from generalisation and abstraction over time (Whitehead, 1929). Abstractions of transient psychical phenomena (e.g., thinking) lead to pertinent concepts, beliefs and knowledge, which are psychical phenomena as well but different from—and necessarily more stable than—those transient phenomena that they are about (conceptualised as EXPERIENCINGS versus EXPERIENCES; Uher, 2015e, 2016a). From experiences made, humans develop concepts of themselves and their world that they use to anticipate future events and to choose among alternative actions (Kelly, 1955; Valsiner, 2012). Humans can develop concepts about anything that they can perceive or conceive of—concrete phenomena observable in the world (e.g., trees, behaviours) and concrete experiencings perceived in oneself (e.g., ongoing emotions) but also other concepts and abstract ideas that they may have developed about themselves and their world (e.g., 'personality', 'mind', 'god'; Uher, 2015a).

2.1.3 What actually are constructs and why are they needed?

Concepts are theoretically constructed by humans, therefore also called CONSTRUCTS[3]. Although constructs are among psychology's most frequent study phenomena (Maraun et al., 2009), some lament "we do not know what constructs are, that is, we have rarely come across a clear description of what something should be like in order to deserve the label 'construct'. Constructs, as far as we are concerned, are truly shrouded in mystery, and not in the good old scientific sense that we currently don't know what they are, but will know when we're finished doing the relevant research, but in the sense that we don't really know what we are talking about in the first place" (Borsboom et al., 2009, p. 150).

This statement is astounding given well-established foundations like Kelly's (1955) Personal Construct Theory, published in a two-volume book already 65 years ago. Formalised in a basic postulate and 11 corollaries, it elaborates how humans develop and use constructs. Specifically, humans formulate from their current interpretation of reality theoretical ideas from which they derive hypotheses about possible future events (basic postulate). Humans use concrete experiences to check their constructs' appropriateness for making viable predictions (experience corollary), and they strengthen, modulate and refine their constructs accordingly in iterative processes (modulation corollary). That way, humans develop complex systems of interrelated yet also domain-specific constructs, organised hierarchically by their level of generality (organisation corollary). This theory lays essential foundations for construct research; indeed, the links to personality taxonomies, nomological networks and validity are obvious (2.2.1, 2.7). Kelly's idea of humans proceeding like naïve scientists to recognise regularities in the world links everyday conceptualising even explicitly with scientific conceptualising. Why is this important theory largely ignored in quantitative psychology?

The above statement also reveals misconceptions of concepts in general—and thus of generalisation, abstraction and categorisation. In GENERALISATION, multiple similar concrete entities (e.g., acts of vocalisation) are—in our minds, thus conceptually—replaced with a single conceptual entity (e.g., the idea of vocalisation). Through ABSTRACTION, we reduce the complexity of entities by emphasising some of their details and deemphasising others (e.g., ideas of chatting, yelling and quarrelling emphasise different aspects of human vocalisation). In CATEGORISATION, we group generalised and abstracted entities by similarity (e.g., yelling and quarrelling as aggression).

[3] For a suggestion to differentiate constructs from concepts, see Markus (2008).

Importantly, we can form concepts on all levels of generalisation, abstraction and categorisation. But in any case, CONCEPTS are ideas, thus psychical phenomena that refer to something and these ideas are not the same as the REFERENTS that they are about. The construct 'extraversion' is a theoretical idea that (following the APA definition) refers to various concrete entities perceivable at a moment, such as behaviours (e.g., expressive acts) and, within oneself, experiencings (e.g., feeling energised)—but also to more abstract concepts that people have developed of themselves and others (e.g., attitudes about others), thus to conceptual entities. Hence, 'extraversion' refers to concrete transient and processual phenomena of behaviour and experiencing but also to conceptual phenomena that people have developed of these former through generalisation and abstraction over time and that are therefore more temporally stable. This highlights that 'extraversion' refers to a conglomerate of various kinds of study phenomena each differing in accessibility (Uher, 2018b). None of these phenomena ever occur all at once in an individual. Their unity as an entity is established only conceptually in the abstract idea of 'extraversion', which thus constitutes a conceptual entity—a construct.

The construal of constructs allowed scientists to turn abstract ideas into entities, thereby making them conceptually accessible to empirical study. But this ENTIFICATION misguided psychologists to overlook their constructed nature—although their labelling as 'constructs' could not make this any clearer (Slaney & Garcia, 2015).

2.1.4 Defining constructs: Inherent challenges

A frequent criticism is that psychological constructs are ill-defined, ambiguous, overlapping, circular and context-dependent ('common-or-garden'; Maraun, 1998; Zagaria et al., 2020). Why is that so? One reason is that many psychological constructs refer to people's beliefs, ideas and concepts about their experiences in the world. Such personal constructs are permeable and iteratively adapted to new experiences, thereby undergoing permanent change and development (modulation corollary; Kelly, 1955). This adaptability is essential to maintain their viability for enabling anticipatory control in a complex world facing an unknown future. The inherent changeability of constructs limits the possibilities for establishing clear-cut scientific definitions that are valid across time and contexts as possible for physical phenomena. Hence, psychologists are struggling with their definitions not because of inferior scientific concepts but because many of their study phenomena are constantly changing in themselves. Psychological constructs must reflect this and be changeable as well (see Hanfstingl, 2019, on assimilation and accommodation processes in scientific construct development). This constitutes a profound difference to the physical sciences and poses fundamental challenges for attempts to quantify psychological study phenomena (Uher, 2020a, 2020b).

2.1.5 The role of language in construct development

In conceptualising, language plays a key role because it enables us to represent perceivable phenomena (e.g., behaviours, feelings) in single words (e.g., 'socialising', 'energised'). This function of language is reflected in the hypothesis that "[t]hose individual differences that are most salient and socially relevant in people's lives will eventually become encoded into their language; the more important such a difference, the more likely it is to become expressed as a single word", which underlies the development of popular personality models (e.g., 16PF, BigFive; John et al., 1988, p. 174; Uher, 2013, 2015d).

Words allow us to make concrete entities (e.g., behavioural acts) independent of their immediate perception and to abstract them into objects of consideration—conceptual entities. This allows us to refer to concrete referents also in their absence, abstracted across time and contexts. We also create words that have only abstract referents, such as phenomena distant from immediate perception (e.g., 'extraversion') or imperceptible in themselves (e.g., 'quantity'). That is, every word denotes a concept in itself (Vygotsky, 1962).

2.1.6 Fallacies from linguistic abstractions: What are *traits*?

Single-word labels for conceptual entities often mislead us to treat abstractions as concrete entities (REIFICATION; HYPOSTATIC ABSTRACTION; Peirce, 1958, CP 4.227). Then we treat abstract terms and the concepts they denote as constituents of reality, thereby attributing them an ontological status (Maraun & Gabriel, 2013; Slaney & Garcia, 2015). This FALLACY OF MISPLACED CONCRETENESS (Whitehead, 1929) underlies many psychologists' beliefs that *trait* constructs are not just summary descriptions of their particular reference phenomena but, indeed, entities inside individuals and causal to the phenomena from which they were first derived. This TRAIT FALLACY entails fundamental circularity in explanation because *traits* serve as both cause and description of phenomena (Uher, 2013).

Failure to distinguish phenomenal observations from causal processes also underlies the erroneous assumption that structures of between-individual differences could reflect within-individual structures (Molenaar, 2004). Common psychological jargon blurs this distinction, thereby fortifying this fallacy; for example, 'personality structure' denotes both population-level patterns and individual-level configurations, although these require different levels and approaches of analysis (variable-oriented versus individual-oriented). The psychological term *traits* variously refers to *a)* individual differences variables, *b)* single scores on such variables; *c)* constructs encoded in such variables and scores, or *d) psychobiological mechanisms assumed to causally underly* a), b) and c) (see 2.9.1).

2.1.7 Everyday language and scientific language: Inevitable interferences in psychology

Such fallacies are widespread also in everyday language. Through socialisation, we acquire a comprehensive everyday psychology including pertinent vocabulary (e.g., Allport and Odbert's 17,953 person-descriptive words). It forms an essential part of our minds and therefore cannot be put aside when doing research (Valsiner, 2012). Moreover, psychologists investigate phenomena of everyday experience that are largely accessible only through language, and must therefore rely on and analyse participants' verbal reports (e.g., questionnaires, interviews). Consequently, psychologists cannot invent new concepts and a scientific language that are completely unrelated to everyday psychology—unlike physicists. Influences on scientific psychology are inevitable and complex (Uher, 2013, 2015c). Psychologists must therefore critically analyse and reflect on their own use of terms and concepts—an established standard in so-called qualitative methods (Flick, 2008).

2.1.8 Language-based research: Inherent challenges

Language is essential for all sciences because scientists must describe and explain their study phenomena. In psychology, explorations of many study phenomena are even intimately bound to languages. Psychologists must therefore, first and foremost, differentiate their study phenomena from the concepts, terms and methods used to explore them, as reflected in the terms PSYCHICAL versus PSYCHOLOGICAL[4]. The common labelling of both study phenomena and scientific tools as *psychological* (e.g., *psychological systems*) blurs this essential distinction. This is problematic because the scientific concepts used to explore psychical phenomena constitute psychical phenomena in themselves, thus they do not exist independently of the objects of research. This constitutes one of psychology's greatest challenges (Uher, 2020b, 2021).

[4] From Greek -λογία, -logia for body of knowledge. This differentiation is used in many languages but not commonly in the English (Uher, 2016a).

In our thinking, words are so inextricably intertwined with the phenomena they denote that we often conflate them. In daily life, this facilitates acting in a complex world. But in science, it entails conceptual fallacies. Their exploration requires concepts of language.

2.1.9 What is language? Conceptualising a key phenomenon

Language is a complex phenomenon in itself. As any sign system, language has three components, *(1)* the SIGNIFYER (e.g., vocalisations like [t͡ʃaɪld] or ink printed on paper forming *CHILD*) that symbolically represents *(2)* the REFERENT, the actual object of consideration to which it refers (e.g., a child nearby), and *(3)* the MEANING that both have for the sign-using persons in a given context (e.g., the concept of children in general or knowledge of a specific child). Signifiers must always be publicly accessible, thus external to individuals' bodies (e.g., printed text, uttered vocalisations). Referents can be everything humans can perceive or conceive, thus can be located internally to individuals' bodies (e.g., feelings, brains) or externally (e.g., children, an object's length). Signifier and referent are linked through the third component, meaning[5]. Meaning is a psychical phenomenon, thus always part of individuals' mind (Salvatore, 2019). This explains why we seldom become aware[6] that a sign's functionality emerges only from the tight interrelations among its three distinct components (Uher, 2015a, 2018a) and why sign systems (e.g., language) are inseparable from the individuals using them (Valsiner, 2012, 2017).

2.1.10 Misconceptions of language and concepts

Lack of differentiation of a sign's three components underlies key fallacies in psychological measurement (Uher, 2021). A primary fallacy is to overlook that what is written or said (signifier) is not the entire sign (SIGNIFIER-SIGN EQUATION). This involves the fallacy to assume that the meaning of words would be inherent to their signifiers (SIGNIFIER-MEANING CONFLATION), such as the widespread belief that standardising signifiers (e.g., writing down item wordings) would allow standardising also their meanings. But items are typically abstract and generalised. Without specifying their intended meanings, raters must rely on their everyday words and concepts, which are, however, fuzzy and context-sensitive (Hammersley, 2013). In everyday life, we intuitively adapt meanings to contexts; we all know that 'small' does not indicate the same magnitude when referred to houses, countries or planets. This is good enough for day-to-day understanding but not for scientific purposes. Signifier-sign equation also involves the fallacy to mistake the signifiers for the referents that they denote (SIGNIFIER-REFERENT CONFLATION; 2.1.6). But signifiers are largely arbitrary and therefore cannot reveal the referents they denote.

Given these fallacies, it is unsurprising that—even if all psychometric criteria are met—rating items are interpreted differently, thus representing broad fields of meaning (Lundmann & Villadsen, 2016; Rosenbaum & Valsiner, 2011; Uher & Visalberghi, 2016). Importantly, in an assessment, every rater considers only part of an item's field of meaning, so that different raters consider—for the same item—different referents. This makes their data incomparable, thus precluding averaging and adding as commonly done (Uher, 2018a).

2.1.11 Psychological terms: Referents with nested conceptual structures

Constructs can refer to all kinds of perceivable or conceivable entities and on all levels of generalisation, abstraction and categorisation (2.1.3.). This entails unparalleled proliferation,

[5] Therefore, sign systems are conceived as SEMIOTIC REPRESENTATIONS (Uher, 2015c, 2015a).

[6] As adults, we become acutely aware of these three components when starting to learn a foreign language because this involves acquiring knowledge about the specific signifiers used in that language (e.g., different word stems, letters, pronunciations), their relations to specific referents and the meanings they have in given contexts.

complexity and thus changeability in the constructs created. Constructs—and the terms denoting them (i.e., their signifiers)—are developed in both science and everyday life. Psychology is a special science because it explores, amongst others, the personal constructs and language terms that laypeople intuitively develop. This entails that psychological terms and constructs overlap and intermingle with those from everyday psychology (Uher, 2013, 2020b), making their referents particularly diverse and complex. Specifically, referent classes of psychological terms can include concrete entities—accessible publicly (e.g., behavioural acts), only privately (e.g., experiencings), or only indirectly through technological measuring instruments (e.g., physiology)—but also conceptual entities that laypeople intuitively construe about these diverse concrete entities as well as about further conceptual ones (e.g., implicit theories). Finally, psychological terms can also refer to the abstract constructs, models and theories that scientists develop about all these entities (e.g., 'dopaminergic system', 'personal constructs').

Consequently, psychological terms have referents with a NESTED CONCEPTUAL STRUCTURE involving both concrete and conceptual entities on various levels of generalisation, abstraction and categorisation and in which they can 'inherit' meanings and referents from other terms and concepts. This complicates the definition of psychological terms, constructs and their referents. Specifically, single-word labels often mask the fact that psychological terms and constructs (e.g., 'extraversion') often denote a conglomerate of different classes of referents, both conceptual (e.g., beliefs, attitudes) and concrete (e.g., behaviours, feelings), thus constituting BLENDED CONCEPTS (Uher, 2018b).

2.2 Operationalising constructs

The nested structure of many construct referents is reflected in the nomological networks that are used (though not always explicitly) to make constructs accessible to empirical investigation.

2.2.1 Nomological networks mistaken as mere semantic networks

Their constructed and abstract nature entails that constructs imply more meaning than their referents (SURPLUS MEANING). This complicates empirical investigations because, no set of referents, however large, can be all-inclusive. Scientists must make interpretive decisions about which particular referents and which of their interrelations to include and in what ways (2.3.3). For these decisions, a NOMOLOGICAL NETWORK is established around a construct, comprising a theoretical and an empirical framework and linkages within and between them (Cronbach & Meehl, 1955). In the theoretical framework, a more abstract construct (e.g., 'extraversion') is broken down to more specific (sub-)constructs (e.g., 'warmth', 'gregariousness'), thus specifying the construct's THEORETICAL DEFINITION. The empirical framework specifies for each sub-construct various more concrete entities serving as CONSTRUCT INDICATORS (e.g., items for 'warmth'), thus specifying the constructs' OPERATIONAL DEFINITIONS (Figure 1).

Nomological networks are surrounded by numerous fallacies, derived from common misconceptions of concepts and language (2.1.10). One fallacy is their equation with semantic networks. SEMANTIC NETWORKS are a logic-based formalism to describe the mental organisation of knowledge representations and information retrieval pathways (Pirnay-Dummer et al., 2012). Persons' mental association strength between concepts (studied e.g., with reaction times) are illustrated in nodes and directed edges. These networks describe multidimensional associations among constructs and their nested structure (2.1.11)—thus, the fields of meaning that are common in given sociolinguistic communities.

Semantic networks necessarily underlie nomological networks; but they are not the same. Semantic networks describe conceptual structures that people have in their minds and that are implicit to everyday language. Nomological networks, by contrast, are established by

scientists to explicitly link abstract conceptual (and therefore 'unobservable') entities with concrete entities that are accessible and thus potentially measurable in a study (Uher, 2020a). Hence, the theoretical framework of a nomological network describes the hierarchical organisation of construct systems based on their level of generality (organisation corollary; Kelly, 1955; i.e., a conceptual relational system), specifying at its bottom narrow constructs that can be linked to accessible entities (construct indicators), which are specified in the empirical framework (i.e., an empirical relational system; Figure 1 and Section 2.4.2). This also underlies empirically established 'personality' taxonomies or *trait* hierarchies and their linkages to specific sets of items (without necessarily being labelled nomological networks).

By equating nomological with semantic networks, some erroneously assumed that "the *meaning of a theoretical term is solely determined by the place of that term in the nomological network* … and not by a reference to an actual entity. Thus, in this view one can have *meaning without reference*, and can *invoke theoretical terms without automatically engaging in ontological claims*" (Borsboom, 2005, p. 155). A theoretical term denotes a construct; thus, it serves as the construct's signifier. The meaning of a construct (and thus of its term) is drawn from its referents and the particular aspects emphasised in them (2.1.3). Not all terms (signifiers) denote specific concepts (meanings) that refer to concrete entities (referents). Indeed, many psychological terms denote abstract constructs with nested referents on various levels of abstraction—and thus broad fields of meaning and central positions in large semantic networks (2.1.11; similarly, Maraun & Gabriel, 2013).

2.2.2 Disparate notions of 'unobservability'

Central to many fallacies is the alleged 'unobservability' of psychology's study phenomena, often blamed for the unclarity of psychological concepts. But the term is used variously for *1)* constructs (e.g., 'extraversion'), which are 'unobservable' because they are conceptual entities, existing as entities only in our minds; *2)* psychical phenomena (e.g., feelings), which are fundamentally imperceptible in others but directly accessible in ourselves (Uher, 2016a); and *3)* phenomena that are causal to observable ones but 'unobservable' in themselves because they are located internal to individuals' bodies or accessible only indirectly through technological measuring instruments (e.g., physiology). The conflation of these disparate notions of 'unobservability' entails numerous fallacies in which constructs, their referents and the ontological causes of these latter are conflated with one another (Figure 1).

2.2.3 Fallacies around constructs

Cronbach and Meehl (1955) introduced some fallacies themselves. They variously described constructs as theoretical concepts or heuristics or *attributes possessed by individuals* (reification). This fallacy of misplaced concreteness (2.1.6) and common misconceptions of concepts and language (2.1.10) misled them to conflate constructs, thus conceptual entities, with the concrete entities they are meant to denote (CONSTRUCT-REFERENT CONFLATION). The common nominalised description of constructs as if they were real objects with properties, structures and causes promotes their reified presentation and blurs their constructed connotation and the construct-referent distinction. Fostered by empirical realist beliefs, this misled Cronbach and Meehl to see in nomological networks the primary theoretical approach for making inferences about the phenomena that are assumed to be *causally responsible for variations in the constructs studied*—thus, for studying psychical phenomena (Maraun, 2012; Maraun & Gabriel, 2013; Slaney & Garcia, 2015).

Construct-referent conflation is difficult to recognise because many construct referents are constructs themselves (2.1.11), especially when language-based methods are used, in which the phenomena of interest are simply described but need not even be present during data generation. Indeed, whereas ethological variables for behavioural observation describe concrete entities observable at a moment (e.g., behavioural acts), rating items often denote constructs,

thus conceptual entities 'unobservable' in themselves (e.g., *habitual behaviours* in 'personality' ratings; Uher, 2020a). By mistaking words for the phenomena they denote (2.1.10), promoted by realist beliefs, scientists may erroneously assume they are studying the described phenomena in themselves, even if these no longer exist in themselves (e.g., past behaviours in *habitual behaviour* judgements; Uher, 2013, 2015d; Uher, Werner, & Gosselt, 2013).

2.2.4 What is a 'psychological variable'?

Variability in phenomena and their properties is informative about their structures and functionings and therefore a key focus of investigation. Likely for this reason do many psychologists "not speak of studying phenomena but rather of *studying variables*" (Maraun & Gabriel, 2013, p. 38). Psychologists *measure and manipulate variables* and invoke *concepts of personality and intelligence as causal variables* (Barrett, 2003), thus granting "*a serious ontological status to variables"* (Borsboom, 2008, p. 41). But apart from natural constants, everything varies and can thus be considered a *variable*. This undifferentiated notion shifts the focus away from the particular phenomena and properties studied (variability of what?) and blurs their distinction (Uher, 2021).

The confusion is further increased because psychologists denote as *variables* also the semiotic encodings of variability in their data (e.g., statistical variables). DATA[7] are the sign (symbol) systems that scientists use to (mathematically) represent (signifier; e.g., Arabic numerals) information about the study phenomena and properties (referents; e.g., behaviours, duration) as conceived by the data-generating persons (meanings; e.g., qualities, mathematical properties; 2.5.2). In this sense, variables are the semiotic encodings (mathematical representations) of measurement results, not the properties to be measured in themselves. Labelling both the constructs, phenomena and properties under study as well as their semiotic encodings in data uniformly as *variables* entails VARIABLE-REFERENT CONFLATION, which underlies a key fallacy in psychometric modelling (2.8; Uher, 2021).

2.2.5 What is an 'attribute'?

The term 'attribute' is another example of psychological jargon that codifies fallacies derived from conflations of terms, concepts and study phenomena. Psychologists speak of *theoretical attributes* (Haig & Borsboom, 2008), *non-observable, inferred, hypothesised attributes* (Barrett, 2003), and *attributes possessed by individuals* (Cronbach & Meehl, 1955). They assume that *attributes exist* and that "*variations in the attribute causally produce variations in the outcomes of the measurement procedure"* (Borsboom, 2005, p. 150). Psychologists debate whether an *attribute is quantitative and measurable* (Michell, 1999), what structures *latent attributes* might have (Michell, 2008) and how to *place attributes in models* (Borsboom, 2005). In these disparate notions, 'attribute' sometimes refers to *1)* constructs (conceptual entities), sometimes to *2)* ontological phenomena (concrete entities, often considered internal and not directly accessible by researchers), and sometimes to *3)* signifiers encoding these different entities in data models (2.1.-2.2). Scientific language must enable clear differentiations of study phenomena from one another and from the concepts, methods, terms and data used for their exploration.

2.3 Developing multi-stage rating scales as measuring 'instruments'

Like 'variables', 'attributes' are assumed to vary. But variations of what?

[7] The term DATA is also used inconsistently. It denotes semiotic encodings (e.g., item variables, variable values) but also seems be used to denote the sample itself, thus the specific set of phenomena, properties and individuals studied from the universe of all possible ones (e.g., in Borsboom & Mellenbergh, 2004).

2.3.1 Triple conflations of phenomenon, quality and quantity

Psychologists commonly ignore that quantities are always of something. QUANTITY denotes divisible properties of entities of the same kind—the same quality, whereas QUALITY denotes properties of different kind (Hartmann, 1964). Thus, quantities are qualitatively homogenous; adding or dividing the magnitude of specific entities does not change their meaning as entities of the given quality. Any divisible properties of the same quality differ only quantitatively but never qualitatively (Michell, 2012).

Hence, quantities always refer to particular qualitative properties (e.g., length, mass, temperature, time). Qualities, in turn, do not exist as such but occupy a study phenomenon (or object), each featuring properties[8] of diverse quality (e.g., behaviours' temporal and spatial properties). Likely misguided by the ambiguous term *trait* (2.1.6), psychologists seldom specify the particular properties to be measured in their study phenomena. In his widely-cited definition of measurement, Stevens (1946) even omitted the fact that its aim is to generate quantifications OF SOME PROPERTY.

This problem is also masked by the agreement, frequency and other judgements widely-used to generate numerical data with language-based methods, which typically fail to specify the property to be quantified in the diverse phenomena described. Agreement and frequency judgements are not specific properties but abstractions in themselves. This lack of specification entails widespread PHENOMENON-QUALITY-QUANTITY CONFLATIONS, most prominently in the jargon around 'psychological variables' and 'attributes' (2.2.4-2.2.5). But without specifying the particular qualitative properties to be measured in the study phenomena, assumptions that an *attribute is quantitative* (Michell, 1999) cannot be scrutinised and quests for establishing an additive-metric base-unit (Barrett, 2018) are futile.

2.3.2 Basic methodological principles of measurement and measuring instruments

Measuring instruments require that variations in the study phenomena's properties causally influence variations in the measurement results—in psychology (Borsboom, 2005), philosophy of science (Woodward, 2003) and metrology, the science of measurement. Unlike psychologists, metrologists elaborated this requirement into a structural framework of measurement (Mari et al., 2017; Mari et al., 2015). Methodological analyses revealed that this metrological framework is based on two basic principles, data generation traceability and numerical traceability (Uher, 2018a, 2020a). These principles are—on their abstract level of consideration and considering inherent peculiarities and limitations—applicable also to psychological study phenomena and can therefore inform psychological measurement theories.

2.3.2.1 Data generation traceability

Data generation traceability requires that the ways in which results are assigned to the specific entities to be measured in the target property (called MEASURANDS) are made fully transparent, and thus traceable. For justified attributions of results to measurands, measurement processes must be designed from knowledge about the study objects and their properties (called OBJECT-DEPENDENCE or OBJECT-RELATEDNESS in metrology). Scientists must explain how the specific operative structures allow to make numerical assignments such that they reveal reliable and valid information about these measurands, and only about them and not also on other influence properties (Mari et al., 2017). This knowledge must be implemented in UNBROKEN DOCUMENTED CHAINS OF COMPARISONS between the properties connecting measurand with result. Similarly, some psychologists demanded "for a procedure to count as a

[8] The term PROPERTIES here denotes particular qualities (some of which may feature divisible properties, thus quantities; 2.3.1); it should not be mistaken for *traits*, which denote constructs in psychology (2.1.4 and 2.1.6).

measurement procedure, it must *yield measurements* [sic: quantifications[9]] of something; that is, it requires that there be a certain connection between the observations and *some theoretical attribute*" (Haig & Borsboom, 2008, p. 2).

"How does one know that the measurement procedure […] indeed measures this particular *attribute* and not something else or nothing at all?" (Haig & Borsboom, 2008, p. 2). This is established through unbroken connection chains because they start from the measurand's EMPIRICAL INTERACTION with a first property connected to it, which, in turn, may interact with a further property and so on. Each connection step establishes proportional relations between the divisible properties (quantities) of the interconnected qualitative properties, thereby establishing causal QUANTITATIVE RELATIONS along the documented chain from measurand to result (Uher, 2020a). In spring scales, for example, scientists chain the target property 'mass' stepwise with the qualities 'gravity force' and 'length of spring deflection' (through physical laws), the latter with 'length of extension over scale' (through visual comparison), and this, in turn, with 'variable and variable value' as results (through semiotic encoding)[10].

2.3.2.2 Numerical traceability

This second methodological principle of measurement requires that the numerical values assigned to measurands are furthermore linked to known standards in documented, thus traceable ways. Measurement process designs must ensure that results are invariant regarding the persons involved (called SUBJECT-INDEPENDENCE or INTER-SUBJECTIVITY in metrology), reliably interpretable, and always represent the same information about the measurands across time and contexts (Mari et al., 2017). To ensure that results have the same meaning everywhere (e.g., specific weight of 1 gram), metrologists establish unbroken documented connection (calibration) chains from internationally agreed primary references (e.g., prototype kilogram) to all working references (e.g., weighing scales) used in non-metrological research and everyday life (JCGM200:2012, 2012). Psychologists must establish (and have in parts already done so) analogous ways to establish intersubjective meanings for numerical assignments (e.g., achievement tests featuring answer categories with universally agreed meanings of correctness or time-based measurements of behaviour; Uher, 2013; Uher, Addessi, & Visalberghi, 2013).

2.3.3 Construct operationalisation is not measurement

For empirical investigations, scientists must make interpretive decisions about which indicators are representative of a construct, this is called FIAT MEASUREMENT[11] (Cicourel, 1964). Reification and construct-referent conflation (2.2.3), however, mislead scientists to attribute to constructs a causal role in data generation. In conjunction with the conflations of phenomena and properties with their semiotic encodings (variable-referent conflation; 2.2.4), this entails the erroneous belief that identifiable (statistical) features of item variables could allow to explore the structure of psychical phenomena's properties and that, consequently, sets of items could constitute *measuring instruments for constructs*.

But connections between constructs, sub-constructs and items cannot be proven or 'discovered' because constructs do not exist as real entities in themselves. These connections are decreed by scientists given a theory, face validity or common sense, so that operational

[9] Psychologists commonly conflate measurement with quantification, thus the process with its result (Mari et al., 2017; Uher, 2018a, 2020a). This may contribute to the focus on result-dependent approaches of data generation.
[10] Developing such unbroken documented connection chains (e.g., for instrument development) requires knowledge of systematic (lawful) connections among properties. They can be identified experimentally, whereby knowledge generation and instrument development are often entangled (Mari et al., 2017).
[11] From Latin fiat for "let it be done".

definitions inevitably vary (e.g., different items for same construct). Importantly, these decisions concern only the QUALITATIVE properties of constructs and their referents; but they cannot establish documented and unbroken connection chains from hypothetical measurands in the constructs to possible QUANTITATIVE properties in the concrete phenomena used as indicators, as required for measurement (2.3.2). This is because construct referents commonly involve different kinds of phenomena each featuring heterogeneous qualities in which particular aspects are emphasised or deemphasised. But only entities featuring the SAME kind of property—the SAME quality—can be compared with one another regarding their divisible properties in this quality—their quantities. But what qualities featuring divisible properties could be identified in conceptual entities with heterogeneous[12] referents? In a hypothetical 'bodiliness' construct; operationalised with persons' height, weight, foot length and age; what divisible properties could we seek in this conglomerate of heterogeneous physical qualities? This example, although bizarre, is structurally analogous to many psychological constructs (e.g., 'extraversion' refers to various concrete and conceptual phenomena featuring heterogeneous qualities; 2.1.3) and highlights that, in blended concepts, data generation traceability cannot be established.

Thus, contrary to common beliefs, construct operationalisation constitutes no step of measurement and constructs cannot be measured in themselves. Measurement processes can be established only for construct indicators provided these are specific properties that are accessible (directly or indirectly; 2.9.3), such as the specific bodily properties in our fictitious example. Rating items, however, often specify neither particular qualitative properties nor quantitative properties to be measured in them but instead only abstractions, thus constructs (Uher, 2020a).

2.4 Asking persons to complete rating scales

Quantitative psychologists did not yet devote much research to this step, in which information about the study phenomena is encoded into data (Uher, 2018a).

2.4.1 Language-based 'instruments': Triadic interactions

Measurement requires an empirical interaction with the target property (2.3.2). In person-executed measurement (e.g., observations, ratings), these interactions involve the data-generating persons' perceptions and interpretations of the phenomena and properties under study and of the semiotic systems used for data generation, as well as the meaning both have for these persons in order to establish systematic matching relations between them. These interactions reflect the interrelations among the three components of signs systems (2.1.9).

2.4.2 Representation theorems

As semiotic systems, data allow to store, manipulate, decompose and recompose; thus, analyse information about the study phenomena IN LIEU OF and in ways not possible with these latter. But inferences to the study phenomena can be made ONLY IF the data represent relevant properties of these phenomena in appropriate ways. For measurement, this requires scientists to specify *1)* the phenomena and the qualitative and quantitative properties to be studied in them—the EMPIRICAL RELATIONAL SYSTEM; *2)* the system to be used for generating data about

[12] Confusingly, the hierarchical organisation of constructs is sometimes called HETEROGENOUS ORDER and contrasted with HOMOGENEOUS ORDER (e.g., Michell, 2012), thereby implying that these disparate notions of 'order' would be somehow comparable. The undifferentiated use of the term 'order' for both hierarchical conceptual abstractions and linear arrangements of magnitudes (2.3.1)—thus, taxonomic order versus magnitude order—may have contributed to the widespread belief constructs could be measurable.

them—the SYMBOLIC RELATIONAL SYSTEM, and *3)* a set of structure-preserving mappings between both (e.g., homo-morphisms)—thus, systematic ASSIGNMENT RELATIONS.

This highlights the necessity of REPRESENTATION THEOREMS, formalizing axiomatic conditions by which empirical relational structures can be mapped to symbolic relational structures. This idea (together with uniqueness theorems; 2.5.1) is basic to representational theory of measurement (Krantz et al., 1971; Vessonen, 2017). However, this theory provides no concepts and procedures for implementing such theorems (Mari et al., 2017). Its vague concepts, variable-referent conflation (2.2.4) and psychometric developments (2.7-2.8) may have fuelled the erroneous belief that representation would be irrelevant for measurement (e.g., Borsboom & Mellenbergh, 2004; Michell, 1999). But representation theorems are basic to any data generation (Uher, 2018a, 2021). In construct research, they establish systematic mapping links from the empirical framework of a construct's nomological network, specifying the concrete entities used for construct operationalisation, to the empirical data (symbolic relational system) generated about them (Figure 1).

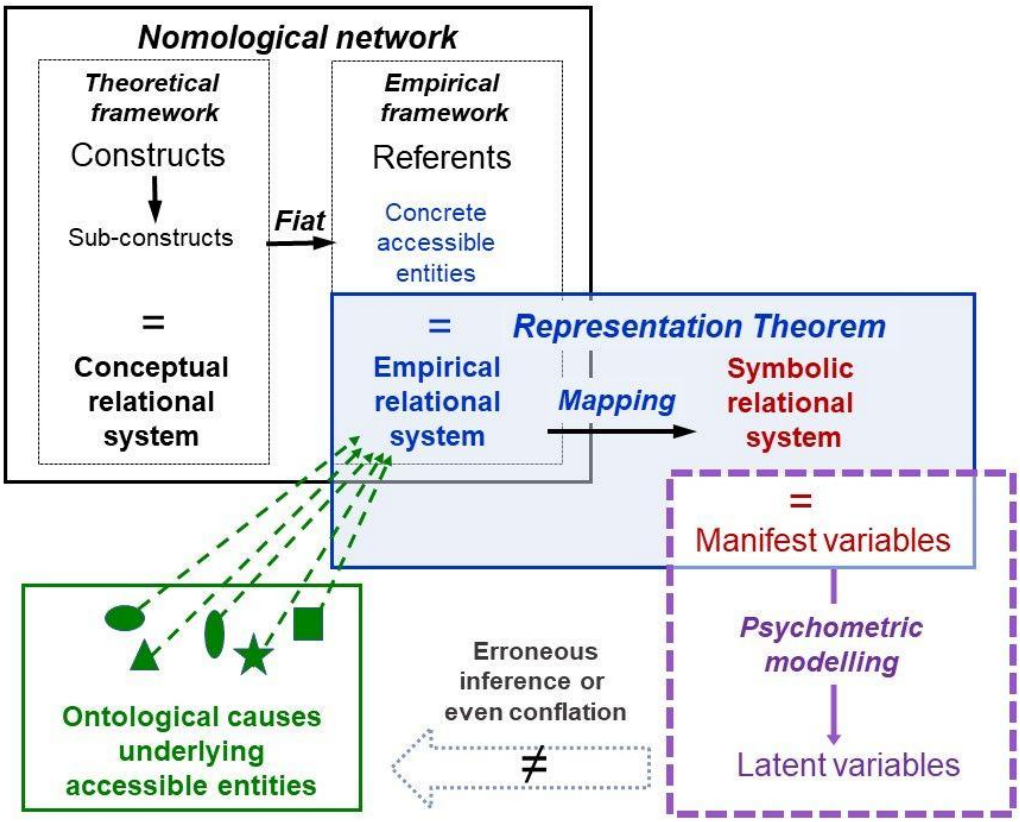

Figure 1 Conceptual relations between nomological networks, representation theorems and psychometric modelling

Explicit representation theorems are needed to establish data generation traceability and numerical traceability (2.3.2). In ethological behaviour observation, the standardised implementation of representation theorems through the data-generating persons is trained and traceable (e.g., in video analyses; Uher, Addessi, et al., 2013). Rating scales, however, serve BOTH as descriptions of the empirical relational system (e.g., behaviours) AND as symbolic relational system (item variables, answer categories), leaving the two systems and their mapping relations unspecified. Instead, their specification is left to respondents' implicit decisions. But

apart from general assumptions about inattention, faking and response bias, psychologists still know little about how raters generate rating data. Pronounced variations in raters' interpretations of items (2.1.10) and answer categories show, however, that rating scales preclude standardised data generation (Uher, 2018a).

2.5 Recoding answer categories into 'numbers'

Multi-stage rating scales are intended to indicate degrees (e.g., of agreement) and are therefore used to generate quantitative data.

2.5.1 Uniqueness theorems

In representational theory of measurement, permissible ways for transforming the symbolic relational systems created without breaking their relations to the empirical relational systems onto which they are mapped are specified in UNIQUENESS THEOREMS (Vessonen, 2017). The permissible transformations specified by Stevens' (1946) four scale types (nominal, ordinal, interval, rational) received much attention in psychology. Together with Stevens' simplified definition of measurement as the "assignment of numerals to objects and events according to rules" (p. 677), they laid the foundation for psychologists' common practice of transforming the (lexical) answer categories of rating scales (e.g., 'agree', 'strongly agree') into numerals (e.g., '3', '4').

Given that rating scales commonly describe diverse and often conceptual phenomena without specifying any particular quality to be quantified, it is unsurprising that 90% of 78 raters in a study interpreted the answer categories not quantitatively but only qualitatively (Figure 13 in Uher, 2018a). Still, psychologists rigidly recode scale categories into numerals—in the SAME way across ALL items regardless of the phenomena and qualities described.

2.5.2 Numeral-number conflation

Numerals are not numbers. Numerals are arbitrary signifiers (e.g., 2, 5; II, V), whereas numbers are mathematical objects arising from ontological interrelations among real phenomena (Hartmann, 1964). Numerals can be used to represent numbers, but also only order (e.g., 2nd, 5th) or only categorical—thus, qualitatively different—properties that have no quantitative meaning at all (e.g., phone *numbers*; Campbell, 1919/2020).

This important distinction if often ignored (NUMERAL-NUMBER CONFLATION), such as when assuming observations would be *represented in numbers* (e.g., Borsboom & Mellenbergh, 2004; Michell, 1999). In measurement, scientists assign not numbers to properties but numerical values (Mari et al., 2015). Their meaning is established through (international) conventions (numerical traceability; 2.3.2.2) and is, necessarily, quality-specific—3 metres is not the same as 3 grams or 3 minutes. Psychologists' rigid recoding of answer categories creates numerals that are devoid of meaning—and thus of mathematical information. Ratings of 'often' recoded into '3' denote neither 3 times nor 3 hours nor even the same level of occurrence across the referents described because occurrence rates generally vary for different phenomena and properties. The meaninglessness of these numerals is masked by psychologists' term SCORES, which implies they could be used flexibly for just any quality and quantity.

2.5.3 Differentiability: Creating quantitative meaning for meaningless scores

Psychologists create quantitative meaning for scores primarily by analysing them for RELATIVE DIFFERENCES. This requires items that allow to generate scores that differentiate well (DIFFERENTIABILITY, differential discrimination), such as between individuals on given variables (between-individual ranking) or between variables within the same individual (within-individual ranking, e.g., Q-sort of typicality). In variable-oriented approaches, quantitative information is thus created from relative differences between scores generated for DIFFERENT individuals—thus, in DIFFERENT mapping processes. In individual-oriented approaches, it is

created from differences in scores generated for DIFFERENT variables—thus, DIFFERENT phenomena and properties[13]. But without unbroken and traceable links to the measurands and to specified standards of comparison (2.3.2), such between-score differences have no counterpart in the real phenomena studied (empirical relational system). These between-score differences can therefore represent ONLY CONCEPTUAL ENTITIES, which cannot reflect magnitudes attributable to the measurands (e.g., individuals' properties) as required for measurement.

2.5.4 Result-dependent processes of data generation

Differential approaches are basic to quantitative psychology. Through iterative item analysis and selection, psychologists establish item scales that reliably produce results with desirable empirical structures (e.g., high differentiability). But without establishing unbroken measurand-result connections using substantive theories about the study phenomena and their qualitative and quantitative properties, these RESULT-DEPENDENT processes of 'instrument' development radically match data generation to desirable results rather than to the objects and properties under study, as required for measurement (2.3.2). But "[t]he route from theory or law to measurement can almost never be travelled backwards. *Numbers* gathered without some knowledge of the regularity to be expected almost never speak for themselves. Almost certainly they remain just *numbers*[14]" (Kuhn, 1961, p. 174-175).

2.6 Analysing reliability: Consistency in the results generated

Consistency in results refers to the structural constraint of homo-morphic mapping relations between empirical and symbolic relational system stipulated in representation theorems (2.4.2). Accordingly, consistency in the symbolic relational system (data set) provides one necessary but insufficient condition for showing that it appropriately represents the empirical relational system under study.

2.6.1 Inter-rater reliability: No evidence for agreement in data generation

Psychologists commonly interpret inter-rater reliability as reflecting intersubjectivity (subject-independence). But comparing results (signifiers) provides evidence neither for raters' agreement in the specific phenomena and properties considered (referents), nor for agreement in raters' interpretations of them (meanings), nor for agreement in raters' assignments to the symbolic relational system (mapping execution). Instead, psychologists commonly explore raters' agreement in ranking DIFFERENT individuals, whereby, in self-ratings, every rater assesses even ANOTHER individual (e.g., Model 1 estimation; Shrout & Fleiss, 1979). Thus, inter-rater reliability refers to DIFFERENTIAL PATTERNS among results generated in DIFFERENT mapping systems, which provides no evidence of traceable data generation processes as needed for measurement.

By contrast, inter-coder or inter-observer reliability in behavioural research (e.g., ethology) compares the results generated by independent persons for the SAME property in the SAME behaviours and SAME individuals at the SAME occasions—thus, for the SAME referents and using the SAME assignment rules (Uher, Addessi, et al., 2013). Reliability thus-conceived compares results obtained from independent executions of the SAME mapping process targeted at the SAME measurands.

[13] Within-individual variations can refer to the same property if considered over time, but their meaning is commonly still derived from between-individual comparisons rather than from specified standards as required for numerical traceability (2.3.2; see also 2.6.3).
[14] Note the correct term here is numerals.

2.6.2 Internal reliability: Operationalist epistemologies for construct definition

Internal reliability denotes how well each item's scores relate independently to the scores of the other items operationalising a construct and how they are related overall. Again, consistency in differential patterns across items provides no evidence for consistency in the processes of their generation as needed for measurement.

Internal reliability is often used to operationally develop and define constructs (e.g., using factor analysis to create *independent constructs*), thereby adapting a nomological network's theoretical and empirical frameworks to the results generated (2.5.4; Figure 1). But this result-dependent OPERATIONALIST approach fundamentally contradicts the REALIST epistemology underlying measuring instruments according to which variations in the target property must be causally related to the results obtained (object-dependent process; 2.3.2)—and not vice versa.

2.6.3 Test-retest reliability: No evidence for consistency in mapping processes

The level of temporal reliability is commonly assumed to indicate how much of the variation in the numerical relational system (data) is attributable to variation in the empirical relational system under study (individuals' properties). But without implementing at least a basic representation theorem, consistency in assessment results obtained at different times cannot provide any evidence for traceable and reproducible mapping processes. Indeed, implicit semantic structures of language (e.g., related and redundant meanings) alone inevitably produce some consistency in assessments (Maul, 2017; Shweder & D'Andrade, 1980; Uher, Werner, & Gosselt, 2013).

2.7 Analysing validity

Misconceptions and fallacies (2.1-2.6) entail conceptual confusions about the empirical processes established as well as controversies about validity.

2.7.1 Pragmatic framework for validation: Established practices

Given that phenomena, qualities, quantities and constructs are often conflated with one another and with their semiotic encodings in terms, variables and scores, and given that language-based 'instruments' largely constitute linguistic abstractions and constructs in themselves (e.g., rating scales), psychologists' debates on whether validity refers to 'instruments' or scores (Kane, 2013; Newton, 2012) are unsurprising. Insufficient elaboration of the study phenomena's nature, the qualitative and quantitative properties to be studied in them, the challenges of their accessibility and the resulting methodological implications entail that unbroken documented connection chains establishing proportional relations in the divisible properties (quantities) between measurand and results, as required for measurement (2.3.2), can be neither theoretically conceived nor empirically implemented.

Consequently, psychologists (can only) determine validity on the basis of the RESULTS generated rather than of the PROCESSES of their generation. Such validity concepts explore the results' convergence with the results generated for theoretically-related constructs or for concurrent or future events (Cronbach & Meehl, 1955). Or they focus on the inferences and actions derived from the scores created, such as their plausibility, coherence and appropriateness (Kane, 2013) or social and ethical consequences of their use (Messick, 1995).

In these concepts, validity properties are determined on the basis of the results' interrelations with results generated for DIFFERENT measurands in DIFFERENT properties and DIFFERENT phenomena that are considered to be relevant and meaningful for a given purpose (Figure 2). Within this PRAGMATIC FRAMEWORK, numerical assignments are useful indices for making predictions in empirical practice (Barrett, 2003; Dawes, Faust, & Meehl, 1989). But this OPERATIONALIST approach aligns 'instrument' development to these purposes rather than to the actual phenomena and properties under study, which would require evidence of

convergent results for the SAME measurands in the SAME properties and SAME phenomena obtained with different 'instruments' or methods. But, contrary to common assumptions, this key principle cannot be applied in construct research because not constructs in themselves but only their diverse indicators are (potentially) measurable (2.9.3). Therefore, different 'instruments' and methods used for studying the SAME constructs necessarily involve investigations of DIFFERENT study phenomena and properties.

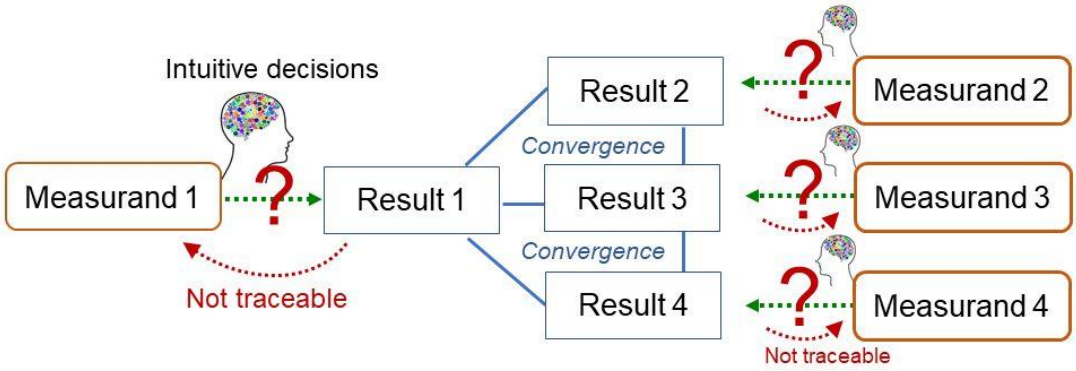

Figure 2 Result-dependent processes of data generation: Example of construct validation using rating-based data

2.7.2 Causal framework of measurement: Intended but not implemented

Nevertheless, following explicitly empirical realist epistemologies, many psychologists interpret validity in terms of causal relations between the presumed target property and the results, e.g., as "a property of measurement instruments that codes whether these *instruments are sensitive to variation in a target attribute*", which is "broadly consistent with the view that a test is valid if it measures what it should measure" (Borsboom et al., 2009; p. 135). "The primary power of causality lies in the theoretical opportunity to think directionally rather than in terms of similarity or correlation" (Borsboom, 2005, p. 160).

This theorising corresponds to metrological and philosophy-of-science concepts of measuring instruments (2.3.2)—but not to psychologists' empirical practice (2.7.1) given that most psychological study phenomena are constructs. As conceptual entities, constructs cannot have any causal effects; such assumptions derive from the reification of constructs as real entities internal to or *possessed by* individuals (e.g., *traits*, *attributes*; 2.1.6). Links between constructs and concrete, thus (potentially) measurable entities are established through interpretive decisions about their qualitative properties (operationalisation), not through empirically detectable causal relations.

Psychological 'instruments' are necessarily language-based, thus constitute semiotic systems comprising signifiers, referents and meanings. Such systems cannot be *sensitive to variations of a property* as physical properties can (therefore called sensitive transducer in metrology; Mari et al., 2015). Instead, language-based 'instruments' involve triadic interactions (perceptions and interpretations) of the data-generating persons (e.g., raters) with BOTH the 'instruments' AND the phenomena and properties under study. But these complex interactions are still largely unknown (2.4) and traceable connection chains implementing proportional relations between quantitative properties, as needed for measurement, cannot be established. Rigid recoding of answer categories and processes of result-dependent data generation cannot provide any evidence for causal measurand-result relations either.

2.8 Psychometric modelling of latent variables

Given these challenges, psychologists aim to *quantify constructs* by exploring structures underlying the scores generated for a construct's indicators in the symbolic relational system (MANIFEST VARIABLES; Figure 1). Statistical models allow to transform cumulative (nominal or ordinal scaled) raw scores generated for manifest variables (e.g., derived for one person across items or for one item across persons) into continuous LATENT VARIABLES of person parameters (e.g., ability) and item parameters (e.g., difficulty) that are assumed to be interval or even ratio scaled (depending on the model used). Following the assumption that the properties of latent variables reflect properties of the study phenomena in themselves, psychometric modelling is considered the ultimate means for exploring quantitative properties in psychology's elusive study phenomena.

These assumptions, however, derive from common misconceptions of language and concepts (2.1.10) and the fallacies around constructs that are codified in psychological measurement jargon (2.2-2.7). Latent variables *are* neither the constructs (conceptual entities) they encode (LATENT VARIABLE-CONSTRUCT CONFLATION) nor these constructs' possible referents in themselves (variable-referent and construct-referent conflations; similarly, Maraun & Halpin, 2008; Slaney & Garcia, 2015). Latent variables (e.g., *factors, dimensions*) are mere statistical concepts used to describe structures in data sets once these have been generated by implementing representation theorems (Figure 1).

In psychological measurement, unbroken measurand-result connections are established neither theoretically nor empirically. Instead, latent properties are derived from differential analyses of scores obtained for DIFFERENT individuals, DIFFERENT item variables, or both. Such differences constitute mere conceptual entities that, without unbroken measurand-result connections, cannot reflect magnitudes attributable to the single individuals studied (2.5.3). Consequently, the quantitative properties identified through psychometric modelling can have only statistical and methodical origins (e.g., error structure; Michell, 2008; simplistic encoding format of rating scales; Uher, 2013, 2015d, 2018a).

2.9 Interpreting findings

Conflation of latent variables with the study phenomena in themselves is also driven by psychologists' intention to not only describe or predict but also to causally explain the phenomena observable in individuals.

2.9.1 Disparate notions of 'causality'

Psychologists' use of 'causality' concepts, however, is inconsistent. They denote either *1)* measurand-result connections establishing data generation traceability in measurement; *2)* statistical relations between variations on latent and manifest variables; or *3)* psycho-bio-social phenomena governing the emergence of observable study phenomena. The first notion refers to the realist framework intended but not implemented (2.7.2); the second to the operationist framework empirically implemented in statistical modelling (2.7.1); the third to ontological claims. Their conflation in psychological jargon entails a repeated switching between these disparate notions and, in consequence, their underlying epistemologies. This results in frequent leaps in argumentation, misinterpretation of psychometric modelling as measurement (Uher, 2021), conflation of description with explanation, and thus of measurement theory with explanatory theory.

2.9.2 Description is not explanation

Description presupposes explanation (Uher, 2016b). Indeed, coherent empirical investigation (e.g., measurement), hypothesis construction and theory building cannot even begin without first describing and defining the phenomena to be studied, thus without developing

an understanding of the meaning of the concepts therefore used (Maraun, 1998). Knowledge and explanatory theories about the study phenomena and properties and their systematic interactions are needed for instrument development. But they provide only justification for establishing connection chains; they are not the same as measurement theories (Woodward, 2003).

2.9.3 We cannot measure 'intelligence' in itself, only intellectual performances

Misconceptions and fallacies also affect result interpretation, such as when overlooking that not constructs in themselves but only their indicators are (potentially) measurable and thus reflected in the findings—if these constitute concrete entities that are accessible in themselves or structurally connected with accessible entities (2.3.2). Hence, contrary to widespread belief, we cannot measure 'intelligence' in itself because it is an abstract construct that refers to different abilities that we assume to underly observable intellectual performances. Moreover, these abilities in themselves, given their transient and processual nature, can be determined only through generalisation and abstraction over time, thus only conceptually (2.1.2). Quantitative properties measured in 'intelligence test' are, for example, number of correct answers and speed of the performances shown. But importantly, these are quantifications of the OUTCOMES of intellectual abilities—not of these abilities in themselves (Toomela, 2008; Uher, 2020a, 2021). Any test performance involves qualitatively different intellectual abilities; but this is not reflected in the results. More intelligent individuals may use qualitatively different (e.g., more efficient) processes than less intelligent ones leading to quantitatively different results in performance (e.g., correctness, speed). Relations between hypothesised abilities and measurable test performances are specified in theoretical and interpretive assumptions on the basis of their QUALITATIVE properties (operationalisation). Inferences from test results to individuals' intellectual abilities and 'intelligence' must mirror these interpretive decisions. This precludes one-to-one inferences because traceable connections of (hypothetical) QUANTITATIVE properties between constructs ('intelligence', 'mental processes') and their indicators (specific test performances) cannot be established.

3. Consequences and future directions

Various lines of critique, both well-established and new, were used to highlight and explain a complex network of misconceptions and fallacies codified in psychological measurement jargon. They entail serious consequences for research and applied practice.

3.1 Consequences for quantitative psychology

3.1.1 Frameworks enabling measurement of psychical phenomena not yet developed

Contrary to common beliefs, the practices established in quantitative psychology in terms of operationist frameworks that support consistent and monotonic data structures with desired statistical properties do not enable measurement of psychological study phenomena and their properties. Rating methods, widely used to generate 'quantitative' data in psychology, do not even enable standardised data generation, not to mention measurement.

3.1.2 Biased inferences about the study phenomena and their properties

The radical alignment of 'instrument' development and data generation to statistical assumptions and desired data structures rather than to the study phenomena and their properties allows, at best, only biased inferences on these latter, leading to inaccurate explanations (Barrett, 2008). The validity of quantitative psychological data in applied settings

(e.g., as legal evidence) may therefore be challenged in courts similar to forensic psychologists' and psychiatrists' diagnostic practices before (Barrett, 2018).

3.1.4 Replicability problems may be even larger than assumed

Psychology's replicability problems may be even more profound than currently thought because the operationalist tailoring of 'instruments' to produce consistent (reliable) data structures artificially increases the replicability of results (Uher, Addessi, et al., 2013). But this is achieved at the expense of accurate structure-preserving mappings and may therefore lead to systematic errors (Hammersley, 2013), thus reducing replicability in terms of measurement.

3.2 Future directions

This article is intended to be an eye-opener for quantitative psychologists. It aims to support critical analyses of the field's current framework of thinking and practice as well as critical reflection of own terminologies, concepts and presumptions to make way for new developments that help avoid the misconceptions and fallacies shown above. Rethinking one's own thinking, language and practice is hard work, much harder than acquiring any new ones. Who could know that better than psychologists themselves!

3.2.1 Theoretical thinking, not statistical handicraft

Statistics has become a preoccupation for many psychologists. Its advent triggered an enormous empirical action-taking, promoted through the efficient large-scale production of allegedly quantitative data through rating methods. But this misled many psychologists to select their study phenomena and research questions on the basis of the methods available, rather than vice versa (method-centrism; statisticism; Barrett, 2003; Lamiell, 2019; Valsiner, 2012). Empirical and conceptual work got out of balance.

3.2.2 Philosophy of science for quantitative psychology

The greatest need for conceptual development—before any theories and methods of measurement could be devised—concerns elaboration of the particular qualities and quantities to be studied in psychological study phenomena rather than to assume that just any conceivable 'more' or 'less' in vaguely defined concepts would constitute a quantity in itself, as currently done. Quantifications can serve pragmatic purposes, but they are meaningless if the phenomena and their qualities are constantly changing in themselves (Uher, 2020a).

Many quantitative psychologists avoid dealing with philosophical questions. But every scientific activity builds on a philosophy of science (for personality psychology, see Rychlak, 1968), and be it implicitly. Quantitative psychology also shields itself fairly strictly from fundamental critique. Criticism and controversial dialogue are, however, needed to promote the inherently dialectical development of knowledge.

3.3 Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals (TPS-Paradigm)

The above analyses rely on the *Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals* (TPS-Paradigm[15]; Uher, 2015a, 2015c, 2018c). Its purpose, basic structure and previous key topics are briefly outlined now.

The TPS-Paradigm comprises coherently interrelated philosophical, metatheoretical and methodological frameworks (therefore *paradigm*). Established concepts from historical and recent lines of thought in psychology and various other disciplines that are relevant for exploring

[15] http://researchonindividuals.org

phenomena in (relation to) individuals were systematically integrated and complemented by novel ones, thereby creating unitary frameworks that transcend disciplinary boundaries (therefore *transdisciplinary*). This highlights connections and starting points for collaboration across disciplines. Moreover, the TPS-Paradigm is targeted toward making explicit and scrutinising basic assumptions made in scientific systems to help scientists critically reflect on, discuss and further develop their theories and practices (therefore *philosophy-of–science*).

3.3.1 Philosophical framework

Three sets of presuppositions are made about the nature and properties of individuals and the phenomena studied in (relations to) them and about the notions by which knowledge about them can be gained (Uher, 2015c, 2015a).

1) ALL SCIENCE IS DONE BY HUMANS and therefore inextricably entwined with and limited by human's perceptual and conceptual abilities, which entails risks for fallacies and biases. A critical-realist epistemology and concepts from theoretical, cognitive and cultural psychology, sociology, and other fields are used to open up meta-perspectives on research processes.

2) INDIVIDUALS ARE COMPLEX LIVING ORGANISMS, conceivable as open nested systems. On each level, they function as organised wholes from which new properties emerge not predictable from their constituents, showing complex patterns of upward and downward causation, resulting in ever more complex systems that are less rule-bound, highly adaptive and historically unique. Concepts from thermodynamics, physics of life, philosophy, theoretical biology, medicine, psychology and sociology about dialectics, complexity and nonlinear dynamic systems were used to elaborate their relevance for research on individuals.

3) THE CONCEPT OF COMPLEMENTARITY, originating from quantum physics and used in philosophy, anthropology and theoretical psychology, was applied to elaborate peculiarities of physical versus psychical (i.e., "non-physical") phenomena, emphasising the necessity for a methodical dualism to account for observations of two categorically different realities that require different frames of reference, approaches and criteria of truth.

These presuppositions underlie the metatheoretical and methodological frameworks.

3.3.2 Metatheoretical framework

At its core are three metatheoretical properties formalising a phenomenon's modes of accessibility to human perception (internality-externality, temporal extension, spatiality conceived complementarily as physical [spatial] versus "non-physical" [without spatial properties]). They are used to metatheoretical differentiate and define various kinds of phenomena studied in (relation to) individuals: morphology, physiology, behaviour, psyche, semiotic representations, outer-appearance modifications and contexts.

These concepts formed the basis for integrating and further developing established concepts from various disciplines to elaborate the peculiarities of psychical phenomena (Uher, 2016a), their functional connections with other phenomena (Uher, 2013) and ontogenetic development; to explore their fundamental imperceptibility by others and its role in the development of agency, language, instructed learning, culture and societies in human evolution (Uher, 2015a); and to clarify when language itself constitutes behaviour (Uher, 2016b).

The metatheoretical definition of 'personality' as individual specificity in ALL kinds of phenomena studied in (relation to) individuals highlighted the unique constellation of probabilistic, differential and temporal patterns that merge together in this concept, the challenges that this entails, arbitrary determination of causes and consequences, and the role that language plays in 'personality' concepts, thereby enabling novel conceptual integrations of existing 'personality' theories (Uher, 2015c, 2015d, 2015e, 2018b, 2018c). The concept of semiotic representations highlighted the composite nature of sign systems (e.g., data; Uher, 2018a, 2019), which is important for scrutinising research practices.

3.3.3 Methodological framework

The methodological framework directly builds on the metatheoretical framework. It features three main areas.

1) PHENOMENON-METHODOLOGY MATCHING. The three metatheoretical properties describing general modes of accessibility were used to derive clear methodological implications and new concepts to help identify fallacies and mismatches (e.g., introquestive versus extroquestive[16] methods to avoid conceptual problems inherent to introspection; Uher, 2016a, 2018a, 2019).

2) APPROACHES FOR TAXONOMISING INDIVIDUAL DIFFERENCES in various kinds of phenomena in human and nonhuman populations were classified by their underlying rationales. Novel, especially behavioural approaches were developed for systematic comparisons with lexically-derived models as well as methodological concepts for comparing individual specificity within and across situations, groups and species (Uher, 2015d, 2015c, 2015e, 2015b, 2018b, 2018c).

3) THEORIES AND PRACTICES OF DATA GENERATION AND MEASUREMENT from psychology, social sciences, metrology and physics were compared, identifying two basic methodological principles of measurement that underlie metrological frameworks and that are applicable also in psychology (data generation traceability, numerical traceability; Uher, 2020a). Analyses scrutinised the involvement of human abilities in data generation across the empirical sciences (Uher, 2019), raters' interpretation and use of standardised assessment scales (Uher, 2018a), and the foundations of psychometrics (Uher, 2021).

Empirical applications were demonstrated in studies with humans of different sociolinguistic backgrounds and various species (e.g., Uher, 2015e, 2018a; Uher, Addessi, et al., 2013; Uher & Visalberghi, 2016; Uher, Werner, et al., 2013). Grounded in established concepts from various disciplines, the TPS-Paradigm provides rich conceptual frameworks as starting points for new developments—such as to explore the possibilities and limitations of measurement in psychology that space does not allow to outline here.

References

Barrett, P. (2003). Beyond psychometrics. *Journal of Managerial Psychology*, *18*(5), 421–439. https://doi.org/10.1108/02683940310484026

Barrett, P. (2008). The consequence of sustaining a pathology: Scientific stagnation— a commentary on the target article "Is psychometrics a pathological science?" by Joel Michell. *Measurement: Interdisciplinary Research and Perspectives*, *6*(1–2), 78–83. https://doi.org/10.1080/15366360802035521

Barrett, P. (2018). The EFPA test-review model: When good intentions meet a methodological thought disorder. *Behavioral Sciences*, *8*(1), 5. https://doi.org/10.3390/bs8010005

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511490026

Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research and Perspectives*, *6*(1–2), 25–53. https://doi.org/10.1080/15366360802035497

Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., & Franić, S. (2009). The end of construct

[16] These methodological concepts were developed to remedy conceptual leaps in concepts of introspection versus extrospection, commonly defined on the basis of the individual under study as inward versus outward perspectives. But this definition does not allow to distinguished them as methods because both perspectives are always contained in any perception. The new concepts of INTROQUESTION and EXTROQUESTION (from Latin "quaerere" for to seek, enquire), by contrast, are differentiated on the basis of *a)* the particular phenomena under study and *b)* the person perceiving them and generating data about them. These concepts highlighted that psychophysical experiments, commonly classified as introspective, constitute extroquestive methods (Uher 2016a, 2019).

validity. In *The concept of validity: Revisions, new directions, and applications*. (pp. 135–170). IAP Information Age Publishing.
Borsboom, D., & Mellenbergh, G. J. (2004). Why psychometrics is not pathological. *Theory & Psychology*, *14*(1), 105–120. https://doi.org/10.1177/0959354304040200
Campbell, N. R. (1957). *Foundations of science: The philosophy of theory and experiment*. New York, NY: Dover Publications.
Campbell, N. R. (2020). *Foundations of science*. Frankfurt am Main: Salzwasser Verlag.
Cicourel, A. (1964). *Method and measurement in sociology*. New York: The Free Press of Glencoe.
Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957
Dawes, R., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668–1674. https://doi.org/10.1126/science.2648573
Fahrenberg, J. (2019). *Wilhelm Wundt (1832 –1920). Introduction, quotations, reception, commentaries, attempts at reconstruction*. Lengerich: Pabst Science Publishers.
Flick, U. (2008). *Managing quality in qualitative research.* Sage Publications.
Haig, B. D., & Borsboom, D. (2008). On the conceptual foundations of psychological measurement. *Measurement: Interdisciplinary Research and Perspectives*, *6*(1–2), 1–6. https://doi.org/10.1080/15366360802035471
Hammersley, M. (2013). *The myth of research-based policy and practice*. London: SAGE Publications. https://doi.org/10.4135/9781473957626
Hanfstingl, B. (2019). Should we say goodbye to latent constructs to overcome replication crisis or should we take into account epistemological considerations? *Frontiers in Psychology*, *10*, 1949. https://doi.org/10.3389/fpsyg.2019.01949
Hartmann, N. (1964). *Der Aufbau der realen Welt. Grundriss der allgemeinen Kategorienlehre [The Structure of the Real World. Outline of the General Theory of Categories] (3. Aufl.)*. Berlin: Walter de Gruyter.
JCGM200:2012. (2012). *International vocabulary of metrology – Basic and general concepts and associated terms (VIM 3rd edition)*. Working Group 2 (Eds.), Joint Committee for Guides in Metrology. Retrieved from https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf
John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality*, *2*(3), 171–203. https://doi.org/10.1002/per.2410020302
Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. https://doi.org/10.1111/jedm.12000
Kelly, G. (1955). *The psychology of personal constructs (Volume 1 & 2)*. London, UK: Routledge.
Krantz, D., Luce, R. D., Tversky, A., & Suppes, P. (1971). *Foundations of measurement Volume I: Additive and polynomial representations*. San Diego: Academic Press.
Kuhn, T. S. (1961). The function of measurement in modern physical science. *Isis*, *52*, 161–193. https://doi.org/10.2307/228678
Kyngdon, A. (2008). The Rasch model from the perspective of the representational theory of measurement. *Theory & Psychology*, *18*(1), 89–109. https://doi.org/10.1177/0959354307086924
Kyngdon, A. (2011). Plausible measurement analogies to some psychometric models of test performance. *British Journal of Mathematical and Statistical Psychology*, *64*(3), 478–497. https://doi.org/10.1348/2044-8317.002004
Lamiell, J. (2019). *Psychology's misuse of statistics and persistent dismissal of its critics*. Springer International. https://doi.org/10.1007/978-3-030-12131-0
Lundmann, L., & Villadsen, J. W. (2016). Qualitative variations in personality inventories: subjective understandings of items in a personality inventory. *Qualitative Research in Psychology*, *13*(2), 166–187. https://doi.org/10.1080/14780887.2015.1134737
Maraun, M. D. (1998). Measurement as a normative practice. *Theory & Psychology*, *8*(4), 435–461. https://doi.org/10.1177/0959354398084001
Maraun, M. D. (2012). Validity and measurement. *Measurement: Interdisciplinary Research and Perspectives*, *10*(1–2), 80–83. https://doi.org/10.1080/15366367.2012.682523
Maraun, M. D., & Gabriel, S. M. (2013). Illegitimate concept equating in the partial fusion of construct

validation theory and latent variable modeling. *New Ideas in Psychology*, *31*(1), 32–42. https://doi.org/10.1016/J.NEWIDEAPSYCH.2011.02.006

Maraun, M. D., & Halpin, P. F. (2008). Manifest and latent variates. *Measurement: Interdisciplinary Research and Perspectives*, *6*(1–2), 113–117. https://doi.org/10.1080/15366360802035596

Maraun, M. D., Slaney, K. L., & Gabriel, S. M. (2009). The Augustinian methodological family of psychology. *New Ideas in Psychology*, *27*(2), 148–162. https://doi.org/10.1016/J.NEWIDEAPSYCH.2008.04.011

Mari, L., Carbone, P., Giordani, A., & Petri, D. (2017). A structural interpretation of measurement and some related epistemological issues. *Studies in History and Philosophy of Science*, *65–66*, 46–56. https://doi.org/10.1016/j.shpsa.2017.08.001

Mari, L., Carbone, P., & Petri, D. (2015). Fundamentals of hard and soft measurement. In A. Ferrero, D. Petri, P. Carbone, & M. Catelani (Eds.), *Modern measurements: Fundamentals and applications* (pp. 203–262). Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9781119021315.ch7

Markus, K. A. (2008). Constructs, concepts and the worlds of possibility: Connecting the measurement, manipulation, and meaning of variables. *Measurement: Interdisciplinary Research & Perspective*, *6*(1–2), 54–77. https://doi.org/10.1080/15366360802035513

Maul, A. (2017). Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary Research and Perspectives*, *15*, 51–69. https://doi.org/10.1080/15366367.2017.1348108

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Michell, J. (1999). *Measurement in psychology. A critical history of a methodological concept*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511490040

Michell, J. (2008). Is psychometrics pathological science? *Measurement: Interdisciplinary Research & Perspective*, *6*(1–2), 7–24. https://doi.org/10.1080/15366360802035489

Michell, J. (2012). Alfred Binet and the concept of heterogeneous orders. *Frontiers in Psychology*, *3*, 261. https://doi.org/10.3389/fpsyg.2012.00261

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Newton, P. E. (2012). Clarifying the consensus definition of validity. *Measurement: Interdisciplinary Research & Perspective*, *10*(1–2), 1–29. https://doi.org/10.1080/15366367.2012.669666

Peirce, C. S. (1958). *Collected papers of Charles Sanders Peirce, Vols. 1-6, C. Hartshorne & P. Weiss (eds.), vols. 7-8, A. W. Burks (ed.)*. Cambridge, MA: Harvard University Press.

Pirnay-Dummer, P., Ifenthaler, D., & Seel, N. M. (2012). Semantic networks. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 3025–3029). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-1428-6_1933

Rosenbaum, P. J., & Valsiner, J. (2011). The un-making of a method: From rating scales to the study of psychological processes. *Theory & Psychology*, *21*(1), 47–65. https://doi.org/10.1177/0959354309352913

Rychlak, J. F. (1968). *A philosophy of science for personality theory*. Boston, MA, US: Houghton, Mifflin and Company. https://doi.org/10.1037/10535-000

Salvatore, S. (2019). Beyond the meaning given. The meaning as explanandum. *Integrative Psychological and Behavioral Science*, *53*(4), 632–643. https://doi.org/10.1007/s12124-019-9472-z

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Shweder, R. A., & D'Andrade, R. G. (1980). The systematic distortion hypothesis. In R. A. Shweder (Ed.), *Fallible judgment in behavioral research: New directions for methodology of social and behavioral science (Vol. 4)* (pp. 37–58). San Francisco: Jossey-Bass.

Slaney, K. L., & Garcia, D. A. (2015). Constructing psychological objects: The rhetoric of constructs. *Journal of Theoretical and Philosophical Psychology*, *35*(4), 244–259. https://doi.org/10.1037/teo0000025

Smedslund, J. (2016). Why psychology cannot be an empirical science. *Integrative Psychological and Behavioral Science*, *50*(2), 185–195. https://doi.org/10.1007/s12124-015-9339-x

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 667–680.
Toomela, A. (2008). Variables in psychology: A critique of quantitative psychology. *Integrative Psychological & Behavioral Science*, *42*(3), 245–265. https://doi.org/10.1007/s12124-008-9059-6
Trendler, G. (2009). Measurement theory, psychology and the revolution that cannot happen. *Theory & Psychology*, *19*(5), 579–599. https://doi.org/10.1177/0959354309341926
Trendler, G. (2013). Measurement in psychology: A case of ignoramus et ignorabimus? A rejoinder. *Theory & Psychology*, *23*(5), 591–615. https://doi.org/10.1177/0959354313490451
Trendler, G. (2019). Conjoint measurement undone. *Theory & Psychology*, *29*(1), 100–128. https://doi.org/10.1177/0959354318788729
Uher, J. (2013). Personality psychology: Lexical approaches, assessment methods, and trait concepts reveal only half of the story-Why it is time for a paradigm shift. *Integrative Psychological and Behavioral Science*, *47*(1), 1–55. https://doi.org/10.1007/s12124-013-9230-6
Uher, J. (2015a). Agency enabled by the psyche: Explorations using the Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals. In C. W. Gruber, M. G. Clark, S. H. Klempe, & J. Valsiner (Eds.), *Constraints of agency: Explorations of theory in everyday life. Annals of Theoretical Psychology (Vol. 12)* (pp. 177–228). New York: Springer International Publishing. https://doi.org/10.1007/978-3-319-10130-9_13
Uher, J. (2015b). Comparing individuals within and across situations, groups and species: Metatheoretical and methodological foundations demonstrated in primate behaviour. In D. Emmans & A. Laihinen (Eds.), *Comparative neuropsychology and brain imaging (Vol. 2), Series Neuropsychology: An interdisciplinary approach* (pp. 223–284). Berlin: Lit Verlag. https://doi.org/10.13140/RG.2.1.3848.8169
Uher, J. (2015c). Conceiving "personality": Psychologist's challenges and basic fundamentals of the Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals. *Integrative Psychological and Behavioral Science*, *49*(3), 398–458. https://doi.org/10.1007/s12124-014-9283-1
Uher, J. (2015d). Developing "personality" taxonomies: Metatheoretical and methodological rationales underlying selection approaches, methods of data generation and reduction principles. *Integrative Psychological and Behavioral Science*, *49*(4), 531–589. https://doi.org/10.1007/s12124-014-9280-4
Uher, J. (2015e). Interpreting "personality" taxonomies: Why previous models cannot capture individual-specific experiencing, behaviour, functioning and development. Major taxonomic tasks still lay ahead. *Integrative Psychological and Behavioral Science*, *49*(4), 600–655. https://doi.org/10.1007/s12124-014-9281-3
Uher, J. (2016a). Exploring the workings of the Psyche: Metatheoretical and methodological foundations. In J. Valsiner, G. Marsico, N. Chaudhary, T. Sato, & V. Dazzani (Eds.), *Psychology as the science of human being: The Yokohama Manifesto* (pp. 299–324). New York: Springer International Publishing. https://doi.org/10.1007/978-3-319-21094-0_18
Uher, J. (2016b). What is behaviour? And (when) is language behaviour? A metatheoretical definition. *Journal for the Theory of Social Behaviour*, *46*(4), 475–501. https://doi.org/10.1111/jtsb.12104
Uher, J. (2018a). Quantitative data from rating scales: An epistemological and methodological enquiry. *Frontiers in Psychology*, *9*, 2599. https://doi.org/https://doi.org/10.3389/fpsyg.2018.02599
Uher, J. (2018b). Taxonomic models of individual differences: A guide to transdisciplinary approaches. *Philosophical Transactions of the Royal Society B*, *373*(1744), 20170171. https://doi.org/10.1098/rstb.2017.0171
Uher, J. (2018c). The Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals: Foundations for the science of personality and individual differences. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *The SAGE Handbook of Personality and Individual Differences: Volume I: The science of personality and individual differences* (pp. 84–109). London, UK: SAGE. https://doi.org/10.4135/9781526451163.n4
Uher, J. (2019). Data generation methods across the empirical sciences: differences in the study phenomena's accessibility and the processes of data encoding. *Quality & Quantity. International Journal of Methodology*, *53*(1), 221–246. https://doi.org/10.1007/s11135-018-0744-3
Uher, J. (2020a). Measurement in metrology, psychology and social sciences: data generation traceability and numerical traceability as basic methodological principles applicable across sciences. *Quality & Quantity. International Journal of Methodology*, *54*(3), 975–1004. https://doi.org/10.1007/s11135-

020-00970-2
Uher, J. (2020b). Psychology's status as a science: Peculiarities and intrinsic challenges. Moving beyond its current deadlock towards conceptual integration. *Integrative Psychological and Behavioral Science*. https://doi.org/10.1007/s12124-020-09545-0
Uher, J. (2021). Psychometrics is not measurement: Unraveling a fundamental misconception in quantitative psychology and the complex network of its underlying fallacies. *Journal of Theoretical and Philosophical Psychology*, *41*(1). https://doi.org/10.1037/teo0000176
Uher, J., Addessi, E., & Visalberghi, E. (2013). Contextualised behavioural measurements of personality differences obtained in behavioural tests and social observations in adult capuchin monkeys (Cebus apella). *Journal of Research in Personality*, *47*(4), 427–444. https://doi.org/10.1016/j.jrp.2013.01.013
Uher, J., & Visalberghi, E. (2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality*, *61*, 61–79. https://doi.org/10.1016/j.jrp.2016.02.003
Uher, J., Werner, C. S., & Gosselt, K. (2013). From observations of individual behaviour to social representations of personality: Developmental pathways, attribution biases, and limitations of questionnaire methods. *Journal of Research in Personality*, *47*(5), 647–667. https://doi.org/10.1016/j.jrp.2013.03.006
Valsiner, J. (2012). *A guided science: History of psychology in the mirror of its making*. New Brunswick, NJ: Transaction Publishers.
Valsiner, J. (2017). *From methodology to methods in human psychology*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-61064-1
Vessonen, E. (2017). Psychometrics versus representational theory of measurement. *Philosophy of the Social Sciences*, *47*(4–5), 330–350. https://doi.org/10.1177/0048393117705299
Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.
Whitehead, A. N. (1929). *Process and reality*. New York, NY: Harper.
Woodward, J. F. (2003). *Making things happen: A theory of causal explanation*. Oxford, UK: Oxford University Press.
Zagaria, A., Ando, A., & Zennaro, A. (2020). Psychology: A giant with feet of clay. *Integrative Psychological and Behavioral Science*, *54*, 521–562. https://doi.org/10.1007/s12124-020-09524-5
***

