



Measurement in metrology, psychology and social sciences: data generation traceability and numerical traceability as basic methodological principles applicable across sciences

Jana Uher^{1,2}

© The Author(s) 2020

Abstract

Measurement creates trustworthy quantifications. But unified frameworks applicable to all sciences are still lacking and discipline-specific terms, concepts and practices hamper mutual understanding and identification of commonalities and differences. Transdisciplinary and philosophy-of-science analyses are used to compare metrologists' structural framework of physical measurement with psychologists' and social scientists' flat measurement of constructs. The analyses explore the functions that measuring instruments and measurement-executing persons in themselves fulfil in data generation processes, and identify two basic methodological principles critical for measurement. (1) *Data generation traceability* requires that numerical assignments depend on the properties to be quantified in the study objects (object-dependence). Therefore, scientists must establish unbroken documented connection chains that directly link (via different steps) the quantitative entity to be measured in the study property with the numerical value assigned to it, thereby making the assignment process fully transparent, traceable and thus reproducible. (2) *Numerical traceability* requires that scientists also directly link the assigned numerical value to known standards in documented and transparent ways, thereby establishing the results' public interpretability (subject-independence). The article demonstrates how these principles can be meaningfully applied to psychical and social phenomena, considering their peculiarities and inherent limitations, revealing that not constructs in themselves but only their indicators (proxies) can be measured. These foundational concepts allow to distinguish measurement-based quantifications from other (subjective) quantifications that may be useful for pragmatic purposes but lack epistemic authority, which is particularly important for applied (e.g., legal, clinical) contexts. They also highlight new avenues for establishing transparency and replicability in empirical sciences.

Keywords Measurement · Constructs · Psychometrics · Quantitative methods · Replicability · Traceability

✉ Jana Uher
mail@janauher.com

¹ School of Human Sciences, University of Greenwich, Old Royal Naval College, Park Row, London SE10 9LS, UK

² London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

1 Introduction

In times of dramatic global changes (e.g., migration, climate) and profound real-world problems (e.g., mental health, populism), collective efforts are needed from all sciences, involving the physical sciences (e.g., chemistry, physics), life sciences (e.g., ecology, medicine), psychology and social sciences (e.g., sociology, linguistics, education), and their applied fields (e.g., engineering, economics, management, social policy). For these efforts, measurement is essential because it concerns the processes for producing accurate, reliable, comparable—thus, trustworthy quantifications (Mari et al. 2015).

Measurement and quantification are considered key to the physical sciences' successes over the last 300 years (Hand 2016). But controversies arose over the idea psychology and social sciences could capitalise on the advantages of quantification in similar ways (Fechner 1860; Stevens 1946; Thurstone 1928). Specifically, not all quantification occurs through measurement. *Quantification* denotes the assignment of numerical values; *measurement* is a purposeful multi-step process, comprising operative structures for making such assignments in reliable, valid and explicitly justified ways. Thus, measurement defines a process structure, quantification its result (Mari et al. 2017).

In the physical sciences, measurement and quantification build on concepts of *metrology* (science of measurement), which involve explicit and internationally accepted definitions, principles and standards (Czichos 2011). Metrological concepts are valid for all physical sciences, engineering and many life-science fields, and are used to design technologies that minimise direct involvement of humans in measurement processes. But physical technologies cannot be applied to the intangible research objects studied in psychology, social sciences and their applied fields. Instead, data about these study phenomena are often generated *directly by persons*¹ (e.g., interviews assessments, observations), and pertinent measurement theories and quantification practices (e.g., psychometric theories, rating scales) have been developed largely independently from those of metrology (Michell 2008; Stevens 1946; Torgerson 1958).

But how can concepts as fundamental to science as measurement and quantification be understood and applied in entirely different and even incompatible ways? What is science at all without some unifying ideas framing scientists' approaches to quantify the—necessarily different—research objects explored in different fields? Increasingly, scholars compare and aim to integrate metrological with psychological and social-science measurement concepts (Berglund et al. 2012; Finkelstein 2003; Fisher and Wilson 2020; Tobi 2014; Wilson et al. 2015). But, although all researchers using quantifications aim to exploit the powers of mathematics, a unified framework on which all sciences can build is still missing. Instead, diverse concepts, terminologies and practices are used, which hampers mutual understanding, identification of commonalities and differences, and establishment of common frameworks (Uher 2018a).

¹ Measurement directly by persons is variously conceived by metrologists as measurement with persons, humans as measurement instrument, human-based measurement, or persons as data generation systems (Berglund 2012; Pendrill 2014).

1.1 A transdisciplinary approach to comparing measurement practices

To compare different sciences' measurement theories and quantification practices regarding their most-basic underlying principles, this article relies on the frameworks of a transdisciplinary and philosophy-of-science paradigm. Its more abstract perspectives help reveal layers of information that are different from those commonly considered, thereby providing new insights to help advance current debates.

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

The *Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals (TPS-Paradigm*²; Uher 2015a, c; 2018c) suits the present purposes well because it is aimed at making explicit the presuppositions, metatheories and methodologies underlying given scientific systems (therefore *philosophy-of-science*) to help researchers critically reflect on; discuss and refine their theories and practices; and to derive ideas for new developments. It comprises a coherent system of interrelated philosophical, metatheoretical and methodological frameworks (therefore *paradigm*). In these frameworks, concepts from psychology, life sciences, social sciences, physical sciences and metrology that are relevant for exploring research objects in or in relation to *individuals* have been systematically integrated, refined and complemented by novel ones, thereby creating unitary frameworks that transcend disciplinary boundaries (therefore *transdisciplinary*). Moreover, the TPS-Paradigm puts into focus the individuals who are doing the research and generating the data to help open up a meta-perspective on research processes, as done in this article.

The TPS-Paradigm has already been applied (1) to integrate and expand on previous concepts of individuals' psyche, behaviour, language and contexts (Uher 2013; 2015a, c; 2016a, b); (2) to refine concepts and methodologies for comparing and taxonomising individual differences in various phenomena and populations (Uher 2015b, c, d, e; 2018b, c), and (3) to critically analyse the involvement of human abilities in data generation across the empirical sciences (Uher 2019) as well as raters' use of standardised assessment scales (Uher 2018a). These conceptual developments and analyses are demonstrated in various empirical studies (e.g., Uher et al. 2013a, b; Uher 2015d; Uher and Visalberghi 2016).

The present article expands on these works by comparing the epistemological, metatheoretical and methodological foundations of structural frameworks of measurement and measuring instruments from metrology with psychological and social-science theories and practices, focussing on constructs and their fiat measurement. But the aim is neither to comprehensively review a broad range of theories and practices from each field nor to provide full descriptions of those discussed because such are available in discipline-specific publications. Instead, important concepts are selected that serve to highlight commonalities and differences in the most basic principles underlying theories and practices used to quantify properties of physical versus psychical and social study phenomena. The aim is to elaborate a unified framework of metatheoretical and methodological concepts that will be needed to identify ways in which the most basic principles of measurement can be met in *all* sciences, while considering their study phenomena's inherent differences.

² Researchonindividuals.org.

1.1.2 Human factors: the lowest common denominator across all empirical sciences

To enable transdisciplinary comparisons, and in line with the TPS-Paradigm's focus on scientists' own role in research processes (Uher 2015a, c), analyses start from the fact that, in all sciences, measurement instruments and quantifications are created and used by humans (Berglund 2012). *Human factors* constitute the *lowest common denominator* in the making of all empirical sciences, which are, by definition, experience-based (from *Greek* empeiria meaning experience).

Every concrete experience has two aspects, the content given and individuals' apprehension of it—thus, the objects of experience in themselves and the subjects experiencing them (Wundt 1896). Accordingly, scientists treat experiences in two fundamental ways. Natural scientists consider the objects of experience in their properties as conceived independently of the subjects; this requires subtracting from the concrete experience the subjective aspects always contained in it using the perspective of *mediate experience* (mittelbare Erfahrung; Wundt 1896). Therefore, natural scientists develop theories, approaches and technologies that help minimise these human factors' involvement and filter out their effects.

Psychologists and social scientist, by contrast, explore the experiencing subjects and their apprehension of the experiential contents using the perspective of *immediate experience* (unmittelbare Erfahrung; Wundt 1896). They study subjects' understanding and interpretation of the experiential contents and how this mediates individuals' concrete experience of 'reality'. Their research object is (inter-)subjectivity with all its complexity, diversity and possible irrationality—thus, human factors in themselves. This entails challenges because scientists can never step outside of their own position in their socio-linguistic and cultural world. Psychologists and social scientists must therefore critically reflect on their *own* human factors and how these may (unintentionally) influence their explorations of others' experience and (inter-)subjectivity. This constitutes a major difference to the natural sciences and requires fundamentally different approaches and methods. It also affects the meaning and utility that quantifications could have for investigating these phenomena, and the possibilities for establishing measurement processes, as explored in this article.

1.1.3 Terminological fallacies

Transdisciplinary comparisons are complicated by various terminological fallacies that arise because, in different disciplines, the same term may refer to different concepts (jingle-fallacies; Thorndike 1903) or different terms to the same concept (jangle-fallacies; Kelley 1927). To facilitate cross-scientific understanding, such fallacies will be highlighted. Terms will be used that may express the essential ideas of given concepts most clearly, rather than favouring the terms of just one science (discipline-specific terms will be put in parentheses where this might be helpful). This requires readers to tolerate and deal with a terminology that, necessarily, diverges from any monodisciplinary standard. The aim is to make accessible discipline-specific concepts to readers from various fields to help build bridges, promote cross-scientific exchange and collaboration, and jointly develop measurement concepts applicable to all sciences.

1.2 Article outline

As foundation for the transdisciplinary comparisons of measurement practices, the article first highlights peculiarities of the different sciences' research objects using metatheoretical concepts from the TPS-Paradigm. It introduces a metatheoretical definition of data and methodological concepts that highlight how human perceptual and conceptual abilities are involved in data generation in any science (Sect. 2). These foundations are then applied to metatheoretically analyse the structural measurement processes established in metrology and physical sciences. The article elaborates two basic methodological principles by which physical scientists, starting out from nothing but their human abilities, have developed technical instruments that help overcome limitations in human perceptual abilities to enable measurement of a broad range of physical properties, including imperceptible ones (Sect. 3). These methodological principles are then compared with those underlying measurement theories and quantification processes used in psychology and social sciences. These transdisciplinary analyses focus on constructs and their measurement by fiat, revealing fundamental differences to metrological concepts not yet well considered (Sect. 4). But the two methodological principles also highlight important commonalities in the ways in which measurement-based quantifications can be generated across all sciences, considering their research objects' inherent peculiarities (Sect. 5).

2 Relevant metatheoretical and methodological foundations

2.1 The sciences' objects of research: peculiarities and concepts

The intangible properties of many psychical and social phenomena (e.g., psyche, social relationships) complicate their definition, differentiation and investigation. Moreover, psychologists' study phenomena involve also those (e.g., conceptualising) by which any science is made (Valsiner 2012); therefore, psychologists must distinguish their study phenomena from the means for exploring them, as reflected in the terms *psychical* versus *psychological* (from Greek -λογία, -logia for body of knowledge; Lewin 1936; Uher 2016a). The TPS-Paradigm provides metatheoretical concepts, integrated and refined from various disciplines and historical lines of thought, that define, describe and differentiate properties of various kinds of phenomena³ studied in or in relation to individuals (for details, Uher 2015a, c; 2018a, c; 2019). The following outlines some relevant concepts.

2.1.1 Formalising modes of accessibility, conceptual differentiations and methodological implications

To highlight essential differences among the sciences' study phenomena and to help formalise their modes of accessibility to human perception under everyday conditions (and thus also the ways to make them accessible under research conditions), the TPS-Paradigm

³ A *phenomenon* is defined as anything that humans can perceive or make (technically) perceivable and/or that humans can conceive (Uher 2015c). Note this notion differs from some theoretical and philosophical definitions (e.g., Kant's 1998).

considers three metatheoretical⁴ properties. These are (1) *location* relative to the studied individual's body (internal–external dimension), (2) *temporal extension* (transient–temporally stable dimension), and (3) *spatial extension*, conceived complementarily as physical (spatially extended) *versus* “non-physical” (without spatial properties). Physicality⁵ denotes corporeal, bodily properties of material phenomena as well as properties that are not corporeal in themselves but become manifest in material phenomena with which they are systematically connected, thus immaterial physical.

Physical phenomena can be described in terms of their spatial properties (even if only subatomic), whereas spatial properties cannot be conceived at all for “non-physical” phenomena (e.g., psyche), which are therefore not simply contrasted against the physical but conceived as complementary instead (indicated by the quotation marks). This distinction resembles Descartes' *res extensa* and *res cogitans* (Descartes et al. 1983) but implies only a methodical and not also an ontological dualism (Uher 2015c, 2016a, 2019). This follows the concept of *complementarity*,⁶ which emphasises the necessity to account for the observation of two categorically different realities that require different approaches, frames of reference and criteria of truth, such as the wave-particle duality of light and matter (Bohr 1937; Heisenberg, 1927) and psyche-physicality (body-mind) properties (Brody and Oppenheim 1969; Fahrenberg 1979, 2013; Walach and Römer 2011).

The particular constellation of metatheoretical properties that can be conceived for study phenomena also enables their *conceptual differentiation* as well as derivation of methodological concepts for investigations. This is now briefly illustrated in three study phenomena relevant for the present analyses—behaviours, psyche and constructs.

2.1.2 Behaviours: immaterial but physical phenomena external to individuals

Behaviours, defined as the “external changes or activities of living organisms that are functionally mediated by other external phenomena in the present moment” (Uher 2016b, p. 490), involve properties that are externally located, transient and (mostly immaterial) physical (e.g., movements, vocalizations, secretions). Their public accessibility enables multiple persons to jointly perceive the *same* behavioural acts and the *same* entities of the properties studied in them using so-called *extroquestive*⁷ methods.⁸ Extroquestive accessibility helps establish inter-subjectivity, an important meta-condition of measurement (see

⁴ Metatheoretical here indicates their abstract level of consideration; only time and space constitute ontological categories.

⁵ Physicality here refers to classical physics concepts, because they match everyday experiences in contrast to those of quantum physics.

⁶ The TPS-Paradigm builds on the concept of *complementarity*, describing two mutually exclusive properties of a research object that are irreducible and maximally incompatible with one another but both required for an exhaustive understanding of it, and that may thus be regarded as complementary to one another (Bohr 1937; Brody and Oppenheim 1969; Fahrenberg 1979; Heisenberg 1927).

⁷ The concepts of *extroquestion* and *introquestion* (from the Latin *extro* for beyond, outside; *intro* for in, within; and *quaerere* for to seek, enquire) were developed to remedy conceptual leaps in previous concepts of *introspection* versus *extrospection*, commonly defined as inward versus outward perspectives. These latter concepts cannot be differentiated from one another as methods because both perspectives are always contained in any given perception. The new concepts highlighted that psychophysical experiments, commonly classified as *introspective* methods, actually constitute *extroquestive* methods (for details, Uher 2016a, 2019).

⁸ Extroquestive methods are defined as “procedures for studying phenomena that individuals can perceive as from outside of themselves (e.g., behavioural observation) and that are (or can be made) perceptible by multiple individuals (e.g., using invasive methods like surgery; Uher 2019, p. 231).

below). Behaviours' spatial properties enable application of physical methods⁹ of investigation (e.g., pedometer). Their transience and processual nature requires methods enabling their real-time capture, called *nunc-ipsum* methods¹⁰ (Uher 2019). This constellation of metatheoretical properties differs from those conceived for the psyche.

2.1.3 The phenomena of the psyche: experiential processes

The psyche is defined as the “entirety of the phenomena of the immediate experiential reality both conscious and non-conscious of living organisms” (Uher 2015c, p. 431), with immediacy indicating absence of phenomena mediating their perception (Wundt 1896). The particular forms regarding the three metatheoretical properties that can be conceived for psychical phenomena highlight peculiarities that complicate their accessibility to investigation (Uher 2016a). Their lack of spatial properties and of systematic relations to the physical phenomena with which they are connected (e.g., brain morphology, physiology)—reflecting complementary psyche-physicality (body-mind) relations—make psychical phenomena inaccessible to physical technologies (Fahrenberg 1979, 2013). Psychical phenomena are conceived as located entirely internal to individuals’ bodies, directly perceptible by each individual itself but inaccessible to others (Locke 1999), requiring so-called *introquestive*⁷ methods¹¹ of investigation.

Temporal properties vary. Ongoing psychical events (e.g., thoughts, emotions) are transient, therefore called *experiencings* (Erleben). Temporally more extended phenomena (e.g., beliefs, knowledge, mental abilities) are called *memorised psychical resultants* (experiences, Erfahrung), with memorisation referring to any retention process. But, although temporally extended, they are accessible only in individuals’ *experiencings* and must be reconstructed in each moment anew within the given context, whereby they are adapted and changed before becoming memorised again (Schacter and Addis 2007). Therefore, psychical phenomena must be conceived as *occurrents* (perdurants in formal ontology)—as *processes*.

Of processual entities, only a part exists at any moment so that they cannot be determined without knowledge of previous occurrences. Occurrents are opposed to *continuants* (endurants in formal ontology), which do exist in their entirety at any moment (e.g., material objects). As processes, psychical phenomena can be conceived only through *abstraction* from their occurrences over time. This leads to beliefs and knowledge *about* them, which are psychical phenomena in themselves as well, but not the same as those they are about (see Uher 2015d, 2016a; similarly Whitehead 1929).

The psyche’s capacities for abstraction are essential for thinking, and thus for the making of science. Abstractions also constitute important study phenomena in themselves.

⁹ Physical methods are defined as “procedures relying on the spatial extension of materials that are systematically related to the physical phenomena [and their properties] under study, directly perceptible and easier to demarcate and categorise than the study phenomena [and properties] (e.g., length of a spring scale to measure weight; Uher 2019, p. 231).

¹⁰ Nunc-ipsum methods (from the Latin *nunc ipsum* for at this very instant) are defined as “procedures enabling the real-time recording of transient phenomena (e.g., behavioural observation, experience sampling”; Uher 2019, p. 231).

¹¹ Introquestive methods are defined as “procedures for studying phenomena perceptible only from within the individual itself and that cannot be made perceptible by multiple individuals under all possible conditions (e.g., inner self-observation, self-report”; Uher 2019, p. 231).

2.1.4 Constructs as study phenomena: abstract conceptual entities

Many psychological and social-science objects of research are abstractions and complex ideas that are theoretically constructed by humans, therefore called *constructs* (Slaney 2017). Examples of these conceptual entities are ‘intelligence’, ‘socio-economic status’, ‘populism’ but also ‘climate’, ‘biological fitness’, and ‘heritability’ studied in the life sciences. Their abstract theoretical nature entails that any given construct always refers to several concrete entities, which may involve occurrences and continuants of physical phenomena (e.g., behaviours, temperature, material objects) but also various “non-physical” phenomena (e.g., emotions, thoughts, social relationships). Abstraction involves that some aspects of the concrete entities to which a construct refers are emphasised and others deemphasised (Whitehead 1929). Differences in the particular referents, aspects and levels of abstraction that persons (implicitly) consider enable unparalleled proliferation, complexity and thus changeability in the constructs created. Therefore, *theoretical definitions* of constructs meant to denote the same conceptual entity can vary (e.g., different definitions of ‘socio-economic status’ or ‘intelligence’) and, as a consequence, also the operational definitions devised for generating data about them (see below).

2.2 Data generation across the sciences: metatheoretical and methodological concepts

To enable transdisciplinary comparisons and considering the role that human factors play in all empirical sciences, both technical measuring instruments and the data-generating persons *in themselves* must be analysed for the functions they fulfil in measurement processes. This is seldom done in any science. For this purpose, a metatheoretical definition of data and methodological principles of data generation highlighting the involvement of human abilities are now briefly outlined and then applied to pinpoint key differences in measurement practices among sciences (Sects. 3 and 4).

2.2.1 What are data? A semiotic definition

The signs used to indicate quantifications (e.g., Arabic numerals, Latin letters) and to which particular scientific communities attribute particular meanings (e.g., mathematical properties) are commonly called *data*.¹² As signs (e.g., variable names, values), the function of data is to represent in physically persistent ways (e.g., print, digital) information about properties of the study phenomena as conceived by the data-generating persons. These representational functions of signs are so deeply engrained in our everyday language and thinking that we seldom become aware that any sign comprises three constituents. These are (1) a physical constituent (e.g., visual ink patterns) used as *signifier* that symbolically represents (2) the *referent*, the actual object of consideration to which it refers (e.g., property, physical object), and (3) the *meaning* (the *signified*) that both have for the sign-using persons, which in itself is a psychical phenomenon (Fig. 1; similarly Ogden and Richards 1923).

¹² This semiotic notion of data corresponds to the linguistic habits of many empirical researchers but differs from information-theoretical notions, in which data are conceived as elements of information (e.g., Floridi 2019).

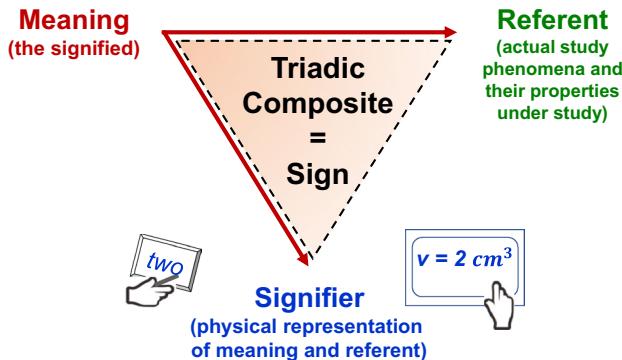


Fig. 1 Data as semiotic representations comprising three composites

These triadic interrelations among signifier, referent and meaning, involving both physical and psychical phenomena, are conceived in the TPS-Paradigm's metatheoretical concept of *semiotic representations*. It specifies, on an abstract level, the basic ideas underlying sign systems (e.g., written and spoken language; Uher 2015a, 2016b, 2018a, 2019). The term *representation* highlights that it is persons' psychical representation (meaning) that connects a signifier with its referent, thereby establishing the triadic relationship that first turns this composite into a sign and creates its functionality. This highlights that a sign is more than just its signifier (as common parlance often implies) because its meaning is not inherent to the signifier itself but only assigned to it. Therefore, the same signifier (e.g., visual patterns like I , V , X) can have different meanings (e.g., Roman numbers or letters). Which particular meaning a signifier has for particular persons and which particular referents it represents for them is not directly apparent from the signifier itself (with very few exceptions, e.g., icons).

Semiotic representations have important functions for abstract thinking—and for measurement. They allow humans to represent perceivable phenomena and their properties (e.g., one green bean) in single words (e.g., written or spoken as ‘one’, ‘green’, ‘bean’). Words enable us to make concrete entities (referents) independent of their immediate perception and to abstract them into objects of consideration (conceptual entities, the signified)—thus, *reifying* them (e.g., ‘quantity’, ‘green colour’, ‘beans’). Through this so-called *hypostatic abstraction* (Peirce 1958, CP 4.227), we develop words that refer to concrete referents not only while we can perceive them but also in their absence, thus abstracted from the ‘here and now’. It also allows us to develop abstract words that have not concrete but abstract referents, such as concepts and ideas describing phenomena and properties that are distant from immediate perception (e.g., ‘vegetables’) or imperceptible in themselves (e.g., ‘quantity’, ‘nutrition’)—thus, constructs. Hence, every word is a concept in itself (Khanam et al. 2019; Vygotsky 1962)—an important point for language-based methods of data collection (see below).

These metatheoretical analyses highlight that lexical and numerical data (e.g., variable names and values) can be used to represent information about research objects (referents) in various degrees of abstraction, ranging from properties directly perceivable at given moments, over those that can only be inferred from perceivable ones, up to abstract ideas that are only construed by humans but do not exist as concrete entities in themselves (constructs). But the level of abstraction represented by particular data is not apparent from the

signifiers in themselves (e.g., written words, mathematical symbols). This has important implications for data generation, analysis and interpretation, especially regarding psychological and social-science constructs (see below).

2.2.2 Conversions of information: the essence of data generation

When assigning quantitative values during measurement execution, information about the objects and properties under study are encoded into the signs used as data. When information¹³ from one kind of phenomenon (e.g., physical objects, behaviours) are represented in another kind of phenomenon (e.g., signifiers printed on paper), this is called *conversion of information* in the TPS-Paradigm. This term is very broad; for any specific case, it requires specification of what kind of information is converted in what ways into what other kind of information. This is commonly done explicitly in metrology, but not so in psychology and social sciences (Uher 2018a). In metrology, engineering and also in psychophysics, information conversion is commonly called transduction; in other fields, also translation or transcription (e.g., molecular biology). But unlike those, the concept of information conversion explicitly refers to person-executed processes and specifies possible sources of the (considerable) losses and inaccuracies that may inevitably occur in them (detailed in Uher 2019). Information conversions are the essence of any data generation, whether executed by persons directly or using technical measuring instruments (see below).

2.2.3 Person-based measurement execution and data generation: abilities and decisions required

For every person-executed conversion of quantitative information (e.g., reading scale displays of measuring devices, observing and encoding behaviours), persons must make decisions about how to identify the information of interest in the study phenomena. In all sciences, however, measurement theories commonly do not explicitly consider the role that measurement-executing persons in themselves must fulfil in measurement processes and what abilities and decisions are required from them.

Three important tasks must be accomplished in any data generation: demarcation, categorisation and encoding (information conversion; Uher 2018a). First, in the multifaceted perceptions available at any moment, data-generating persons must be able to reliably *demarcate* the entities of interest using similarities and dissimilarities in the study phenomena's properties. For measurement, this must involve *both* qualitative and quantitative properties. This is because *quantity* denotes divisible properties of entities of the same kind, thus of the same quality, whereas *quality* denotes properties of different kind (Hartmann 1964). Accordingly, measurement-executing persons must first determine the study properties' *quality* and then compare entities of the same quality regarding their divisible properties. This presupposes that the qualitative and quantitative properties used for demarcation are (made) directly and accurately *perceivable* for the data-generating person. Temperature, for example, is directly perceivable but not accurately enough so that entities cannot be reliably demarcated, whereas directly perceivable material phenomena (e.g., tubed

¹³ In the TPS-Paradigm, the term information denotes relations among elements (in the mathematical sense as set members, not as chemical elements); these elements are not considered data in themselves (as done in information-theoretical definitions of data). Instead, data are conceived as semiotic representations.

mercury), given their corporeal and temporally more extended properties, enable reliable demarcations. Some study phenomena feature considerable variations in their perceivable properties (e.g., spatial extensions of biological cells and behavioural acts vary). This complicates demarcations and requires data-generating persons to make decisions about what constitutes *one* entity (e.g., individual cells, single acts).

Second, measurement-executing persons must *categorise* the entities thus-demarcated (e.g., individual cells into cell types, single behavioural acts into behavioural categories). This involves not only consideration of their perceivable properties (e.g., qualities, different structures) but often also *theoretical and contextual interpretations*, especially in psychical and social phenomena. Behaviours, for example, are commonly categorised by their known (or assumed) functions because perceptually similar acts can have different meanings in different contexts (Uher 2015b).

Third, measurement-executing persons must convert information from the entities thus-categorised into information encoded in the data. For systematic and standardised encoding, scientists must specify all three constituents of the particular signs used as data. That is, they must specify the decisions that the measurement-executing persons have to make about which pieces of information from the phenomena and properties under study should be demarcated and categorised in what ways, and the rules by which these should be assigned to the signifiers (e.g., mathematical symbols, lexical descriptions). These specifications must be made explicit and involve properties that are directly and accurately *perceivable* by data-generating persons *during* measurement execution (Uher 2018a, 2019).

Developing inter-subjective agreement in demarcation, categorisation and encoding is facilitated when the study phenomena are (or can be made) publicly accessible, thus *extroquestively*. But in phenomena that cannot be made publicly accessible by any means and are accessible only to each individual, thus only *introquestively* (e.g., psychical phenomena), inter-subjective agreement can be developed only indirectly and always involves uncertainty about the actual entities and properties considered. This has particular implications for psychological and social-science measurement (see below).

These metatheoretical and methodological foundations are now applied to scrutinise and compare the different sciences' theories and practices of measurement.

3 Metrological concepts of measurement: metatheoretical and methodological foundations

Metrologists emphasise that *measurement* specifies not only a functional relationship connecting a measurement system's input with its output. Measurement is a purposeful multi-stage process characterised by its structure that guarantees the reliability of the results produced and that justifies the ways in which this is achieved (Maul et al. 2018). Basic metrological concepts are now explored using a terminology and examples that are primarily aimed at making them accessible to psychologists and social scientists.

3.1 Methodology versus methods: different levels of scientific enquiry

For metrologists, the generic description of the structure of a measurement process, including the logical organisation of all operations involved, is called a *measurement method*. It is a component of a more complex system, which starts from a measurement principle and also includes a measurement procedure (JCGM200:2012 JCGM200 2012, def. 2.5). This

metrological use of the term method, however, is confusing for psychologists and social scientists for whom it has a different meaning (jingle-fallacy).

In psychology and social sciences, method is distinguished from methodology. *Methodology* denotes the system of principles underlying the conduct of scientific enquiry. It provides the philosophical and theoretical underpinning of the ways (approaches) in which research objects can be explored and that make particular operations suited for this purpose and others not, together with explanations of what their results indicate and why. *Method*, in turn, denotes the selection and construction of specific behaviours and instruments (practices, techniques) used to perform particular research operations (e.g., observing, videotaping, interviewing, self-reporting). Hence, methodology is the higher-order concept; it comprises the classification of methods together with their underlying philosophical and theoretical rationales (Kothari 2004).

In metrology, this important differentiation is often not made explicit, likely because epistemologies, which can lead to fundamentally different methodologies, are much less diverse than those used in psychology and social sciences (Sect. 4). Nevertheless, metrological concepts incorporate methodology as well, though under different terms, and, confusingly for other scholars, labelled as method.

3.2 Structure of measurement processes

Metrologists conceive measurement as a complex structured process in which a measurement task is fulfilled by applying theoretical and methodological principles and by executing procedures and techniques (methods) for data generation (Mari et al. 2017).

The *measurement task* must be defined by specifying the *objects under measurement* (e.g., individuals I , their bodily entities E and behaviours like talking activities T), the *property of interest* (e.g., body length L , temporal duration D), and the *measurands*, thus the specific entities to be measured in the study property (JCGM200:2012, 2012, def. 2.3), such as the body lengths l_a , l_b and l_c of target individuals i_a , i_b and i_c , and the temporal durations d_a , d_b and d_c and average sound volumes v_a , v_b and v_c of their talking activities t_a , t_b and t_c during a meeting m_j .

Then, the *general model of the measuring system*—the measurement methodology—must be specified involving the design of measurement procedures and measuring instruments as well as explanations of how they enable capturing the study property (e.g., loudness of talking). Therefore, scientists design process structures that enable *empirical interactions* with the study property (e.g., sound volume). Measuring systems are based on the (necessarily idealised) identification of systematic (lawful) structural connections among properties or at least specific assumptions about such connection networks, which allow scientists to check measurement results via experimental cross-validation (Mari et al. 2017).

From the general model (methodology), metrologists then derive a *specific model of the measuring system* regarding the specific study objects that bear the measurands (e.g., target individuals i_a , i_b and i_c , their bodily entities e_a , e_b and e_c , and talking activities t_a , t_b and t_c in meeting m_j). This also involves a model of the measurands, thus the specific entities to be measured in the study property, such as the lengths l_a , l_b and l_c of these individuals' bodies or the temporal durations d_a , d_b and d_c of their talking behaviours. In such models, necessarily, the objects and the properties studied in them—as they occur in the actual world—are idealised and approximated (abstracted) to demarcate them as entities from the

actual world, corresponding to the particular entities intended to be measured (Mari et al. 2017).

Thereafter, scientists must define the specific parameters to be measured (encoded in variables) as well as the *measurement model* comprising assumptions on their interrelations including any calculation to produce the measurement result (e.g., weighting). This involves the operational definition of the objects and properties studied (called operationalisation in psychology and social sciences). Scientists must also specify the particular procedural operations (methods) that the measurement-executing persons have to perform given the theoretical and methodological specifications made. These operations must enable an empirical interaction between study property and measuring instrument during measurement execution. All this occurs before persons can execute the actual measurement procedure to generate results (e.g., by using technical instruments). For simplicity, the various types of measurement uncertainty involved along this multi-stage process are not considered here (for details, Mari et al. 2015, 2017).

To distinguish measurement from other processes of evaluation (e.g., opinion making), metrologists specify two meta-conditions of measurement and concepts for implementing them in measurement processes.

3.3 Object-dependent and reproducible measurement processes

A first metrological meta-condition is that measurement processes must be designed from knowledge about the objects and properties studied, therefore called *object-dependence*, *object-relatedness* or *object-ivity* (Mari and Wilson 2015). It requires explanations of the ways in which specific operative structures enable the assignment of specific numeric values to the measurands (entities to be measured in the study property) such that these values reveal reliable and valid information about them. An important condition is that these processes are able to convey information specifically on the measurand, and not also on other properties featured by the research object or states of the surrounding. Instead, the process design should minimise such influence properties' effects (Mari et al. 2017). The ability to reproduce a given measurement process on the same or similar objects given particular conditions, including changes in the experimental context in which the results are achieved (e.g., locations, operators, measuring instruments), is called *measurement reproducibility* (Mari et al. 2015). Note, reproducibility here refers to the process, not only to the results.

The term object of research is not commonly used by psychologists and social scientists who mostly study abstract ideas and other intangible phenomena (e.g., psyche, social structures) in or in relation to individuals (formerly subjects, now participants). Therefore, the term object-dependence cannot have the same denotation as in metrology. But independent of that, an analogue concept is lacking in psychological and social-science measurement (see below).

3.4 Subject-independent results

A second metrological meta-condition is that measurement processes do not depend on the opinions of the persons (subjects) operating them. Instead, the process design must ensure that results are invariant with respect to the persons involved; therefore called *subject-independence*, *subject-transparency* (terms uncommon in psychological and social-science measurement) or *inter-subjectivity*. It requires that the quantitative values assigned to the measurands must be univocally interpretable in different places and times, thus in the same

ways by the persons generating and those using them. Hence, subject-independence is a condition of the results' public interpretability, which ensures that results always represent the same information regarding the measurands (Mari et al. 2017).

The term inter-subjectivity is used in similar ways in psychology and social sciences, where it primarily refers to individuals' shared agreement (e.g., in perception, interpretation or meaning of something). But, in these fields, objectivity is opposed to subjectivity and denotes independence of results from the investigator (e.g., test administrator). Accordingly, objectivity is commonly interpreted as inter-subjectivity and not as alignment to the object of research, thus confounding two metrological key concepts of measurement. This highlights a profound cross-scientific jingle-fallacy in the term objectivity, which is therefore not used here.

An important means to establish object-dependence in measurement processes and their results' subject-independence is implementation of two types of *traceability*. They highlight two basic methodological principles underlying metrologists' structural frameworks of measurement that are applicable also across sciences (see below).

3.5 Data generation traceability

To justify that the generated results are attributable to the objects and properties studied, numerical assignments must be systematically connected through *unbroken documented chains of comparisons* to the measurand and a reference (e.g., standard unit). Every step in the chain involves the possibility that the entities of the connected properties (e.g., measurand and measurement unit) can be compared with one another regarding their quantities (Mari and Wilson 2015) so that quantitative information from one property can be converted (transduced) into quantitative information in another property.

By implementing unbroken documented chains of quantitative information conversion, scientists establish object-dependence in the measurement process. This allows to trace the measurement results thus-created, in the inverse direction, back to the measurands and the particular references (e.g., standard units) used to quantify them (Fig. 2); this is called *data-generation traceability* in the TPS-Paradigm (Uher 2018a). Maybe this kind of traceability is so self-evident for metrologists and already implied by their concept of object-dependence, that it is not explicitly mentioned in metrological research, which focusses only on numerical traceability (see next). But data generation traceability underlies all measuring instruments and highlights essential metatheoretical principles for their construction (see below). Moreover, and importantly, it is a key concept in which metrological and physical measurement processes differ from many psychological and social-science practices of quantification (see below) and is therefore conceived as a separate concept in the TPS-Paradigm.

3.6 Numerical (metrological) traceability and references

The universal (subject-independent) meaning of numerical values assigned in measurement processes (e.g., the specific length of 1 m) arises from internationally accepted conventions about explicitly defined standard units that are systematically connected through unbroken documented calibration chains to primary references. A *primary reference* can be a measurement standard or the definition of a measurement unit through its practical realisation, such as an object (e.g., prototypes), a system, or an experiment involving a defined

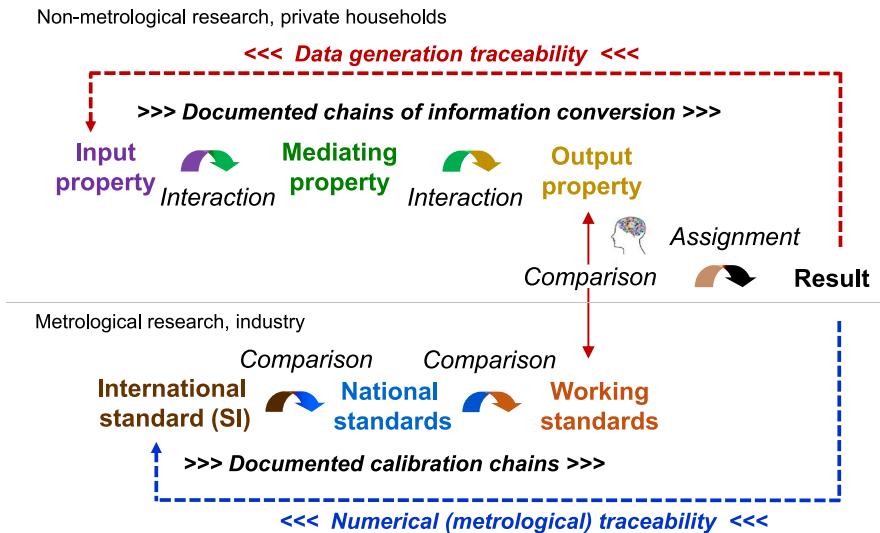


Fig. 2 Data-generation traceability and numerical traceability

relationship to the quantity of interest (JCGM200:2012, 2012). For many physical quantities, metrologists have defined primary standard references (e.g., international prototype kilogram; Quinn 2010). These are linked through *unbroken documented chains of comparisons* (called *calibration chains*), first, to different secondary references (e.g., national standards) and, from there, to working references (e.g., measuring sticks in science labs and private households). For every comparison, metrologists specify *uncertainties* as a quantitative indication of a result's quality and reliability (JCGM100:2008, JCGM100 2008). Documented calibration chains ensure that any comparisons with working references that are traced to the same primary standard reference will produce comparable results for the same measurand (De Silva 2002) that can thus be understood all around the globe in the same ways, thus subject-independently (inter-subjectively). This is called *metrological traceability* (Mari et al. 2015). For applications across the sciences, the underlying methodological concept will be called *numerical traceability* in the TPS-Paradigm.

3.7 Measuring instruments: establishing documented unbroken connection chains between measurands and results

For direct comparison with psychological and social-science practices, it is useful to consider the simplest and historically oldest physical measuring instruments because, in them, involvement of human factors is still greater and more directly apparent than in today's sophisticated measuring technologies. These latter build on knowledge gained from these older instruments yet involve more complex physical processes, many of which imperceptible by humans.

In measuring instruments, metrologists implement documented unbroken connection chains that must start from the specific property to be measured in the study object, the *input property* (measurand), and its interaction with a first *mediating property* that is systematically connected to it (object-dependence). In the simplest case, this mediating

property, in itself, is directly and accurately perceivable by measurement-executing persons. In the above example of measuring individuals' body lengths, material objects are used that feature the property of length in dimensions easily and accurately perceivable for humans (e.g., wooden sticks) and that enable direct comparison with the measurands. The sticks' spatial and temporal extensions enable multiple persons to directly and jointly perceive the same entities of length in them. This extroquestive accessibility facilitates the definition of identical (or at least highly similar) entities that different people can reliably and inter-subjectively (subject-independently) demarcate and that can be marked on sticks in standardised and persistent ways for use as references. To generate results, measurement-executing persons must directly compare the measurands' length with the length of the units marked on the measuring stick. Person-executed comparison is possible because both, measurand and units, constitute quantities of the *same* property (length). They must convert the information obtained from this comparison, such as by counting units, into information encoded in the given signs used as data (e.g., specific lexical and numerical symbols encoding quantity values and length units).

Many physical properties, however, can be perceived by humans either not accurately enough, not easily or not at all (e.g., weight, colour, density). Then physical scientists introduce a further mediating property (called *sensible transducer*) that *a*) is sensitive to and structurally (lawfully) connected with the input property and that *b*) can be connected in turn to another property, thereby establishing a systematic mapping (Mari and Wilson 2015). From the first mediator's interaction with the input property, the quantity information can be converted stepwise into further, likewise systematically connected mediating properties, whereby the result of each conversion step depends on the result of the previous. This unbroken documented connection chain is continued until it is possible to convert the information, on the person-side end of the conversion chain, into a property that

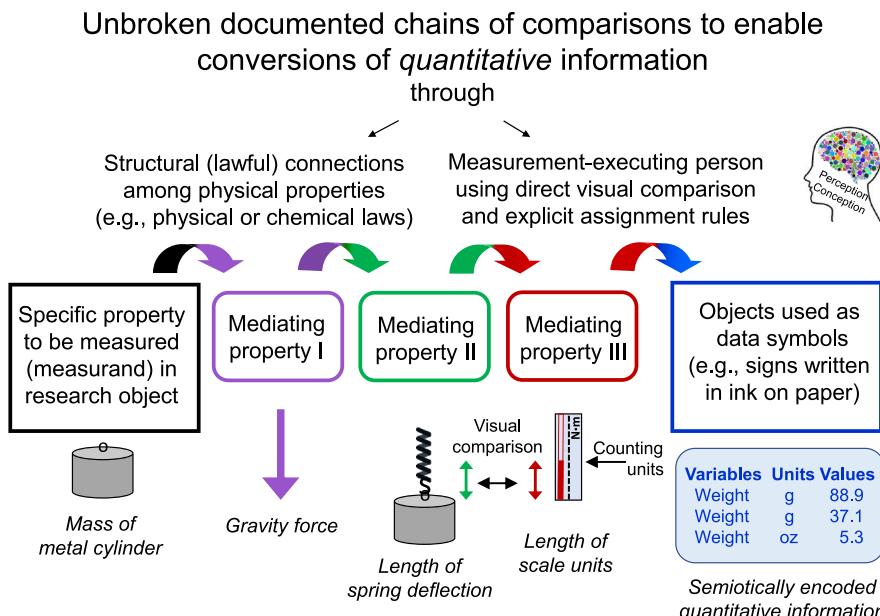


Fig. 3 Methodological principles underlying measuring instruments

persons can directly, reliably and inter-subjectively perceive (e.g., length of tubed mercury) for comparison with measurement units.

Spring scales illustrate this principle. A metal cylinder's specific mass (measurand) is connected to the gravity force (mediator 1) acting on it (on earth). Gravity force acting on the object (called its weight) is also connected to the deflection of a spring (mediator 2) when the object is attached to it and that scientists, in turn, directly connect to a scale display with equal units marked on it (mediator 3). The properties 'mass'—'gravity force'—'length of spring deflection' are chained by physical laws, which establish proportional relations among the specific quantities of these different properties. The connection between 'length of spring deflection'—'length of extension over scale' is established by the measurement-executing person through visual comparison. The person also executes the final step by applying unchanging rules for converting the quantitative information thus-obtained (e.g., by counting scale units) into semiotic information encoded in the lexical and numerical signs (variable names, values) serving as quantitative data (results; Fig. 3). In digital instruments, these last two steps are automatized to further reduce the involvement of human factors in measurement processes. Analogously, when measuring time using sand glasses, it is gravity and sand that provide mediating properties for stepwise conversions of quantitative information about time periods into quantitative information directly, reliably and inter-subjectively (extroquestively) perceptible properties (sand grains in the hourglass compartments).

3.8 Conclusion: two basic methodological principles of measurement

In summary, two basic principles are crucial for measurement. (1) *Data generation traceability* requires that the particular ways in which measurement results are assigned to the specific properties to be measured in the study objects must be fully transparent and therefore traceable. This is achieved by designing measurement processes that systematically connect the measurand through unbroken and documented links with the measurement result assigned to it (object-dependence). Therefore, measurement reproducibility refers to the process, not just to the results generated. (2) *Numerical traceability* requires that the numerical value of the measurement result is also linked to known standards, in documented and transparent ways, thereby establishing its inter-subjective meaning (subject-independence). These two general methodological principles underlying metrologists' structural frameworks of measurement can also be applied in psychology and social sciences, although in different ways and not for all study phenomena. This highlights commonalities, and thus comparability that can be established across sciences, but also fundamental differences as explored now.

4 Psychological and social-science concepts of measurement: metatheoretical and methodological foundations

Study phenomena, theories and research practices in psychology and social sciences are extremely heterogeneous. They involve, for example, natural-science and technology-based investigations of individuals' morphology (e.g., neuro-imaging), physiology (e.g., skin conductance) and behaviour (e.g., life-logging), software-based explorations of behaviour (e.g., video-analysis), investigations of textual data from individuals' verbal

interactions and written documents (e.g., text mining, machine learning), or economic accounts of individuals' material wealth (e.g., income). They also involve explorations of the intangible phenomena of the psyche (e.g., thoughts, emotions) and those emerging from individuals' social and societal interactions (e.g., language, politics, culture) that are mostly studied in constructs and for which physical technologies cannot be applied. These latter are in the focus here.

4.1 Plurality of epistemologies and methodologies

Epistemology and especially methodology correspond to metrologists' specification of the measurement method and the general model of the measuring system, comprising a generic description of the process structure including logical organisation of all operations. But the sciences' different approaches for dealing with experience entail fundamental differences. Metrologists and natural scientists focus on the contents of experiences and therefore develop technologies that help minimise involvement of human factors in empirical investigations and filter out their effects. Psychologists and social scientists, by contrast, explore how subjects understand and interpret the contents of their experiences and how this subjective apprehension mediates their concrete experience of 'reality'. This highlights the "non-physical" and conceptual nature of these study phenomena and their unparalleled complexity, variability and changeability. Moreover, the aims for which these are being studied and the perspectives taken on them vary greatly.

All this entails a plurality of ontological and epistemological concepts, each describing different aspects of and even entirely different perspectives on people's psychical and social 'reality' as well as different general approaches by which knowledge about these phenomena can be gained. *Realists*, for example, assume humans could access 'reality' rather directly and accurately. *Positivists* are less concerned with finding true explanations of 'reality' and focus on empirical evidence and predictive utility instead. *Constructivists* emphasise the pronounced influence that socio-cultural beliefs have on people's perception and conception of 'reality', which therefore differ among individuals and communities (including physicists regarding their own science; Hossenfelder 2018). These are only three of many different epistemological stances, which inevitably influence the methodologies that psychologists and social scientists derive from them and in which they specify the theory and philosophy of operations and techniques (methods) that enable access to and investigation of these study phenomena.

Measurement and quantification are of primary interest to realists and positivists, whereas, to explore people's meaning making and the constructions and interpretations of their individual and social 'realities', quantifications are often uninformative. Therefore, constructivist epistemologies and methodologies entail so-called *qualitative methods*, which involve analytical techniques to extract the qualities, structures and interrelations of meanings from textual and other verbal data (e.g., transcripts, social media texts) as well as systematic techniques and operational strategies to generate (or select) therefore suitable data (e.g., interviews, narratives)—often without aiming to obtain numerical information as well. Qualitative methods are contrasted with operations and techniques for generating and analysing numerical data, commonly called *quantitative methods* (e.g., rating scales). But this qualitative-quantitative dichotomisation, widespread in these fields, is inaccurate and misleading because it implies the idea

that quantitative data could reflect pure quantities, ignoring that quantities are always *of* something—qualities. Any investigation requires specification of the particular qualitative properties studied and only some methods *additionally* enable collection of quantitative information about them (Uher 2018a).

4.2 Fiat ‘measurement’ of constructs: fundamental challenges for implementing traceable conversions of quantitative information

The study phenomena’s “non-physical” and processual nature and the abstract conceptual level required for their exploration entail particular challenges for measurement.

4.2.1 Conceptual confusions around constructs

Although constructs constitute the primary objects of psychological and social-science research (Maraun et al. 2009), their definition and use are often ambiguous, inconsistent and afflicted with serious conceptual problems still largely ignored. A key fallacy is the common conflation of constructs as *theoretical-logical-linguistic tools* (e.g., abstractions, models, theoretical frameworks) with *their referents*, the concrete entities they are meant to denote (e.g., psychical processes, behaviours, income; Danziger 1997). This *construct-entity conflation* (Slaney and Garcia 2015) occurs, for example, when scientists interpret constructs as reflecting ‘attributes’ or ‘qualities’ that individuals ‘possess’ (e.g., in Cronbach and Meehl 1955), as widely done in ‘trait’ psychology (Uher 2013, 2015e). It contributes to the reification of constructs, ascribing to them an ontological status. The language used by construct developers further contributes to this reification and misleads scientists to overlook the constructed nature of constructs (Slaney and Garcia 2015), and thus also the necessity to clearly distinguish theoretical from operational construct definition.

4.2.2 By fiat definition of measurement models: operationalising constructs in concrete entities

For quantitative empirical investigations, constructs, given their conceptual nature, must be *operationally defined* (operationalised) in concrete entities that are accessible and thus (potentially) measurable (i.e., some of their referents). This corresponds to metrologists’ specification of a measurement model comprising the demarcated entities to be measured (measurands) in the study phenomena’s properties together with assumptions on their interrelations to produce the measurement results. For constructs, scientists use either single concrete entities, called *proxies* (e.g., annual income as single measure of ‘socio-economic status’; citation scores as single measure of ‘research impact’), or multiple concrete entities, called *indicators* (or *items* in language-based methods), from which composite measures are derived (Oakes and Rossi 2003). However, given the complexity and abstract nature of constructs, no proxy and no set of indicators can be all-inclusive, because, as conceptual entities, constructs imply more meaning (*surplus meaning*) than the concrete entities by which they can be operationally defined. No set of indicators could ever fully account for the abstract and complex phenomena construed as ‘social status’, ‘intelligence’ or ‘populism’. Construct developers must therefore decide about which particular indicators to include and which of their interrelations to consider and in what ways. This operational

definition of constructs by decree is called *measurement by fiat*¹⁴ in the social sciences (Cicourel 1964; Torgerson 1958).

To select indicators for construct operationalisation, two different approaches are used, psychotechnical and psychometric engineering. In *psychotechnical engineering* (Vautier et al. 2012), scientists define for the constructs under study a theoretical framework from which they derive an empirical framework specifying the measurable indicators used for operationalisation as well as specified linkages within and between these two frameworks (Cronbach and Meehl 1955; Messick 1995). An example is ‘socio-economic status’, for which social scientists specify various parameters of education, income, wealth and occupation that they consider construct relevant. Similarly, psychologists specify particular intellectual performances that they consider indicative of a given ‘intelligence’ construct.

However, even if correspondences between theoretical definition and empirical results are established through explicitly defined and interlinked frameworks, it cannot be established whether this allows to measure ‘socio-economic status’ or ‘intelligence’ as the actual research objects in themselves because these are conceptual entities that can be construed very differently (e.g., by considering different referents, aspects and levels of abstraction). This impossibility is reflected in social scientists’ conceptual debates about the nature of social stratification and the diversity of ‘socioeconomic status’ definitions and measures developed (Oakes and Rossi 2003). It is also reflected in ‘intelligence’ researchers’ intense debates, led for more than a century already, about which specific abilities (e.g., cognitive, emotional, social, creative) form part of ‘intelligence’ and which ones not (Spearman 1904; Sternberg 2018).

Given these challenges, it is unsurprising that some scientists follow operationalist epistemologies, according to which theoretical concepts could be defined by uniquely specified measurement operations (Chang 2009), such as when defining ‘intelligence’ as that “what an IQ-test measures” (Boring 1923; van der Maas et al. 2014). The idea that operational definition could substitute theoretical definition, and thus define the construct, underlies *psychometric engineering*, which is widely-used in psychology and social sciences (e.g., as alternative approach for defining ‘socio-economic status’; Oakes and Rossi 2003). In psychometrics, construct definitions and their theoretical structure are derived from empirical interrelations among indicators, which have often been selected in ways unrelated to the theoretical constructs established from them (Thissen 2001; Vautier et al. 2012). For example, popular ‘personality’ constructs (e.g., Big Five) were derived from empirical associations among judgements on person-descriptive words, which had been filtered from the English lexicon using an approach unrelated to any ‘personality’ theory (Uher 2015d).

4.2.3 Psychometric instruments and ‘measurement’ theories

In psychometrics, a first aim is to identify suitable sets of indicators (e.g., cognitive tasks, survey items) using empirical structures in the results that can be generated with them. For example, substantial interrelations among results (internal consistency) may suggest that the indicators capture phenomena and properties that can be conceived as forming a coherent entity as construed in a given construct. However, high internal consistency also implies considerable redundancies among the indicators. Redundancies can be

¹⁴ Derived from the Latin *fiat* for let it be done.

created easily through hypostatic abstraction (e.g., by emphasising aspects differently) and are therefore widespread in conceptual thinking and human language (Lahou 1996). But in natural systems (e.g., biological systems), ecological and evolutionary pressures may constrain the occurrence of redundancies. Indeed, associations among functionally-related behaviours are often only moderate or absent such as occurrences of non-contact and contact aggression (Uher et al. 2013a; Uher 2015b). This may explain why redundancy-based methods of data analysis (e.g., factor analysis) are hardly used outside of psychology and social sciences (Uher 2015d; Trofimova et al. 2018).

A second aim of psychometrics is to derive composite measures for constructs from the results obtained for their indicators. For this reason, sets of indicators are commonly called psychometric ‘*measuring*’ instruments (e.g., questionnaires, ‘intelligence’ tests). Various psychometric ‘measurement’ theories were developed for this purpose; most important are classical test theory and probabilistic latent trait theory (e.g., item response theory, Rasch modelling). They allow to define, on the basis of statistical assumptions, psychometric ‘measurement’ models in which the construct, because it is a conceptual entity and thus non-observable in itself, is encoded as a *latent variable*, and the indicators, because they are concrete observable entities, as *manifest variables*. Commonly, the quantity values created for these variables are labelled scores because the term value denotes a quality that renders something desirable or *valuable* (a cross-scientific jingle-fallacy).

These psychometric models build on the assumption that, irrespective of the methods used, invariant quantities exist for constructs (e.g., person ability), therefore called true scores or latent trait scores. Given these naïve realist assumptions, psychometricians aim to develop ideal methods (e.g., purposefully designed rating scales or ‘intelligence’ tests) that allow to empirically implement identity functions that turn these pre-existing true or latent scores into estimated scores that can be derived from the manifest indicator scores (with defined errors or probabilities, respectively; Mari et al. 2017; Uher 2018a). From statistical assumptions and assumptions about particular influencing factors (e.g., item difficulty, guessing, inattentiveness), psychometricians model what manifest scores can, theoretically, be obtained for an indicator given particular true or latent scores on the construct level (e.g., probabilistic variation around the hypothetical construct score). Such test-theoretical models are then used to infer (estimate) from the test persons’ empirically obtained manifest indicator scores their latent or true scores for the underlying (latent) construct (e.g., ‘ability’ score).

This widespread equation of constructs with latent variables, however, is based on the erroneous equation of constructs and with the phenomena they are meant to represent (construct-referent conflation; Maraun and Gabriel 2013), which places the ontology of constructs into their operational definition. This fallacy, which occurs in psychometric and psychotechnical engineering alike (e.g., in Cronbach and Meehl 1955), blurs the nature of the relations between theoretically constructed concepts (constructs) and the phenomena they are intended to represent (their referents; Slaney and Garcia 2015). It may also have contributed to the misconception of construct operationalisation as constituting a step of measurement.

4.3 Construct operationalisation is not measurement

In fiat ‘measurement’, scientists assume, given a particular theory (e.g., statistical or content-related), face validity, common-sense or intuition, that particular concrete entities are representative referents of a given construct. But these assumptions cannot be proven. The

links between construct and indicators can be established only by decree because constructs are socio-culturally constructed and may therefore vary substantially across (also scientific) communities. For example, meaning, composition and operationalisation of ‘socio-economic status’ constructs vary substantially across countries due to geographic, cultural and socio-economic differences (Psaki et al. 2014). For object-dependent investigations of the study phenomena’s *qualities*, consideration of these variations is essential, but they hinder object-dependent investigation of the *quantities* that may occur in these qualities, as required for measurement.

It follows that construct operationalisation cannot constitute a step of measurement¹⁵ as implied by the term fiat ‘measurement’. Instead, it reflects decisions and assumptions made by scientists about which sets of indicators may meaningfully reflect the abstract idea represented by a construct, which can and should be made explicit and inter-subjectively. Selecting indicators to explore how individuals apprehend the contents of their experience of ‘reality’ requires interpretive analysis; this cannot be accomplished experimentally. These interpretive decisions are essential because they concern the *qualitative* properties of the study phenomena to which a given construct is meant to refer. But crucially, these by-fiat decisions do not establish documented and unbroken connection chains from hypothetical measurands in the constructs—i.e., *quantitative* (divisible) properties of the qualities under study—to possible *quantitative* properties in the phenomena used as indicators, as required for measurement (object-dependence). This precludes the establishment of data generation traceability, a methodological key principle of measurement (Fig. 4).

The necessarily decision-based linkage of constructs with their indicators differs fundamentally from the measurement of physical properties, for which neither the parameters nor their interrelations can be defined by decree. Instead, they have to be developed and defined on the basis of existing physical properties and their structural (lawful) connections that scientists must identify experimentally.

But importantly, this is a consequence of these phenomena’s properties, not of the scientists’ concepts. Indeed, when applying their structural framework of measurement (Sect. 3.2) to social-science phenomena, metrologists specified in mathematical equations various relations that equated the construct ‘research performance’ with ‘quality of research products’ and the latter with ‘research impact’, which they then operationalised with citation scores and declared these relations’ “existence is assumed and not further discussed” (Mari et al. 2017). This clearly reflects decisions by decree (fiat) and a psychotechnical approach. The expression of assumed relations in mathematical equations is foundational also for test-theoretical approaches in psychology and social sciences; this is not specific to metrology. But the metrologists failed to recognise the conceptual nature of the abstract ideas like ‘research performance’, ‘research quality’ and ‘innovation’ that they discussed and the inherently interpretive nature of the decreed relations between these constructs and their operationalisations.

¹⁵ Construct operationalisation also differs fundamentally from fuzzy measurement, which specifies ways for representing epistemological plausibilities or probabilities for particular beliefs or events (Billingsley 1995). Thus, fuzzy measurement is concerned not with operationally defining concepts but instead with generating scores, which constitutes a subsequent step in measurement processes.

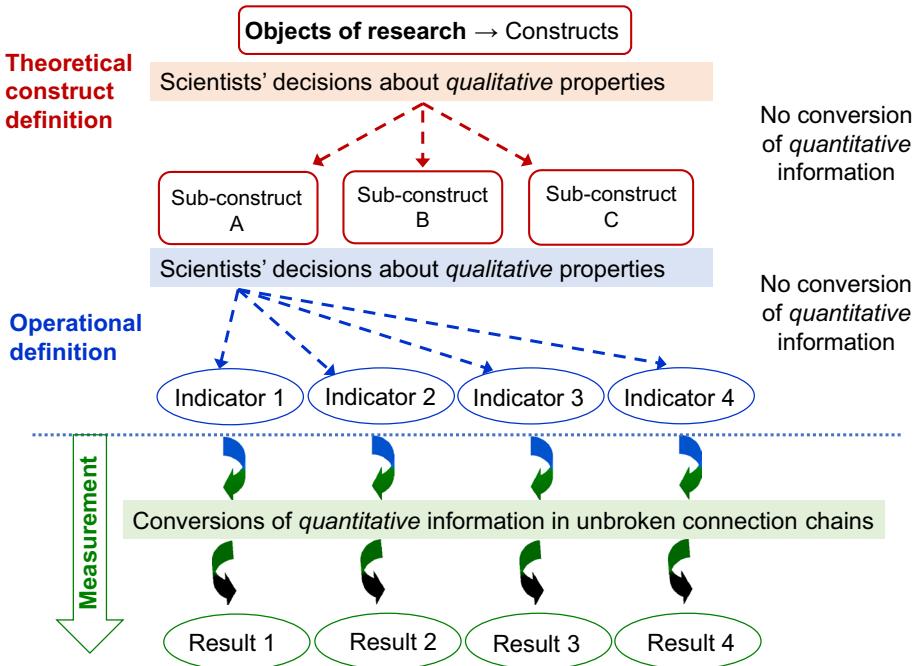


Fig. 4 Construct operationalisation is not measurement. Note: Construct operationalisation involves decisions on *qualitative* properties but not conversions of *quantitative* information as required for measurement

5 Conclusions

5.1 Measurement processes can be established only for construct indicators but not for constructs in themselves

These methodological analyses highlight that constructs in themselves cannot be measured. The frequent notion of ‘construct measurement’ and ‘measuring instruments’ for constructs (e.g., questionnaires) is oversimplifying and misleading, and likely a result of the widespread construct-referent conflation. But measurement processes can be established for many of the concrete entities chosen as construct indicators (referents). Hence, object-dependence as a meta-condition of measurement cannot refer to constructs as the actual research objects but only to the indicators used for their operationalisation—precisely because these are accessible and thus (potentially) measurable (Fig. 4). This differentiation is crucial because indicators are neither the construct in itself nor specific quantities of it. Constructs and their indicators constitute different entities (conceptual vs concrete; e.g., ‘socio-economic status’ vs income).

Results obtained for construct indicators (e.g., performances in ‘intelligence’ test) can be used to draw inferences about the abstract entity construed (e.g., ‘intelligence’). But because the links connecting constructs with their indicators are established not on the basis of lawful empirical connections persisting across time and contexts (as in measurement of physical properties) but by interpretive decisions (by fiat), one-to-one correspondences cannot exist and inferences to the construct level are, necessarily, interpretive as well

and can never be proven. This explains why different operationalisations lead to different results for the same construct and individual (e.g., IQ-scores typically vary across ‘intelligence’ tests). Clear awareness of the inherently interpretive nature of inferences from indicator results to constructs has particular relevance for legal contexts, where the validity of psychometric scores as legal evidence (e.g., persons’ ‘intelligence’) is increasingly being questioned and may soon be challenged in courts (Barrett 2018) similar to psychiatric diagnostic practices before (Faust 2012).

The two basic methodological principles of measurement (data generation traceability and numerical traceability) can, however, be implemented for many construct indicators, though not for all. In behavioural observations¹⁶ (ethological real-time or video-based coding, not ratings, see below), multiple observers can reach inter-subjective agreement in the demarcation, categorisation and encoding of behavioural acts because their extroquestive accessibility enables direct and joint perception of the same occurrences. This allows to establish documented and unbroken links between the entities to be measured and their encoding in data, thus transparency and traceability in data generation. Behaviours’ pronounced variability, transient nature and context-dependent meaning entails, however, that *a*) the demarcation and categorisation of behavioural acts necessarily involves some *defined* scope for interpretation (e.g., what acts are of the same kind), and that *b*) observers, using nothing but their human abilities¹⁶, can hardly ever quantify particular properties directly (e.g., duration of talking). For these reasons, observers commonly encode only occurrences/non-occurrences of defined behavioural acts in nominal data (1/0) from which post hoc—after data generation is completed—or using behavioural coding software, ratio-scaled quantitative data can be derived (e.g., durations, frequencies; for details (Uher 2013, 2015d, 2018a). Measurement of temporal properties is well-established; defined occurrences are discrete quantities (multitudes), which are countable. The meaning of the quantifications thus-derived is therefore documented, transparent and subject-independent, thus establishing numerical traceability as a social-science analogue of metrological traceability.

Person-based measurement of behavioural phenomena, however, is largely confined to their physical properties (e.g., temporal, spatial). Attempts to establish object-dependent and thus traceable processes to generate quantitative data about interpretive and meaning aspects of behaviours (e.g., ‘dominance’ or ‘persuasiveness’ of talking) face limitations.

5.2 Pitfalls of language-based methods used for quantitative data generation: rating items refer to concepts

Psychological and social-science investigations often involve language-based methods for quantitative data generation (e.g., assessment scales) in which persons are asked to respond to standardised descriptions of construct indicators, called items (e.g., single words, short sentences). Words can refer to concrete, directly perceivable entities but also to entities abstracted from the here and now or even imperceptible in themselves (e.g., constructs). The level of abstraction, however, is not directly apparent from the words themselves. This may have obscured the fact that, in assessment methods, many items enquire about entities that are abstract and not even present during data generation (e.g., past events like habitual behaviours as in ‘personality’ assessments). This inevitably precludes the establishment of

¹⁶ “Observational methods involve the targeted, and thus trained, perception of the study phenomena without any mechanism standing between the observer and the observed” (Uher 2019, p. 230).

object-dependent and traceable data generation processes. Consequently, rating methods cannot capture information about such entities *in themselves* but only about persons' pertinent ideas and beliefs, thus their concepts.

As abstractions and generalisations, concepts refer to various entities (referents). Indeed, rating items are often purposefully worded in abstract and decontextualised ways to make them applicable to diverse phenomena and contexts without specifying any particular ones. This requires respondents to use their common-sense knowledge to first interpret the content described and to construct specific meanings for the given context. It is therefore unsurprising that interpretations of psychometrically selected, standardised rating items vary substantially within and between persons, indicating broad fields of meaning and substantial subjectivity in data generation (Lundmann and Villadsen 2016; Rosenbaum and Valsiner 2011; Uher and Visalberghi 2016; de Williams et al. 2000). But unlike in observational methods, scientists commonly neither instruct nor train the data-generating persons to interpret rating items in standardised ways nor do they enquire about raters' item interpretations and the particular referents that raters considered when judging a particular case. This introduces a twofold break in data generation traceability (Uher 2018a).

To create numerical data, raters are asked to indicate their judgements in predefined multi-stage answer categories commonly labelled lexically (e.g., agree, strongly agree, etc.). Researchers then assign to these answer categories numerical values in always the same and thus perfectly traceable ways. But this numerical assignment is only a recoding of data. The actual data generation is accomplished by the raters. Despite their pivotal role in data generation, the ways in which respondents choose their answer categories are still largely unexplored. First studies showed that, when choosing their answer categories on agreement scales, about 90% of 78 respondents considered not quantitative properties as commonly assumed but only qualitative properties instead (Figure 13 in Uher 2018a). But scientists rigidly assign always the same numerical values to the same answer categories regardless of raters' category interpretations and regardless of the item content and thus the different qualities to which they refer. Therefore, the meaning of the numerical values assigned by researchers can be traced back neither to the particular measurands raters may have had in mind nor to some known standards that could create for these values an intersubjective meaning with regard to the study properties (e.g., how often must a particular behaviour occur to be judged as 'often' given that different behaviours generally occur with different frequencies within and across situations; Uher 2015b). This precludes establishment of numerical traceability (for details; Uher 2018a).

By contrast, coding performances in educational and 'intelligence' tests, for which correct and incorrect responses exist, creates subject-independent and traceable meanings for the numerals assigned to test answers. These can be documented and be made transparent and publicly accessible (e.g., for enabling international comparison of test performances in PISA-studies), thereby establishing numerical traceability.

5.3 Composite scores derived from indicator measurements reflect artificial quantifications

From indicator results, psychologists and social scientists aim to derive overall scores for their constructs. But quantifications obtained for different indicators often refer to different properties and therefore cannot be simply summarised (e.g., income in monetary currency, education in years). Therefore, composite scores—whether using explicit rules derived from theoretical construct definitions in psychotechnical engineering or using

statistically-derived test-theoretical models in psychometrical engineering—constitute *artificial quantifications*. This is because, in constructs, scientists aim to summarise entities with *heterogenous* qualities, which precludes the possibility to identify in them divisible properties of the *same* quality, thus quantities. The artificial quantifications created for constructs may still be useful for pragmatic purposes, especially when they are derived from measurement-based and thus traceable indicator results (e.g., composite scores of ‘socio-economic status’). This allows establishing comparability—but only within the limits of the inherently *decision-based* selection of indicators and of algorithms used to merge their results (e.g., different weighting).

5.4 Two basic methodological principles of measurement applicable across sciences

The transdisciplinary analyses identified two basic methodological principles that underlie metrologists’ structural frameworks and are crucial for measurement and that can be meaningfully adapted to psychologists’ and social scientists’ study phenomena, carefully considering their peculiarities. (1) *Data generation traceability* requires that assignments of numerical values solely depend on the properties explored in the study objects (object-dependence) and are made fully transparent, and thus reproducible and traceable. To achieve this, scientists must establish unbroken documented connection chains that directly link (via different steps if needed) the quantitative entity to be measured in the qualitative study property (measurand) with the numerical value assigned to it, thus ensuring equivalence between them. (2) *Numerical traceability* requires that scientists directly link the assigned numerical values also to known standards, likewise in documented and transparent ways, thereby establishing the results’ public interpretability (subject-independence).

These two methodological principles highlight important commonalities in the ways in which measurement-based quantifications can be generated across all sciences. They specify the foundational concepts of measurement that are required to ensure the quality of the quantitative information obtained and to justify the public trust placed in them. They are also needed to distinguish measurement-based quantifications from other (e.g., subjective) quantifications that may be useful for pragmatic purposes but lack epistemic authority. This is of particular importance for applied (e.g., legal, clinical, educational) contexts in which quantifications are used to make decisions about individuals. These two principles also open up new perspectives on the replication crises widely-discussed in various sciences and provide new concepts for the kind of transparency needed in scientific investigations for overcoming them.

5.5 Reconsider if quantifications are meaningful at all to explore given phenomena

But the analyses also highlighted some fundamental differences and limitations. Crucially, possibilities for implementing measurement processes are not a matter of scientific discipline or their ascribed level of scientificity but solely depend on the study phenomena’s properties. To explore individuals (inter-)subjective understanding and interpretation of the contents of their experience that mediate their concrete experience of ‘reality’, scientists must investigate the qualities, interrelations and development of meanings. These study phenomena are highly complex, context-dependent and changeable. Consequently, psychological and social-science concepts are not applicable uniformly across time and space but constantly changing as well. Rather than a deficiency, this reflects the precision and

object-dependence with which these scientists explore the manifold ever-changing *qualities* that constitute the key features of their study phenomena.

Quantifications are meaningful only if the basic qualities to be studied for their possible divisible properties remain rather constant. But quantifications have little value for description and explanation if the qualities *in themselves* are undergoing permanent change and development. Psychologists and social scientists can learn from metrologists' advancements in measuring physical properties when it comes to obtaining precise *quantitative* information about *constant* properties. But metrological approaches are inadequate for exploring the every-changing processes of meaning making and interpretation. Interpretive approaches cannot be replaced by mathematical formalisations and algorithms. Metrology does not provide any pertinent concepts; this is the expertise of psychological and social scientist—and this is why scholars must collaborate across the sciences to tackle the challenges of the twenty first century.

Acknowledgements The author thanks the editor and four reviewers for their constructive comments on earlier drafts. This research was funded by a Marie Curie Fellowship of the European Commission's FP7 Programme awarded to the author (EC Grant Agreement No. 629430).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barrett, P.: The EFPA test-review model: when good intentions meet a methodological thought disorder. *Behav. Sci.* **8**(1), 5 (2018). <https://doi.org/10.3390/bs8010005>
- Berglund, B.: Measurement in psychology. In: Berglund, B., Rossi, G.B., Townsend, J.T., Pendrill, L. (eds.) *Measurement with Persons : Theory, Methods, And Implementation Areas*, pp. 27–50. Taylor Francis, New York (2012)
- Berglund, B., Rossi, G.B., Townsend, J.T., Pendrill, L. (eds.): *Measurement with Persons: Theory, Methods, and Implementation Areas*. Taylor Francis, New York (2012)
- Billingsley, P.: Probability and Measure, 3rd edn. Wiley, New York (1995)
- Bohr, N.: Causality and complementarity. *Philos. Sci.* **4**(3), 289–298 (1937)
- Boring, E.G.: Intelligence as the tests test it. *New Repub.* **36**, 35–37 (1923)
- Brody, N., Oppenheim, P.: Application of Bohr's principle of complementarity to the mind-body problem. *J. Philos.* **66**(4), 97–113 (1969). <https://doi.org/10.2307/2024529>
- Chang, H.: Operationalism. In: Zalta, E.N. (eds.) *The Stanford Encyclopedia of Philosophy* (Fall 2009 ed), (2009). <http://plato.stanford.edu/archives/fall2009/entries/operationalism>; Accessed 23 Oct 2014
- Cicourel, A.: Method and Measurement in Sociology. The Free Press of Glencoe, New York (1964)
- Cronbach, L.J., Meehl, P.E.: Construct validity in psychological tests. *Psychol. Bull.* **52**(4), 281–302 (1955). <https://doi.org/10.1037/h0040957>
- Czichos, H.: Introduction to metrology and testing. In: Czichos, H., Saito, T., Smith, L. (eds.) *Springer Handbook of Metrology and Testing*, pp. 3–22. Springer, Berlin (2011). https://doi.org/10.1007/978-3-642-16641-9_1
- Danziger, K.: Naming the Mind: How Psychology Found Its Language. Sage, London (1997)
- De Silva, G.M.S.: Basic Metrology for ISO 9000 Certification. Butterworth-Heinemann, Oxford (2002)
- Descartes, R., Miller, V.R., Miller, R.P.: Principles of Philosophy (Reprint). Reidel, Dordrecht (1983)
- Fahrenberg, J.: The complementarity principle in psychophysiological research and somatic medicine. *Zeitschrift Für Klinische Psychologie Und Psychotherapie* **27**(2), 151–167 (1979)

- Fahrenberg, J.: Zur Kategorienlehre der Psychologie: Komplementaritätsprinzip; Perspektiven und Perspektiven-Wechsel. Pabst Science Publishers, Lengerich (2013)
- Faust, D.: Ziskin's Coping with Psychiatric and Psychological Testimony. Oxford University Press, Oxford (2012). <https://doi.org/10.1093/med:psych/9780195174113.001.0001>
- Fechner, G.T.: Elemente der Psychophysik. Breitkopf und Härtel, Leipzig (1860)
- Finkelstein, L.: Widely, strongly and weakly defined measurement. *Measurement* **34**(1), 39–48 (2003)
- Fisher, W.P., Wilson, M.: An online platform for sociocognitive metrology: the BEAR assessment system software. *Meas. Sci. Technol.* **31**(3), 034006 (2020). <https://doi.org/10.1088/1361-6501/ab5397>
- Floridi, L.: Semantic conceptions of information. In: Zalta, E.N. (ed.) *The stanford encyclopedia of philosophy* (Winter 2019 Edition) (2019). Retrieved from <https://plato.stanford.edu/archives/win2019/entries/information-semantic>. Accessed 22 Dec 2019
- Hand, D.J.: Measurement: A Very Short Introduction. Oxford University Press, Oxford (2016)
- Hartmann, N.: Der Aufbau der realen Welt. Grundriss der allgemeinen Kategorienlehre, 3rd edn. Walter de Gruyter, Berlin (1964)
- Heisenberg, W.: Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift Für Physik* **43**(3–4), 172–198 (1927). <https://doi.org/10.1007/BF01397280>
- Hossenfelder, S.: Lost in Math: How Beauty Leads Physics Astray. Basic Books, New York (2018)
- JCGM100:2008. Evaluation of measurement data—Guide to the expression of uncertainty in measurement (GUM). Joint Committee for Guides in Metrology (originally published in 1993), (2008). <http://www.bipm.org/en/publications/guides/gum.html>. Accessed 10 June 2018
- JCGM200:2012.: International Vocabulary of Metrology—Basic and General Concepts and Associated Terms (VIM 3rd edn.). Working Group 2 (eds.), Joint Committee for Guides in Metrology, (2012). https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf. Accessed 5 Feb 2017
- Kant, I.: Kritik der reinen Vernunft (J. Timmermann, ed.). Felix Meiner Verlag, Hamburg (1998)
- Kelley, T.L.: Interpretation of Educational Measurements. World, Yonkers (1927)
- Khanam, S.A., Liu, F., Chen, Y.-P.P.: Comprehensive structured knowledge base system construction with natural language presentation. *Hum. Centric Comput. Inf. Sci.* **9**(1), 23 (2019). <https://doi.org/10.1186/s13673-019-0184-7>
- Kothari, C.R.: Research Methodology: Methods and Techniques, 2nd edn. New Age International Publishers, New Delhi (2004)
- Lahlou, S.: The propagation of social representations. *J. Theory Soc. Behav.* **26**(2), 157–175 (1996). <https://doi.org/10.1111/j.1468-5914.1996.tb00527.x>
- Lewin, K.: Principles of Topological Psychology. McGraw-Hill, New York (1936)
- Locke, J.: An Essay Concerning Human Understanding. Electronic Classics Series. The Pennsylvania State University, Hazleton (1999)
- Lundmann, L., Villadsen, J.W.: Qualitative variations in personality inventories: subjective understandings of items in a personality inventory. *Qual. Res. Psychol.* **13**(2), 166–187 (2016). <https://doi.org/10.1080/14780887.2015.1134737>
- Maraun, M.D., Gabriel, S.M.: Illegitimate concept equating in the partial fusion of construct validation theory and latent variable modeling. *New Ideas Psychol.* **31**(1), 32–42 (2013). <https://doi.org/10.1016/J.NEWideAPSYCH.2011.02.006>
- Maraun, M.D., Slaney, K.L., Gabriel, S.M.: The Augustinian methodological family of psychology. *New Ideas Psychol.* **27**(2), 148–162 (2009). <https://doi.org/10.1016/J.NEWideAPSYCH.2008.04.011>
- Mari, L., Carbone, P., Giordani, A., Petri, D.: A structural interpretation of measurement and some related epistemological issues. *Stud. Hist. Philos. Sci.* **65–66**, 46–56 (2017). <https://doi.org/10.1016/j.shpsa.2017.08.001>
- Mari, L., Carbone, P., Petri, D.: Fundamentals of hard and soft measurement. In: Ferrero, A., Petri, D., Carbone, P., Catelani, M. (eds.) *Modern Measurements: Fundamentals and applications*, pp. 203–262. Wiley, Hoboken (2015). <https://doi.org/10.1002/9781119021315.ch7>
- Mari, L., Wilson, M.: A structural framework across strongly and weakly defined measurements. In: 2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, pp. 1522–1526, (2015). <https://doi.org/10.1109/I2MTC.2015.7151504>
- Maul, A., Mari, L., Torres Irribarria, D., Wilson, M.: The quality of measurement results in terms of the structural features of the measurement process. *Measurement* **116**, 611–620 (2018). <https://doi.org/10.1016/J.MEASUREMENT.2017.08.046>
- Messick, S.: Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am. Psychol.* **50**(9), 741–749 (1995). <https://doi.org/10.1037/0003-066X.50.9.741>
- Michell, J.: Is psychometrics pathological science? *Meas. Interdiscip. Res. Perspect.* **6**(1–2), 7–24 (2008). <https://doi.org/10.1080/15366360802035489>

- Oakes, J.M., Rossi, P.H.: The measurement of SES in health research: current practice and steps toward a new approach. *Soc. Sci. Med.* **56**(4), 769–784 (2003). [https://doi.org/10.1016/S0277-9536\(02\)00073-4](https://doi.org/10.1016/S0277-9536(02)00073-4)
- Ogden, C.K., Richards, I.A.: *The Meaning of Meaning: A Study of The Influence of Language Upon Thought And of The Science of Symbolism*. Harcourt, Brace & World, San Diego (1923)
- Peirce, C.S.: Collected papers of Charles Sanders Peirce. vol. 1–6, Hartshorn, C., Weiss, P. (eds.), vol. 7–8, Burks, A.W. (ed.). Harvard University Press, Cambridge (1958)
- Pendrill, L.: Man as a measurement instrument. *NCSLI Meas.* **9**(4), 24–35 (2014). <https://doi.org/10.1080/19315775.2014.1172170>
- Psaki, S.R., Seidman, J.C., Miller, M., Gottlieb, M., Bhutta, Z.A., Ahmed, T., Ahmed, A.S., Bessong, P., John, S.M., Kang, G., Kosek, M., Checkley, W.: Measuring socioeconomic status in multicountry studies: results from the eight-country MAL-ED study. *Popul. Health Metr.* **12**(1), 8 (2014). <https://doi.org/10.1186/1478-7954-12-8>
- Quinn, T.J.: From Artefacts to Atoms: The BIPM and the Search for Ultimate Measurement Standards. Oxford University Press, Oxford (2010)
- Rosenbaum, P.J., Valsiner, J.: The un-making of a method: from rating scales to the study of psychological processes. *Theory Psychol.* **21**(1), 47–65 (2011). <https://doi.org/10.1177/0959354309352913>
- Slaney, K.L.: Validating psychological constructs: Historical, philosophical, and practical dimensions. Palgrave Macmillan, London, UK (2017). <https://doi.org/10.1057/978-1-37-38523-9>
- Schacter, D.L., Addis, D.R.: Constructive memory: the ghosts of past and future. *Nature* **445**(7123), 27 (2007). <https://doi.org/10.1038/445027a>
- Slaney, K.L., Garcia, D.A.: Constructing psychological objects: the rhetoric of constructs. *J. Theor. Philos. Psychol.* **35**(4), 244–259 (2015). <https://doi.org/10.1037/teo0000025>
- Spearman, C.: General intelligence, objectively determined and measured. *Am. J. Psychol.* **15**, 201–293 (1904)
- Sternberg, R.J.: Intelligence in humans. In: Reference Module in Neuroscience and Biobehavioral Psychology. Elsevier, Amsterdam (2018). <https://doi.org/10.1016/B978-0-12-809324-5.21773-2>
- Stevens, S.S.: On the theory of scales of measurement. *Science* **103**, 667–680 (1946)
- Thissen, D.: Psychometric engineering as art. *Psychometrika* **66**(4), 473–486 (2001)
- Thorndike, E.L.: Notes on Child Study, 2nd edn. Macmillan, New York (1903)
- Thurstone, L.L.: Attitudes can be measured. *Am. J. Sociol.* **33**, 529–554 (1928)
- Tobi, H.: Measurement in interdisciplinary research: the contributions of widely-defined measurement and portfolio representations. *Measurement* **48**, 228–231 (2014). <https://doi.org/10.1016/J.MEASUREMENT.2013.11.013>
- Torgerson, W.S.: Theory and Methods of Scaling. Wiley, New York (1958)
- Trofimova, I., Robbins, T.W., Sulis, W.H., Uher, J.: Taxonomies of psychological individual differences: Biological perspectives on millennia-long challenges. *Philos. Trans. R. Soc. B Biol. Sci.* **373**(1744), 20170152 (2018). <https://doi.org/10.1098/rstb.2017.0152>
- Uher, J.: Personality psychology: Lexical approaches, assessment methods, and trait concepts reveal only half of the story—Why it is time for a paradigm shift. *Integ. Psychol. Behav. Sci.* **47**(1), 1–55 (2013). <https://doi.org/10.1007/s12124-013-9230-6>
- Uher, J.: Agency enabled by the psyche: explorations using the Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals. In: Gruber, C.W., Clark, M.G., Klempe, S.H., Valsiner, J. (eds.) Constraints of agency: explorations of theory in everyday life. *Annals of Theoretical Psychology*, vol. 12, pp. 177–228. Springer, Cham (2015a). https://doi.org/10.1007/978-3-319-10130-9_13
- Uher, J.: Comparing individuals within and across situations, groups and species: metatheoretical and methodological foundations demonstrated in primate behaviour. In: Emmans, D., Laihinens, A. (eds.) Comparative neuropsychology and brain imaging. Series Neuropsychology: An Interdisciplinary Approach, vol. 2, pp. 223–284. Lit, Berlin (2015b). <https://doi.org/10.13140/RG.2.1.3848.8169>
- Uher, J.: Conceiving “personality”: Psychologist’s challenges and basic fundamentals of the Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals. *Integr. Psychol. Behav. Sci.* **49**(3), 398–458 (2015c). <https://doi.org/10.1007/s12124-014-9283-1>
- Uher, J.: Developing “personality” taxonomies: Metatheoretical and methodological rationales underlying selection approaches, methods of data generation and reduction principles. *Integr. Psychol. Behav. Sci.* **49**(4), 531–589 (2015d). <https://doi.org/10.1007/s12124-014-9280-4>
- Uher, J.: Interpreting “personality” taxonomies: why previous models cannot capture individual-specific experiencing, behaviour, functioning and development. Major taxonomic tasks still lay ahead. *Integr. Psychol. Behav. Sci.* **49**(4), 600–655 (2015e). <https://doi.org/10.1007/s12124-014-9281-3>
- Uher, J.: Exploring the workings of the psyche: metatheoretical and methodological foundations. In: Valsiner, J., Marsico, G., Chaudhary, N., Sato, T., Dazzani, V. (eds.) *Psychology as the science of human*

- being: the Yokohama Manifesto, pp. 299–324. Springer, Cham (2016a). https://doi.org/10.1007/978-3-319-21094-0_18
- Uher, J.: What is behaviour? And (when) is language behaviour? A metatheoretical definition. *J. Theor. Soc. Behav.* **46**(4), 475–501 (2016b). <https://doi.org/10.1111/jtsb.12104>
- Uher, J.: Quantitative data from rating scales: an epistemological and methodological enquiry. *Front. Psychol.* **9**, 2599 (2018a). <https://doi.org/10.3389/fpsyg.2018.02599>
- Uher, J.: Taxonomic models of individual differences: a guide to transdisciplinary approaches. *Philos. Trans. R. Soc. B* **373**(1744), 20170171 (2018b). <https://doi.org/10.1098/rstb.2017.0171>
- Uher, J.: The transdisciplinary philosophy-of-science paradigm for research on individuals: foundations for the science of personality and individual differences. In: Zeigler-Hill, V., Shackelford, T.K. (eds.) *The SAGE handbook of personality and individual differences: Volume I: The science of personality and individual differences*, Chapt. 4, pp. 84–109, Sage, London, UK (2018c). <https://doi.org/10.4135/978126451163.n4>
- Uher, J.: Data generation methods across the empirical sciences: differences in the study phenomena's accessibility and the processes of data encoding. *Quality & Quantity*. *Int. J. Methodol.* **53**(1), 221–246 (2019). <https://doi.org/10.1007/s11135-018-0744-3>
- Uher, J., Visalberghi, E.: Observations versus assessments of personality: a five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *J. Res. Pers.* **61**, 61–79 (2016). <https://doi.org/10.1016/j.jrp.2016.02.003>
- Uher, J., Addessi, E., Visalberghi, E.: Contextualised behavioural measurements of personality differences obtained in behavioural tests and social observations in adult capuchin monkeys (*Cebus apella*). *J. Res. Pers.* **47**(4), 427–444 (2013a). <https://doi.org/10.1016/j.jrp.2013.01.013>
- Uher, J., Werner, C.S., Gosselt, K.: From observations of individual behaviour to social representations of personality: developmental pathways, attribution biases, and limitations of questionnaire methods. *J. Res. Pers.* **47**(5), 647–667 (2013b). <https://doi.org/10.1016/j.jrp.2013.03.006>
- Valsiner, J.: *A Guided Science: History of Psychology in the Mirror of Its Making*. Transaction Publishers, New Brunswick (2012)
- van der Maas, H., Kan, K.-J., Borsboom, D.: Intelligence is what the intelligence test measures. Seriously. *J. Intell.* **2**(1), 12–15 (2014). <https://doi.org/10.3390/intelligence2010012>
- Vautier, S., Veldhuis, M., Lacot, E., Matton, N.: The ambiguous utility of psychometrics for the interpretative foundation of socially relevant avatars. *Theory Psychol.* **22**(6), 810–822 (2012). <https://doi.org/10.1177/0959354312450093>
- Vygotsky, L.S.: *Thought and Language*. MIT Press, Cambridge (1962)
- Walach, H., Römer, H.: Complementarity is a useful concept for consciousness studies: a reminder. *Neuroendocrinol. Lett.* **21**, 221–232 (2011)
- Whitehead, A.N.: *Process and Reality*. Harper, New York (1929)
- de Williams, A.C.C., Davies, H.T.O., Chadury, Y.: Simple pain rating scales hide complex idiosyncratic meanings. *Pain* **85**(3), 457–463 (2000). [https://doi.org/10.1016/S0304-3959\(99\)00299-7](https://doi.org/10.1016/S0304-3959(99)00299-7)
- Wilson, M., Mari, L., Maul, A., Irribarria, D.T.: A comparison of measurement concepts across physical science and social science domains: instrument design, calibration, and measurement. *J. Phys Conf. Ser.* **588**(1), 012034 (2015). <https://doi.org/10.1088/1742-6596/588/1/012034>
- Wundt, W.: *Grundriss der Psychologie*. Springer, Körner (1896). <http://archive.org/>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.