# Psychological measurement: Theory, paradoxes, and prototypes

## Stephen M. Humphry
The University of Western Australia, Australia

## Abstract
The article focuses on the ongoing debate regarding the measurement of psychological attributes. The aim is to clarify different uses of the term *theory* and key points of agreement and disagreement among participants. In addition, the article addresses misinterpretations of key points in recent articles and notes an apparent paradox arising in the representational theory of measurement. Substantive theory is contrasted with both item response theory and the representational theory of measurement. Emphasis is placed on the direct dependence of the measurement of physical attributes on substantive quantitative theory as opposed to any form of separate measurement theory. It is concluded that the primary challenge faced in quantitative psychology is to posit testable substantive theories or laws which form a foundation for measurement.

There is an ongoing discussion and debate in *Theory & Psychology* and other outlets regarding whether it is possible to measure psychological attributes. Consistent with observations made by Sijtsma and Emons (2013), it seems there is a lack of commonality in the discussion in the usage of the term *theory* with respect to psychological measurement. Participants in the debate have focused on at least three forms of theory: (a) Item response theory (IRT), (b) the representational theory of measurement, and (c) substantive theory about psychological attributes and phenomena.

Previously it has been observed that approaches to measurement in the physical sciences are directly based on substantive physical theory. Specifically, design principles

**Corresponding author:**
Stephen M. Humphry, Graduate School of Education, The University of Western Australia, Crawley, WA 6019, Australia.
Email: stephen.humphry@uwa.edu.au

for physical measuring instruments are invariably based directly on substantive physical theory (Hebra, 2010; Humphry, 2013b). Consequently, successful empirical tests of substantive theory underpin the success of measurement and there is no separate measurement theory (Humphry, 2011, 2013a).

In contrast, Michell (2008) asserts:

> It is really only when tests [of cancellation conditions] are predicted on the basis of a psychological theory that gives explicit content to quantitative structure and confirmed by data that we will have compelling evidence that the relevant latent trait is quantitative. (p. 20)

The cancellation conditions derive from the representational theory of measurement and, therefore, Michell claims that a separate measurement theory is needed to have such compelling evidence. However, tests of this kind did not play a part in the progress of physics and hence representational theory played no direct part in the progress of measurement in the physical sciences. Clearly, though, quantities such as force, power, potential difference, and electrical current have been successfully measured very precisely and accurately using a wide range of instruments.

For this reason, I have criticized representational theory as being divorced from physics and psychology alike. Michell, however, misconstrued my motivation for criticizing representational theory, stating:

> [M]ore dismaying than this is the fact that both Humphry and Sijtsma dismiss the theory of conjoint measurement as a viable research tool. It is dismaying because this theory and IRT are not opposed and no one is forced to choose *either* one *or* the other. (2014, p. 115)

Contrary to this remark, I have never suggested that one is forced to choose either one or the other of the theories. As will be elaborated in this article, my position is that similar issues exist in relation to both forms of "theory."

Given the above considerations, it is important to clarify points of agreement and disagreement with Michell as well as to clarify the senses in which the term theory is used. The article is structured as follows. The term *theory* is first considered in relation to measurement in psychology and more generally. A point of agreement with Michell is clarified, with a focus on deficiencies of psychometrics. Next, it is argued that IRT and representational measurement theory have commonalities. In particular, both purport to provide a theoretical framework for measurement but without direct connection to substantive psychological theories and therefore without obvious avenues to measuring in well-defined measurement units. Both are also quasi-empirical in nature: IRT invokes a quasi-empirical unit and representational measurement theory invokes a quasi-empirical notion of quantitative structure. Questions are then posed about an apparent paradox that appears (or perhaps reappears) in representational theory pertaining to the claim that continuous attributes possess quantitative "structure." A dynamic prototype for measurement elucidates key considerations in measurement that pertain to the ontology of quantities. Lastly, the discussion turns to the kind of evidence that would be required to demonstrate the successful measurement of psychological attributes.

## Usage and meaning of theory

It seems that a key source of miscommunication in the ongoing debate about measurement in psychology is lack of clarity regarding the usage and meaning of the term *theory* in relation to measurement. I will take *theory* to mean a conceptual framework intended to describe and explain one or more empirical phenomena. Sijtsma and Emons (2013) state that "IRT models are prescriptive with respect to measurement but not necessarily descriptive" (p. 787). I wholeheartedly agree and think that the representational measurement theory is especially prescriptive.

Clearly, a conceptual framework intended to explain empirical phenomena does not attempt to *prescribe* anything pertaining to the nature of empirical data. Based on this consideration alone, it is not justifiable to claim that either the Rasch model or representational theory constitute *theories* according to common usage of the term. Neither the representational theory nor Rasch models were developed to explain substantive quantitative phenomena or relations in psychology or education. Proponents of representational measurement theory and Rasch models explicitly *prescribe* empirical conditions they claim are necessary to successfully measure an attribute. In representational theory, it is claimed that the axioms must be satisfied as shown by empirical tests. In the application of Rasch models, it is prescribed that empirical response data must fit the model.

Both Rasch (1960/1980) and proponents of representational theory (Krantz, Luce, Suppes, & Tversky, 1971) offered post hoc claims of compatibility with specific cases of measurement in physics. In lines of development other than Rasch's, IRT models are used to describe response patterns, but the models are still not based on substantive theory and do not purport to explain empirical phenomena.

In contrast, in the physical sciences it has proven possible to measure in well-defined units based on substantive theory that describes quantitative relations and explains physical phenomena. Physical equations express physical theory in abbreviated form (Humphry, 2013a). The original extended forms of physical laws, such as by Newton, stated causal relations among ratios of different kinds of physical quantities, such as between forces, masses, and accelerations. Physical measuring instruments are designed in a manner that is intentionally and directly based on substantive physical theory (Hebra, 2010).

The same cannot be said of IRT or representational theory. As I have stated of IRT: "it seems apt to state that item response models formally emulate physical equations that contain implicit measurement units" (Humphry, 2013a, p. 778). The formal emulation does not have the same basis in empirically testable substantive theory that explains psychological phenomena.

Having said all of this, there seems to be an underlying point of agreement in the ongoing debate: most, if not all, participants agree that if it is to be successful, measurement in Psychology would ultimately need to be based on substantive theory. Sijtsma (2012), Sijtsma and Emons (2013), and Borsboom and Mellenbergh (2004) regard substantive theory as being necessary and/or concerned at least partly with justifications for applying particular models. Similarly, Michell (1999) clearly considers workable substantive theory essential to successful measurement. Although there are potential substantive justifications for IRT models, I have come to the conclusion that without a clear

connection to substantive theory, IRT models have at best quasi-units. It was interesting to find that Sijtsma and Emons (2013, p. 793) concur with this conclusion.

Clearly, it is often useful to order people in terms of attitudes, cognitive performances, and so forth. However, the caveat that there is only a quasi-unit is virtually never given, much less the implications for the application of arithmetic, and mathematics in general, made clear.

## Points of agreement and disagreement with Michell

### *Pathological science*

I agree with Michell that psychometrics has so far failed to convincingly demonstrate that any psychological attribute has been measured and that pathological science is occurring. I have particularly stressed that psychometrics has so far failed to demonstrate that a psychological attribute has been measured in a well-defined *unit of measurement* (Humphry, 2011). If it has not been shown that an attribute can be reproducibly measured in a well-defined unit, it has not been shown that the attributes have been measured on a scale.

In his recent article, Michell (2014) states:

> As I have noted, many psychometricians typically do not seem to be interested in investigating whether the attributes they aspire to measure are really quantitative … Instead they are primarily interested in already claiming that they can measure such attributes. (p. 117)

It is entirely accurate to say that many psychometricians do not seem to be interested in testing whether attributes are really quantitative. There is little indication that most who employ IRT models have considered the fundamental justification for parameters that are purported to be real numbers associated with levels of psychological attributes.

In contrast, the original use of mathematics in science using real numbers evolved from the development of substantive physical theory. This usage took the form of physico-mathematics which served to express, in an abbreviated form, quantitative relations among ratios, such as the relation between ratios of forces and ratios of accelerations (Humphry, 2013a; Roche, 1998).

The absence of systems of units of measurement of psychological attributes and the absence of attempts to develop such systems are key indicators of the lack of interest in testing whether psychological attributes are quantitative. Systems of units in physics depend on established and mature physical theory in the form of quantitative relations (de Boer, 1994–1995). Thus the absence of a system of units and of any progress toward establishing such a system seemingly reflects the absence of the requisite body of workable, quantitative theories and laws. Michell has summarized the situation eloquently, stating:

> Psychology might be on the way to becoming a successful quantitative science, but as a body of workable, quantitative theories and laws, it is so far short of the example set by physics that no one yet has a clear idea of what a successful quantitative psychology would look like. The

history of science teaches us many things, but I do not think that one of them is that we can expect to make progress by ignoring pertinent matters. (1999, p. 217)

I endorse the statement above and, further, stress that if we are to measure in well-defined units, it is likely we will need to have developed at least the beginnings of a body of workable, quantitative theories and laws. Similarly, I fully support Michell's (2014) recommendation for the "critical as opposed to credulous use" (p. 115) of IRT models.

Sijtsma and Emons (2013) express the view that the realization of laws in psychology is light years away. My response to this is twofold. First, I do not think we can say how distant this goal is or even, for that matter, whether it is attainable. Second, if the authors are correct and if the physical sciences serve as any kind of guide, the successful measurement of psychological attributes in well-defined units is also light years away. It behoves us to approach the objective of measuring psychological attributes in a realistic manner and to properly qualify claims regarding progress toward the objective.

## Measurability theory or measurement theory

While I agree that no one has a clear picture of what a successful quantitative psychology would look like, I challenge Michell's (2005, 2008) claims that representational measurement theory is capable of either: (a) demonstrating the failure to measure psychological attributes or (b) guiding endeavours to measure psychological attributes. Thus, as I have said previously, I agree with the main conclusion drawn by Michell but disagree with the *basis* of the conclusion, inasmuch as that basis lies in representational theory (Humphry, 2011).

In a similar vein, it is difficult to see how IRT can achieve these objectives either. Contrary to Michell's (2014) supposition, I do not regard IRT and representational theory as competing models. There are more similarities than differences in the underlying rationales for IRT and representational measurement theory. Specifically, both are *quasi-empirical* theories that have no direct foundation in substantive psychological theory or law (Humphry, 2013b).

Regarding representational theory, Michell (2014) claimed that it "might more aptly be called measurability theory, for it tells what it is for attributes to be measurable" (p. 115). Whether called measurability or measurement theory, it is expressly prescriptive and invokes a quasi-empirical structure, as characterized by Sherry (2011):

> [A] representational theorem establishes at most the existence of a function from a quasi-empirical relational system to a numerical relational system. Here a quasi-empirical system is a set-theoretic counterpart of a domain of empirical investigation, constructed by substituting exact mathematical domains and concepts for inexact empirical ones. (p. 520)

Here lies the similarity between representation theory and IRT. IRT models invoke a quasi-empirical unit. Proponents of Rasch models, in particular, consider it a prescriptive requirement that data fit the model in order to successfully measure an attribute. Thus, for example, Sijtsma and Emons (2013) rightly say that "IRT models are prescriptive with respect to measurement but not necessarily descriptive" (p. 787). Analogously, the

representational measurement theory invokes axioms and specifies quasi-empirical relations, defined in relation to a formal set-theoretical framework. Proponents of the theory consider it a condition that such relations exist in order to infer that an attribute is measured (and measurable). Thus, both bodies of theory attempt to express requirements for successful measurement in purely formal terms; and in both cases it is claimed that to demonstrate successful measurement, researchers must show that empirical data agree with the formally expressed conditions or requirements.

Kyburg (1996) similarly suggested the empirical relational structures of representational measurement theory are overly idealized. Kyburg had also earlier observed that the applicability of representational theory rests entirely on whether the axioms actually apply in nature, stating:

> It is all very well to say that *if* a certain set of the attributes of objects and a certain operation on them obey certain axioms, *then* those attributes can be represented by a function of a certain sort from those objects to the real numbers. But we also want to know that the axioms are satisfied; we must at least face the classical problem of inductive or scientific inference. (1984, p. 3)

This problem does not arise in physical sciences because there exists a coherent body of substantive theory in quantitative form and the soundness of this body of theory can be inferred from a vast array of experimental tests confirming it. The very same body of theory provides the foundation for designing measuring instruments and procedures.

Few sources on measurement in the physical sciences even acknowledge the representational theory of measurement. For example, in a comprehensive account of measurement, *The Mathematics of Measurement: A Critical History*, Roche (1998) did not canvass the axiomatic approach to measurement set out in representational theory. In contrast, he did provide an extensive account of the origins and history of the way in which quantitative relations were understood and expressed in physics. His account of the origins and history of measurement in physics reinforces the fact that relations directly stated in a quantitative manner form the fundaments of physical measurement.

It is important for psychology as a discipline to appreciate that, notwithstanding the apparent simplicity of expressions of relations among physical quantities such as Newton's laws, a great deal of contextual knowledge gives meaning to these expressions of theory and law. As Roche (1998) observes: "The mathematical expression [of a quantitative law], although highly compact, is an incomplete part of the large body of physical information which defines the meaning, application and limitations of the physical law" (p. 232).

## The Rasch paradox

Turning our attention back to attempts to measure in psychology, Michell (2014) stated that Humphry "admits" the "[Rasch] paradox remains a theoretical possibility" (p. 113). This portrayal is inaccurate. Far from admitting it as a possibility, I unequivocally agreed that there *is* a paradox and, furthermore, stated that denial of the paradox implies pathological science as follows:

either there is a paradox or a unit must be presupposed to avoid a paradoxical conclusion. To presuppose a unit effectively presupposes that measurement is possible rather than treating it as a scientific hypothesis. To approach measurement based upon such a presupposition is the core of what Michell terms as pathological science. (Humphry, 2013a, p. 776)

This inescapable paradox pertains to IRT models in general because the models formally emulate physical equations in which terms have an implicit unit. Although IRT models are expressed algebraically like physical equations, the terms in IRT models are not measurements in specific and well-defined units: There are no definitions of units in IRT of the kind that underpin the International System of Units (SI). Psychometricians routinely claim that such models allow them to measure on interval scales, which implies that the parameters represent measurements expressed in units. However, the parameters are really just a means of summarizing patterns in empirical response data and therefore serve a relatively narrow epistemological function. In contradistinction with physical equations, IRT models are not founded in substantive theory.

Following from these points, contrary to the assertion by Sijtsma and Emons (2013), my explanation of the Rasch paradox does not depend on whether the units are physical or perceptual. As set out in sources such as Humphry (2011), the unit that I have made explicit in the Rasch model denotes a scientifically hypothesized *unit of a psychological attribute*. I do not find "implicit measurement" invoked by Sijtsma and Emons (2013) a sufficiently clear or adequate basis for progress in the endeavour to measure psychological quantities in well-defined, reproducible units. However, I agree that the Rasch paradox is somewhat tangential to the fundamentals of the debate and I will leave the remarks here for the purpose of the present article.

## Apparent paradoxes and quandaries in representational measurement theory

Related to earlier criticisms about the quasi-empirical, idealized nature of representational theory, a curious feature of Michell's arguments is his use of the term "quantitative structure" in relation to continuous quantities. At least on the surface, it seems paradoxical that a quantity like length should be continuous and undivided, yet simultaneously possess the "internal structure" referred to by Michell (1997, p. 356).

Let us adopt the following definition: structure is the arrangement of and relations between the parts or elements of something complex. If length is continuous and undivided, it contains no parts and so there cannot be relations *among the parts* of the attribute length. According to this definition of structure, it is difficult to conceive of the attribute length as having "internal structure."

It is apparent that Michell (1997) thinks of *quantity* as being literally a range of magnitudes or instances that stand in relation with each other. Now, although it is clearly useful to think in terms of specific precisely determinable magnitudes at instants in time, this is an idealization and we must remember this is the case. As Lynds (2003) observes, because all physical processes are dynamic and time cannot be frozen, physical magnitudes cannot literally be determined at an instant in time. This is well-established in the Heisenberg uncertainty principle. The concept of a precisely determined, instantaneous

magnitude is an idealization—a useful fiction. There is nothing wrong with this per se; however, we need to take care in how useful fictions are invoked, particularly where it comes to claims about the fundamental *nature* of quantitative attributes, i.e., in ontology. As will be explained, while it may be *useful* in various situations, it is by no means clear that we *must* invoke precisely determined instantaneous magnitudes in order to characterize or understand measurement.

## The concatenation of magnitudes in the representational theory of measurement

In building the representational theory of measurement, the concatenation of magnitudes is the prototype of measurement (Krantz et al., 1971; Luce & Tukey, 1964). The purported objective of the representational theory is to form a foundation for the measurement of quantities where concatenation operations are not possible, as in the social sciences. Representational measurement theory is also ambitiously referred to by Krantz et al. (1971) in the title of their work as the "Foundations of Measurement." As observed by Kyburg (1984), the representational theory is a set of formal theorems premised on axioms. Thus, the foundations of the "Foundations of Measurement" are its axioms.

The axioms inherently invoke the idealized, quasi-empirical concatenation operation. As Luce and Tukey (1964) stated, representational theory will be seen as valuable by "those who *grant* [emphasis added], in the situations where it is natural, the fundamental character of measurement axiomatized in terms of concatenation" (p. 4). Concatenation operations are certainly characteristic of common surface characteristics of instrumental measurement processes, i.e., they are relevant to the *epistemology* of measurement, particularly as it relates to the historical measurement of length.

However, the real foundation of measurement is the actual nature of physical quantities, and the nature of relations among quantities. It can by no means be taken for granted that the *nature* of physical quantities can be "axiomatized" in terms of the concatenation operation. Thus, I question the foundations of the "Foundations of Measurement." Concatenation operations are relevant to the *epistemology* of the measurement of physical quantities but they are not obviously relevant to the *ontology* of quantities. This distinction is elaborated in the following section.

Thus, the representational theory may offer insights into the considerations that arise when the quasi-empirical concatenation operation is employed in measurement. It would be perfectly justifiable to offer representational theory as a basis for offering such insights through formal means. However, there is little reason to think that the theory can offer insights into the ontology of continuous quantities, or that it will help researchers in psychology to develop substantive quantitative theory.

## A dynamic prototype for measurement

In physics, there is an obvious alternative to concatenation as a prototype for measurement and measurability, namely the enumeration of wave cycles that span a region of space in a given direction. Consider, for example, a length traversed by an electromagnetic wave such as a light wave, as represented in Figure 1. Length can be measured by
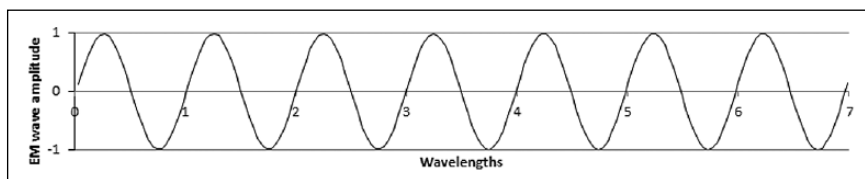
**Figure 1.** An electromagnetic wave propagating in a straight line.

enumerating wave cycles spanning a given region of space. The unit of length is either a single wavelength or a multiple of the wavelength.

Between 1960 and 1983, the metre was actually defined as a specific number of wavelengths of a specific electromagnetic emission in a vacuum. Key features of this prototype are as follows: the length traversed is continuous and unbroken; the wave cycles are continuous/unbroken and dynamic; and the passage of time is also continuous. Nevertheless, it is possible to measure length in units of wavelength by enumerating wave cycles spanning a region.

For example, if a beam of light is closely adjacent to an object and it spans the same length as the object, it is possible to infer the length of the object in units of wavelength. An interferometer can be used in practice to measure length in this fashion. No concatenation operations are necessary operationally and concatenation is irrelevant to the ontology of the physical phenomenon.

Furthermore, it is possible to establish the ratio of two lengths by determining the ratio of the number of wavelengths spanned by light traversing a length A to the number of wavelengths spanned by light traversing a length B. Relations can be determined in this manner because a wave propagates through a continuous region of space and its cycles can be enumerated. It is neither necessary nor useful to think of distance as possessing an internal structure, or to think of relations among wavelengths in terms of any internal structure. Instead, it is useful to think of such relations among numbers of cycles of the propagating wave and the associated numbers of wavelengths. The relations among wavelengths exist *due to* the *nature* of continuous space and they can be measured due to the cyclic nature of the propagation of waves through space.

In contrast, Michell makes the assertion that "half the structure of … a quantity is *due to* [emphasis added] the merely ordinal relations between magnitudes and the other half is due to additive relations between magnitudes" (Michell, 2008, p. 18). Inasmuch as the *structure of a quantity* is supposed to refer to its ontology/nature, Michell here apparently claims that the *nature* of a continuous quantity is due to relations between magnitudes. In the case of length, the above example shows that the relation of one length to another can be determined due to the ontological continuity of space and cyclic propagation of a waveform. It is unnecessary in this prototype to refer to the internal structure of length or any other quantity.

On the surface it may seem that wave phenomena have a limited scope of relevance in measurement. This, however, is not the case. Energy is related to wave frequency through the Plank relation $E = hv$, where $h$ is the Plank constant. Consequently, wave phenomena are directly relevant to the measurement of length, energy, mass, time, and temperature, and indirectly (if not also directly) to electrical and other quantities.

*Dynamic processes in psychology*

Psychological processes and phenomena cannot be frozen in time any more than physical processes can be frozen in time. It seems therefore that a dynamic prototype for measurement is relevant to both physics *and* psychology. In addition, wave phenomena are directly involved in substantive theory related to visual perception, mental states, stages of sleep, and circadian rhythms. Time is also a key variable in experimental psychology, and the SI definition of time is stated in terms of periods of radiation, a wave phenomenon.

It is worth noting that applications of Rasch's (1960/1980) Poisson model involved dynamic processes. The Poisson was used to model numbers of errors in real-time oral reading of texts and to model reading speed. Errors and speed were modelled as a function of the ratio of reader ability to text difficulty. This is one possible avenue to measuring quantitative attributes in psychology that is rarely pursued in recent times, even by proponents of Rasch models.

Notwithstanding the dynamic nature of psychological processes, it may be defensible to conceive of psychological attributes as relatively specific magnitudes over a defined interval. In physics, temperature is a manifestation of thermodynamic processes and its measurement involves continuously dynamic energy exchanges; yet in practical terms, it is fruitful to conceive of a given temperature as a relatively stable magnitude over a small time interval. Similarly, although cognitive processes are dynamic, it may be useful to conceive of mental abilities as relatively stable over a given time interval.

However, it does not seem fruitful to suppose that psychological attributes possess an internal structure that exists due to relations between instantaneous magnitudes. Indeed, it seems likely that to approach measurement with this presupposition is doomed from the outset. If it were absolutely necessary to characterize measurement as it is characterized in the representational measurement theory, it would be necessary to grapple with how to satisfy criteria for successful measurement that have been set out by Krantz et al. (1971). However, given there is a prototype for measurement which gives precedence to continuity and dynamic processes, and does not invoke instantaneous magnitudes, the relevance of representational theory to psychology is questionable.

At the same time, it is stressed that it is also questionable whether psychologists can succeed in measuring psychological attributes using IRT. The chief advantage of IRT is its practical utility; IRT is not superior as (so-called) measurement theory and is not firmly grounded in substantive theory.

## Conclusion

Judging from physics, it is likely that the sine qua non of successful measurement is that it is possible to obtain the same measurements of a given quantity in a given unit based on two or more distinctive approaches, each of which is based on a *substantive* theoretical foundation. This is what has been referred to as the mutual grounding of different measurement methods (Chang, 2004). An example is the measurement of distance using a tape measure or interferometry. Another example is the measurement of temperature using a thermometer and a thermistor.

Even quantities that are said to be directly measured may also be indirectly measured by exploiting quantitative relations between two or more kinds of quantity. For example, length may be directly measured by concatenating an object, and it may also be indirectly measured using an interferometer due to the nature of electromagnetic wave propagation and phenomena. Thus it is even possible to indirectly measure attributes, such as length, that are generally thought of as being measured directly. Consequently, it is reasonable to suggest that the sine qua non of *all* successful measurement is that it is possible to obtain the same or directly comparable measurements of a quantity using distinct approaches and instruments based on substantive theory.

If measurement based on distinctive approaches is indeed the sine qua non of successful measurement, neither representational theory nor IRT are fundamentally necessary to achieving the aim. Nevertheless IRT may well be useful and representational measurement may or may not provide useful insights.

It has proven possible to measure a given physical attribute in different ways because principles of both the design and operation of measurement instruments are based on substantive, quantitative theory. The relevant quantitative theory encompasses a number of specific, causal relations between quantities. Instruments and procedures are designed to: (a) isolate one causal relation from other causal relations and (b) minimize the effect of extraneous variables on the outcomes of the procedures. When the design of instruments and procedures is based on substantive theory, it is invariably possible to measure quantities in the same, well-defined *measurement units* based on separate and distinctive approaches. It seems likely, therefore, that the primary challenge faced in quantitative psychology is to posit testable substantive theories or laws which can form a foundation for measurement.

## Declaration of conflicting interests

## Funding

## References

Borsboom, D., & Mellenbergh, G. J. (2004). Why psychometrics is not pathological: A comment on Michell. *Theory & Psychology*, *14*, 105–120. doi: 10.1177/0959354304040200

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. Oxford, UK: Oxford University Press.

de Boer, J. (1994–1995). On the history of quantity calculus and the international system. *Metrologia*, *31*, 405–429.

Hebra, A. J. (2010). *The physics of metrology. All about instruments: From trundle wheels to atomic clocks*. New York, NY: Springer.

Humphry, S. M. (2011). The role of the unit in physics and psychometrics. *Measurement: Interdisciplinary Research and Perspectives*, *9*, 1–24.

Humphry, S. M. (2013a). A middle path between abandoning measurement and measurement theory. *Theory & Psychology*, *23*, 770–785. doi: 10.1177/0959354313499638

Humphry, S. M. (2013b). Understanding measurement in light of its origins. *Frontiers in Psychology*, *4*(113). Retrieved from http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00113/full

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement: Vol. I. Additive and polynomial representations*. New York, NY: Academic Press.

Kyburg, H. E. (1984). *Theory and measurement*. Cambridge, UK: Cambridge University Press.

Kyburg, H. E. (1996). Quantities, magnitudes and numbers. *Philosophy of Science*, *64*, 377–410.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new scale type of fundamental measurement. *Journal of Mathematical Psychology*, *1*(1), 1–27.

Lynds, P. (2003). Time and classical and quantum mechanics: Indeterminacy vs. discontinuity. *Foundations of Physics Letters*, *16*(4), 343–355. doi: 10.1023/A:1025361725408

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, *88*, 355–383.

Michell, J. (1999). *Measurement in psychology*. Cambridge, UK: Cambridge University Press.

Michell, J. (2005). The logic of measurement: A realist overview. *Measurement*, *38*, 285–294.

Michell, J. (2008). Is psychometrics pathological science? *Measurement: Interdisciplinary Research and Perspectives*, *6*(1–2), 7–24.

Michell, J. (2014). The Rasch paradox, conjoint measurement, and psychometrics: Response to Humphry and Sijtsma. *Theory & Psychology*, *24*, 111–123. doi: 10.1177/0959354313517524

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: The University of Chicago Press. (Original work published 1960)

Roche, J. (1998). *The mathematics of measurement: A critical history*. London, UK: The Athlone Press.

Sherry, D. (2011). Thermoscopes, thermometers, and the foundations of measurement. *Studies in History and Philosophy of Science*, *42*, 509–524.

Sijtsma, K. (2012). Psychological measurement between physics and statistics. *Theory & Psychology*, *22*, 786–809. doi: 10.1177/0959354312454353

Sijtsma, K., & Emons, W. H. M. (2013). Separating models, ideas, and data to avoid a paradox: Rejoinder to Humphry. *Theory & Psychology*, *23*, 786–796. doi: 10.1177/0959354313503724

## Author biography

Stephen M. Humphry is an Associate Professor in the Graduate School of Education at The University of Western Australia. His primary field of research is psychometrics and his broader interests include the history and philosophy of science, with a particular focus on measurement across disciplines.