

On the conceptual foundations of psychological measurement

Andrew Maul¹, Mark Wilson², David Torres Irribarra²

¹ School of Education, University of Colorado, Boulder, CO, USA

² Graduate School of Education, University of California, Berkeley, CA, USA

E-mail: andrew.maul@colorado.edu

Abstract. Measurement has long been an important element of epistemology in the physical sciences and natural philosophy. More recently, the psychological sciences have developed a variety of techniques that purport to be instances of measurement as well. However, it is not clear how the understanding of measurement invoked in psychological science applications accords with the understanding of measurement found in other scientific disciplines. A sharper focus on conceptual clarity and coherence across the psychological and physical sciences has the potential to add a great deal to efforts to improve such practices. In this paper, we argue that it is possible to formulate a philosophically coherent account of how measurement works in both the physical and the human sciences.

1. Introduction

Measurement has long been an important and prominent concept in the physical sciences and natural philosophy. More recently, the psychological sciences have developed a variety of techniques that purport to be instances of measurement as well. However, it is not clear how the way in which the concept of measurement is understood in these disciplines accords with the way it is understood in science more generally, or by the professional and general publics.

In recent years, a number of scholars (e.g., [1-5]) have subjected the conceptual and philosophical foundations of psychological measurement to vigorous investigation and critique. On one reading, this body of work seems to suggest that the way in which measurement is understood by psychological scientists is entirely dissimilar to the way in which it is understood by physical scientists and philosophers of science, and, by the criteria of the latter group, there may not have yet been any instances of successful measurement of psychological attributes, or at least no such instances that can be described in a philosophically coherent manner.

In our experience, psychological scientists who are aware of this body of work tend to react to it one of two distinct ways. The first is to accept the premise that there is but one definition of measurement throughout all of science, and that the activities and conceptual vocabulary of psychological scientists are out of touch with this definition. The second is to deny that the concept of measurement has or needs to have a consistent definition across scientific disciplines, and thus conclude that differences in understandings of measurement are not necessarily problematic. In the first case, the word ‘measurement’, when used by psychological scientists, would be a metaphor at best and a conceptual error at worst; in the second case, the word would be merely a homonym. We argue that it is not necessary for psychological scientists to react in either of these ways. Instead, we claim that it is possible to formulate a philosophically coherent account of how psychological measurement could work, in a manner consistent with mainstream positions in philosophy of science.



The structure of the paper is as follows. First, we review the ways in which three major intellectual strands have influenced thinking about both psychological measurement: first, *empiricism*, second, *pragmatism*, and third, *scientific realism*. We then argue that it is possible to formulate a coherent account of how measurement works that is consistent with certain important and relevant motivations of each of these strands of thought, but that also avoids pitfalls and contradictions associated with more severe formulations of each of them; furthermore, such a new formulation clarifies that there are foundational ideas about measurement that apply in both physical and psychological scientific disciplines, and thus that different fields' conceptions of measurement need not and should not be thought of as incompatible.

2. Empiricism and psychological measurement

A wide range of philosophical positions falls under the broad heading of *empiricism*; they share in common a commitment to direct observation as the basis for knowledge. In the twentieth century, the movement known as *logical positivism* synthesized many ideas from classical empiricism along with then-current advances in the philosophy of language and mathematics. Logical positivism was associated with the position that statements regarding unobservable (theoretical) entities or forces should only be regarded as meaningful if such statements can be linked to observations in a clear and consistent manner. In the fledgling psychological sciences, *behaviorism* captured many of the same intuitions as those behind positivism. At this time, the concept of the human mind was regarded as too metaphysical and unobservable to be a proper object of scientific inquiry.

The positivist project is now widely regarded to have failed, and behaviorism as a logical doctrine has been abandoned [6]; modern philosophies of science typically include a much greater degree of acceptance of the inclusion of unobservable phenomena—such as the human mind—in scientific theories. However, positivism and behaviorism have left a significant legacy on methodological thinking in the psychological sciences [6, 7], including thinking about measurement. Notably, two major strands of thinking about measurement emerged in the early-to-mid 20th century that are consistent with many positivist principles. The first is *operationalism* [8], which is characterized by the stance that the meaning of any theoretical concept is exhausted by the operations undertaken to measure instances of the concept. The second is *representational measurement theory* (RMT; [9]), which is characterized by the stance that measurement is the construction of morphisms between empirical and numerical relations. The most widely-used definition of measurement in psychology, namely Stevens' view that “measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules” ([10], p.667), is consistent with the principles of both operationalism and RMT.

Operationalism had a strong influence on psychology through the influence of early behaviourists [7] such as Boring [11] and his student Stevens [10]. One of the reasons for this is surely the difficulty of precisely defining psychological attributes; stating that ‘intelligence is what the tests test’ ([11], p.35) neatly sidesteps the issue, and also gives at least the appearance of rigor and objectivity by anchoring abstract ideas in observables.

Operationalism has been almost uniformly rejected as irreconcilable with general scientific practice and vocabulary [6]. To see why, note that operationalism has the consequence that each unique set of operations must be associated with a distinct theoretical term; thus, for example, the outcome of the application of an alcohol thermometer and the outcome of the application of a mercury thermometer cannot refer to the same theoretical property. Additionally, the concept of measurement error—a central concept in psychological measurement—is ill-fitting with operationalism: if the results of applying a procedure are by definition a measurement of the theoretical term, what is there to be in error about?

Representational measurement theory shares in common with operationalism an empiricist emphasis on the exhaustion of the meaning of measurement claims by observable facts. More specifically, RMT holds that to measure is to construct a representation of an empirical relational system via a numerical relational system. On this view, the starting point for measurement is the

determination of empirical relations amongst objects (e.g., X is greater than Y and less than Z). This requires that empirical relations be directly observable, or “identifiable” ([12], p.7), though it is not always obvious what this means (c.f. [1, 3]).

In its unadorned form, RMT suffers from many of the same weaknesses as operationalism. RMT cannot account for the role of measurement error except via the introduction of metaphysics [3], and again conflates ontology and epistemology by denying that a measured attribute has an existence beyond the manner in which (relations between) instances of it are observed. Thus, it is very difficult to make sense of how RMT accords with both scientific and lay discourse about measurement and knowledge acquisition more generally [1].

Following the collapse of logical positivism and an associated general retreat from extreme forms of empiricism, many scholars became increasingly willing to accept that the interpretation of concepts like temperature and knowledge outrun their associated measurement procedures or observable relations.

3. Pragmatism and psychological measurement

Philosophically *pragmatist* orientations generally share in common an orientation towards action (i.e., practice) and utility: “The pragmatic method ... is to try to interpret each notion by tracing its respective practical consequences. What difference would it practically make to anyone if this notion rather than that notion were true? If no practical difference whatever can be traced, then the alternatives mean practically the same thing, and all dispute is idle” ([12], p.14).

From the pragmatist's perspective, the purpose of beliefs and theories is to facilitate successful engagement with the world. Thus, a pragmatic orientation towards measurement might de-emphasize concerns such as the ontology of the measured attribute or the manner in which numerical assignments are formally constructed out of empirical relations, and ask instead whether the results of the measurement procedure can be usefully applied. The concept of ‘usefulness’ is, of course, relative to the motivations and values of individuals and society as a whole, and thus as such motivations and values change, so might the usefulness of the measure.

Psychological and educational tests and assessments are generally developed with a certain purpose or set of purposes in mind and are deeply embedded in a complex social structure; thus, a focus on the practical consequences of test use seems sensible. Such a focus is reflected in, for example, the concept of *validity* (in psychological testing contexts), crucial aspects of which are defined with reference to the actions taken on the basis of testing procedures as well as the direct interpretation of the results of such procedures (e.g., [13]).

However, one possible danger of the emphasis on the uses of tests and the consequences of such use is that the concept of *measurement*, as a knowledge-acquisition enterprise, can get buried. Testing and measurement are not the same thing, and testing procedures that are successful by many important pragmatic criteria without any measurement taking place (for example, when they are used as devices for prediction, as formalizations of values, as social signals, and as deterrents of certain behaviors or prods to certain forms of action); however, it seems common in the psychological sciences to speak of testing and measurement as if they were synonymous [14]. Moreover, the lack of a distinction between these concepts makes it more difficult to recognize when genuine claims about the measurement of psychological attributes are implicit in larger claims about the use of tests. For example, it may be claimed that it is appropriate to use the results of educational tests to guide decisions about the retention or firing of teachers. It is very difficult to make sense of this claim without the additional claims that (a) educational tests measure important attributes of students, (b) teachers have causal potency over these student attributes, and so (c) educational tests can be used to measure the effectiveness of teachers. Articulation and evaluation of such measurement claims are completely missing from most discussions of the ‘validity’ of the use of educational tests for the purposes of teacher evaluation (e.g., [15]).

Thus, a great deal of thinking about measurement in psychology is characterized by the views that tests should be evaluated with reference to their intended purposes, and that testing is synonymous

with measurement. The first view we take to be quite reasonable; the second, however, is clearly false, and surely contributes both to the seeming incompatibility of conceptions of measurement in the physical and psychological sciences.

4. Realism and psychological measurement

As with empiricism and pragmatism, there are a broad range of positions that fall under the broad heading of *scientific realism*; they share in common the commitments that (a) there is a (single) natural world, which exists regardless of what any conscious being thinks or perceives (the ‘metaphysical commitment’); (b) scientific claims about the world are to be taken at face value, as possessing truth-values (the ‘semantic commitment’), and (c) so interpreted, true scientific claims constitute knowledge of the world (the ‘epistemological commitment’).

Michell has argued that the ‘classical’ understanding of the concept of measurement—the discovery of ratios of magnitudes of quantity—entails realism. More specifically, to the extent that “scientists see themselves as investigating independently existing natural systems and see their theories and hypotheses as attempts to capture something of the structure and ways of working of such systems” [2, p.286] the common understanding of the act of scientific discovery—and, *a fortiori*, measurement—invokes all three of the aforementioned realist commitments.

Michell’s account is consistent with (but does not explicitly require) *metaphysical realism*. Metaphysical realism entails the *correspondence theory of truth* (the position that statements are true if they directly correspond to facts in the world), and implies that scientific objectivity depends on the subject matter of science itself being mind-independent, or *ontologically objective* [16]. In the context of psychological measurement, the implication that measured attributes of objects must possess mind-independent existence in order to be measurable is often interpreted as implying that they must possess physical (perhaps specifically neurophysiological) identity, and in some contexts a genetically-determined biological basis for variation in the attribute.

Further, for attributes to be measureable in the relevant population, there must be differences in the ways in which the attribute is instantiated in the relevant population. Whether these differences must themselves possess a certain sort of structure is a point of some contention in the literature. Michell argues that, by definition, measurement requires that the measured attribute possess a quantitative structure; other sources do not impose this restriction: for example, in the International Vocabulary of Metrology [17] it is stated that ordinal attributes are measurable, and some realists (e.g., [3]) use the concept of measurement in reference to attributes with purely nominal differences, such as gender.

A number of challenges have been raised to metaphysical realism. In our view, the most compelling of these invoke the observation that there are simply too many ways in which beliefs and symbols can be mapped onto the world for it to be plausible that there is a single complete and true description of the way the world really is (e.g., [18]). A commitment to scientific realism need not entail the belief that the world consists of a fixed totality of mind-independent objects and their properties; rather, it is possible to maintain a realist view of measurement while acknowledging that *knowledge* is constructed by humans, and can be constructed in multiple ways depending on the pragmatically determined frames of reference.

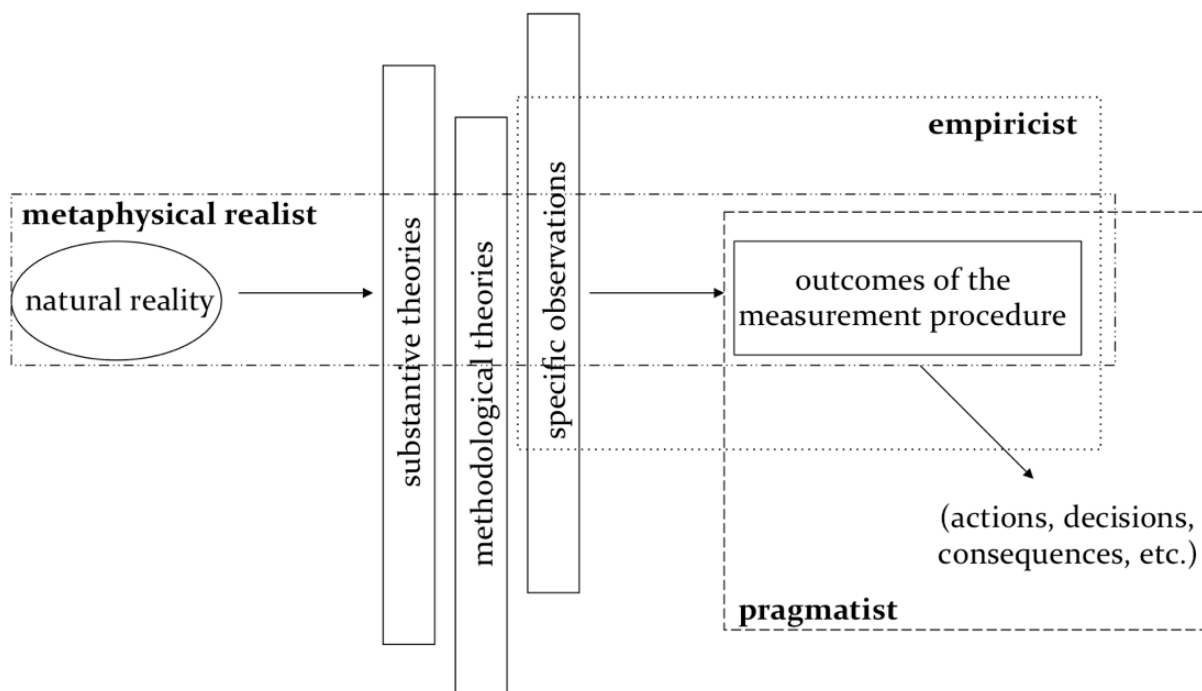
5. Synthesis: A pragmatic-realist view of measurement

Empiricism is motivated by the intuition that the preferred method of acquiring knowledge is through observation and experience. Pragmatism is motivated by the intuition that theories are made useful via their links with practice. Realism is motivated by the intuition that that scientific inquiry seeks to gain knowledge about a natural world. None of these intuitions contradicts the others; however, severe formulations of each position may wind up saying something false by denying the correct intuitions of the other views. In our view, extreme proponents of operationalism and representational measurement theory overreach in denying that attributes exist independently of observations; extreme advocates of pragmatism overreach in denying that epistemic tools such as measurement seek to gain knowledge of a natural world; and finally, extreme supporters of

metaphysical realism overreach in denying that knowledge is constructed by humans, and that it can be constructed in multiple correct ways depending on the observer's pragmatic concerns.

An example of a philosophic framework that is consistent with each of these intuitions is found in Putnam's recent writings on *pragmatic realism* [18], which acknowledge that conceptual relativity is not at odds with realism, but rather, "to use a Wittgensteinian idiom, *seeing* is always *seeing as*, and it is the interface between the world and the rich fabric of our concepts that jointly determines what we see." Thus realism does not necessarily entail a correspondence view of truth, nor does it deny that our conceptual schemes, models, and linguistic frameworks actively shape our experience of the world and frame our knowledge of it. Our application of the pragmatic-realist view of measurement is illustrated in Figure 1. The dotted boxes show how severe formulations of each of the aforementioned philosophical stances see a smaller part of a bigger picture. In this account, the existence of natural reality is not denied, but neither is it seen as directly presented to our senses; instead, our various substantive and methodological theories and pragmatic concerns cause us to organize and prioritize experience in a particular way.

Figure 1: A pragmatic-realist view of measurement
epistemic layers



With regards to psychological attributes, contrary to the claims of metaphysical realism, it is not necessary for such attributes to be ontologically objective for them to be real components of the natural world. Elsewhere [19], we have discussed how the distinction between ontological and epistemic subjectivity and objectivity [16], and the recognition of the existence of intentionality-dependence of objects and attributes of objects, permit the formulation of a coherent realist account of the ontology of psychological attributes. Briefly, psychological attributes can (a) involve conscious phenomena with subjective first-person ontology, and (b) have conceptual boundaries delineated by contextually and pragmatically-driven frames of reference, rather than being natural kinds (or attributes) in the classic sense. Further, the connection between natural reality and the outcomes of a measurement procedure is not in itself compromised by the fact that we choose to privilege certain contrast classes, levels of explanation, methods of summarization, and modes of description. Notably,

scientific models (including statistical models) serve precisely the purpose of organizing experience. This implies that scientists have the responsibility to be aware of and acknowledge the role that their conceptual frameworks play in shaping their investigations, and to explicate and empirically investigate the hypothesized connections between the psychological realities under investigation and the outcomes of purported measurement procedures.

We think the account of measurement given here shows that measurement *is* concerned with gaining knowledge about the real world, that it *is* an empirical process, and that it *is* shaped by the pragmatic concerns of the agents involved. Further, there is nothing logically distinct about the structure of this process when applied in the psychological sciences and the physical sciences.

References

- [1] J. Michell, An introduction to the logic of psychological measurement. Lawrence Erlbaum Associates, 1990.
- [2] J. Michell, The logic of measurement: A realist overview, *Measurement*, 38, 285–294, 2005.
- [3] D. Borsboom, Measuring the mind: Conceptual issues in contemporary psychometrics. Cambridge University Press, 2005.
- [4] G. Trendler, Measurement theory, psychology and the revolution that cannot happen. *Theory & Psychology*, 19, 579–599, 2009.
- [5] N. Cliff, Abstract measurement theory and the revolution that never happened. *Psychological Science*, 1992.
- [6] M. Bickhard, The tragedy of operationalism. *Theory and Psychology*, 11, 35–44, 2001
- [7] C. Green, Of immortal mythological beasts: Operationism in psychology. *Theory and Psychology*, 2, 291–320, 1992.
- [8] P. Bridgman, The logic of modern physics. New York: Macmillan, 1927.
- [9] D.H. Krantz, R. D. Luce, P. Suppes, A. Tversky. Foundations of measurement volume I: additive and polynomial representations. Vol. 1. Dover Publications, 2006.
- [10] S.S. Stevens, The operational basis of psychology. *The American Journal of Psychology*, 1935.
- [11] E. Boring, Intelligence as the tests test it. *New Republic*, 35–37, 1923.
- [12] W. James, Pragmatism: a new name for some old ways of thinking. Cambridge MA: Harvard University Press, 1907.
- [13] S. Messick, Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp.13–103). New York: American Council on Education/Macmillan, 1989.
- [14] D. Borsboom, Educational measurement, 4th edition: Book review. *Structural equation modeling*, 16, 702–711, 2009.
- [15] L. Shepard, Evaluating test-based teacher evaluation: Validity as a theory-of-action framework. *Applied Measurement in Education* (forthcoming).
- [16] J. Searle, The rediscovery of the mind. MIT Press, 1992.
- [17] JGCM 200:2012, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM), Joint Committee for Guides in Metrology, 2012 (2008 editions with minor corrections), <http://www.bipm.org/en/publications/guides/vim.html>.
- [18] H. Putnam. The threefold cord: Mind, body, and world. Columbia University Press, 1999.
- [19] A. Maul, The ontology of psychological attributes. Paper presented at the annual meeting of the National Council for Measurement in Education, Vancouver, April 13–16, 2012.