

Validity and Measurement

Michael D. Maraun

*Department of Psychology
Simon Fraser University*

As illuminated forcefully by Professor Newton's provocative analytical and historical excursion, as long as tests are employed to practical ends (prediction, selection, etc.) there is little cause for the metatheoretic angst that occasions rounds of papers on the topic of validity. But then, also, there seems little need, within this context of test employment, for the concept validity itself, its presence in discourse merely serving to obscure what are typically well-defined technical issues.

Perplexity and angst over the issue of validity arises from its role as a back door to measurement, the psychologist's wanting to have his or her cake and eat it too; cleave to the belief that his or her tests measure, while demurring in the use of the full-blooded measurement language that has, in the past, lead to his or her censure.

Despite its lengthy history, the volumes that have been written about it, the study of test validity is neither evolving, nor approaching resolution of a core set of defining problems, certainly none such bearing on the topic of measurement. It continues, rather, to spin its wheels, as evidenced by the lack of "precision, consistency, and clarity" associated with so many of its central offerings.

The explanation for this state of affairs is that the back door did not lead to measurement at all but was, instead, an illusion, a projection of a primitive metaphysical picture the commitment to which distorts understanding of measurement through a fundamental conflation of its conceptual and empirical components and leads to incoherence. Because incoherence is irresolvable, perplexity forever attends the consideration of test validity, to be encountered afresh by scholars every few years.

The issue of whether construct validity should be taken to be the whole of validity is largely irrelevant in light of the fact that the psychologist sees all of science through the lenses created by, among notable others, Cronbach, Meehl, Thurstone, Lazarsfeld, Lord, and Novick, out of materials provided by the empirical realist philosophy of Feigl, Hempel, and Sellars that rose up midcentury to crush positivism and its offspring, operationism.

Cronbach's and Meehl's (1955) shotgun-scatter account of construct validity is simply their reading of empirical realist philosophy, and Borsboom's causal account of measurement is just

a repackaging of the empirical realist–inspired causal account of measurement on which was erected construct validity.¹

But although Cronbach and Meehl drew their inspiration from empirical realist philosophy, they, apparently, misunderstood a number of its key details. These misunderstandings were most saliently manifested in their failure to maintain the distinction between a *theoretical term* (a concept, hence, an element of language, which is employed in the articulation of a theory, and, on the empirical realist account, denotes the elements of a class of causal entities hypothesized to exist) and its *referents* (the denoted, causal entities hypothesized, under the theory, to exist and that may be, if they do exist, selected as targets of scientific investigation).

The failure to maintain this distinction was enshrined within construct validity as a blanket annihilation of the distinction between the conceptual and empirical components of science in general, and measurement in particular. Contrary to what Professor Newton supposes, the *raison d'être* of the concept construct is not to allow for the expression of “the difference between how the world is and how we currently understand it” (this issue). The questionableness of this explanation is suggested in a simple noting of the fact that no other science has seen the need for such a concept.

The concept construct is, in fact, what became of *concept* and *referent* when the distinction between the two was eradicated. It is an illegitimate melding of the two, an allowance for the psychologist to slide back and forth ambiguously, and obliviously, between the, now sublimated, conceptual, and the empirical components of science. Thus, in claiming that “we currently don’t know what *anxiety* means . . . [we] are engaged in scientific work to answer this question” (1955, p. 20), Cronbach and Meehl ascribe to (what they take to be) a construct *both* a meaning (something a concept has) and status as a target of empirical investigation (something a constituent of nature can have). Or, as Strauss (1999, p. 19) puts it, “One major concern in psychology is to define and describe psychological constructs.” Concepts are defined; the constituents of nature denoted by some concept are described.

The historical record indicates that the psychologist has never succeeded in grasping the place of language in science, hence, where meaning comes from, hence, what is meant by *conceptual*. He, variably, misconstrues *conceptual* as having to do with hunches, speculations, or theories (as does Professor Newton in his target article); but the latter are propositional forms, hence, are *about* the empirical (hence, *presuppose* the conceptual).

Let us review fundamentals. Concepts are elements of language. A concept’s range of correct employments is fixed by linguistic rules (the concept’s grammar). The linguistic rules that fix the range of correct employments of *certain* concepts (e.g., neutrino, bachelor, rock) grant ascription of these concepts to entities potentially extant in the world. Such a concept (its meaning fixed in language, by linguistic rules) denotes its referents (which, should they be extant in the world, have properties that can be investigated as a matter of science).

A very small subset of concepts φ (among them, mass, length, velocity) are embedded in normative practices of measurement, practices the constituent parts of which include rules RM_{φ}

¹As the reader will come to realize upon studying the historical lineage of statements such as the following: “The problem faced by the investigator is, ‘What constructs accounts for variance in test performance’” (Cronbach & Meehl, 1955, p. 2); “Determining what psychological constructs account for test performance is desirable for almost any test” (1955, p.2); “A numerical statement of the degree of construct validity would be a statement of the proportion of the test score variance that is attributable to the construct variable” (1955, p. 8).

by which the “ φ ” of objects is measured; units in terms of which measurements are expressed; rules for the translation of one system of units into another.² A concept φ that is embedded in a practice of measurement can be correctly ascribed to numbers, and *is* correctly ascribed to a number φ^* just in the case in which φ^* was produced in conformity with RM_φ (in which case φ^* is a measurement, in particular units, of the “ φ ” of whatever was measured).

One takes a measurement φ^* of the “ φ ” of something (whose “ φ ” can be measured) when one’s actions in producing φ^* are in conformity with RM_φ . Meaning is conferred through concept ascription; meaning is conferred upon a number (*as* a measurement of the “ φ ” of something) when concept φ is correctly ascribed to the number and the concept’s grammar grants such an ascription, just in the case in which the number was taken in conformity with RM_φ .

Measurement manifests the standard conceptual/empirical autonomies, hence (a) the empirical properties of the distribution of the measurements taken of the “ φ ” of a collection of objects has no bearing on RM_φ (eo ipso, what it is to correctly measure the “ φ ” of objects); and (b) RM_φ (eo ipso, what it is to correctly measure the “ φ ” of objects) has no bearing on the empirical properties of the distribution of the measurements taken of the “ φ ” of a collection of objects.

That men are, on average, say, 167 pounds has no implications for what it is to correctly measure mass nor, certainly, whether the numbers on which the claim is based are, in fact, measurements of mass. On the contrary, the claim is not a claim *about* mass at all unless the numbers are masses, hence, unless they were taken in conformity with the rules for the measurement of mass.

In biology, chemistry, physics, etc., there is the careful separation of definition (of concept) and scientific, empirical investigation (of extant referents of concepts) and of the taking of measurements (in accordance with rules that must be learned by all natural scientists) and the employment of these measurements to investigate the objects and entities measured.

In psychology, on the other hand, there is a raging metaphysics under which are generated incoherences of the following sort: (a) science is the proper tool to both resolve issues of meaning and reveal the nature of the world; (b) a putative measurement instrument (a test) measures the cause of variation in the scores it assigns to respondents on the basis of their behavior;³ (c) from (b), because test behavior is always caused, a putative measurement instrument (a test) always measures *something*; (d) the something the instrument measures is unknown (it is as if the instrument were a net thrown into the sea of psychological unobservables, the task being to make guesses as to the nature of the catch it yields); (e) from (b), measurement is a matter of detection, and measurement instruments are detectors (they pick up on the action of extant, unobservable causal entities); (f) measurement cases are empirical cases; (g) psychological concepts such as intelligence, dominance, happiness, etc., can be freely paired with a test and some psychometrics to yield measurement (of the phenomena that they denote).⁴

²It is not the case that such a concept and its associated measurement practice are separable; rather, the concept originated with a “method of projection” (Baker & Hacker, 1980).

³The cause of (what brings about) variation in the power outputs (in horsepower) of, say, a set of combustion engines is certainly not power but, rather, the fact that the engines were built differently. The empirical investigation of power output, including the impact upon power output of engine design, presupposes the independent capacity to measure the power outputs of work-producing entities.

⁴In fact, as their grammars make clear, psychological concepts are not embedded in normative practices of measurement, hence, the phenomena they denote cannot be measured; their handling by the psychologist constitutes a systematic *deformation* of their meanings.

Though I admire Professor Newton's efforts to bring clarity to thinking on the topic of test validity, as long as test validity is informed by a metaphysics, clarity will remain an unattainable goal, and incoherence a constant companion.

REFERENCES

- Baker, G., & Hacker, P. (1980). *Wittgenstein: Meaning and understanding*. Chicago, IL: University of Chicago Press.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Strauss, B. (1999). Latent trait and latent class models. *International Journal of Sport Psychology*, 30(1), 17–40.