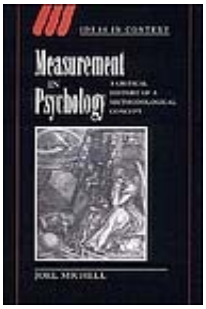
PsycCRITIQUES
Contemporary Psychology: APA Review of Books
American Psychological Association

Empirical Quantification Is the Key to Measurement—Including Psychological Measurement

A review of



Measurement in Psychology: A Critical History of a Methodological Concept
by Joel Michell

New York: Cambridge University Press, 1999. 246 pp. ISBN 0-521-62120-8. $54.95

Reviewed by
R. Duncan Luce

Any psychologist engaged in measuring psychological attributes should read this very readable, scholarly book. Will she or he like it? With a few exceptions, probably not. Should she or he, nonetheless, pay attention to it? I think so. But then, as you will see, I am far from an unbiased reviewer.

The theme of the book is that beginning with Fechner and including such measurement scientists as Spearman, Cattell, Thorndike, Kelley, Thurstone, Lord, Novick, and, especially, S. S. Stevens, measurability has been presumed rather than tested. The focus of most of psychometrics and psychophysical measurement has been, and continues to be, how best to infer measures when it is postulated, largely without discussion, that the attribute of interest is quantitative in nature. Michell contends that this involves placing the cart before the horse in that it is an empirical matter to ascertain whether an attribute is actually quantitative. “If ever there was, in the history of any science, a theory accepted just because it answered questions rather than investigated because it raised them, [mental testing] is it” (p. 107).

This, of course, is not a new issue. It was, after all, the focus of the so-called Ferguson Committee of the British Association for the Advancement of Science whose desultory deliberations lasted a decade, ending with a report in 1940 (pp. 143-155) that was totally split.[1] The physical scientists, heavily dominated by the philosopher of physics N. R. Campbell, contended that an attribute's being quantitative entails two things. The first thing

(Condition 1) is an empirical demonstration that the attribute in question exhibits properties sufficient for a numerical representation. It is not a matter to be decided by fiat, in particular, not by an operational definition such as the subjective equality of just-noticeable differences, the normal distribution of intelligence scores, or the assumption that magnitude estimates form a ratio scale. The second thing (Condition 2) is that the only empirical ways to decide the issue are either to have (a) a binary operation on entities that can be shown to have an additive representation, as in length and mass measurement, or (b) two such additive measures that are discovered to exhibit an empirical invariance that then serves as a derived measure. An example of Item b is density: the invariant ratio of the mass to the volume of a homogeneous substance. Moreover, all committee members, including the psychologists, agreed that (Condition 3) psychology has no such additive operations distinct from the physical ones. Therefore, the physicists concluded there was no possibility of quantitative estimates of sensory events. Because Condition 3 was undeniable at this time, the psychologists of the time had three options: agree with the physicists that legitimate, fundamental, psychological measurement is impossible; ignore Condition 1; or show that the claim of Condition 2 is more limited than it needs to be.

Michell carefully reconstructs the history and concludes that psychologists had been operating to that point by ignoring Condition 1 and that, for the most part, they continue to do so to this day. Boring (1920, p. 104) recognized early on the issue to which Kelley (1923, p. 104) responded:“Boring's conclusions are generally destructive, and tend to leave one with the feeling that there is no sound statistical basis for mental measurement, and little for other psychological measurement” (p. 104). “The founding fathers of modern psychology, almost to a man, simply presumed that measurement was a scientific imperative and, accordingly, thought to contrive quantification” (p. 3). Stevens (1946, 1951, p. 15) enshrined the prevailing view in his famous dictum that measurement is the “assignment of numerals to objects or events according to rules” (p. 15), indeed any operational rule, not just empirical ones. Moreover, Stevens claimed that Campbell had interpreted Condition 2 much more narrowly than was warranted by his own philosophy of measurement, which seems to Michell and to me a remarkable misreading of Campbell's detailed writings on the subject.

Although Campbell had a reasonably clear idea of the mathematics underlying additive measurement, he did not reference the 1887 ideas of Helmholtz (pp. 68-74) which were fully axiomatized by Holder's (1901, p. 74) formulation of what must be empirically true to achieve quantification using an order plus an operation-what has

come to be called *extensive measurement*. Indeed, Michell takes the plausible position that this type of measurement is the basic source of the real number system. He is at some pains to distinguish quantification as provided by extensive structures from representation theories that map structures into the real number system. I do not think the issue is as deep as he does, but this is not the place to discuss it. And, in any event, Michell and I agree that such additive measurement was well understood.

Suppes and Zinnes (1963, p. 198-199) formulated explicitly the measurement perspective that quantification had to be justified empirically and presented Suppes's (1951, p. 197) modified axiomatization of extensive measurement and discussed derived measurement in detail. They “elevated measurement theory to a level of rigour not attained since Hölder (1901) and Weiner (1919)” (p. 199). Axiomatizations justifying quantification in specific situations different from the extensive ones were ordinal, semiorders (which are close to ordinal), and difference and absolute difference structures, which are both fairly easily reduced to extensive measurement. They chose not to cite either the axiomatization of expected utility (von Neumann & Morgenstern, 1947) because the primitives included numbers (probabilities) or that of subjective expected utility (Savage, 1954) because they deemed it at the time too complex to explain (P. Suppes, personal communication, November 1, 1999). It is surprising that Michell continues to ignore this very original work when citing the history of discoveries that broadened appreciably the scope of Condition 2 to structures not having an extensive operation. He flatteringly, but not entirely accurately, credits the Luce and Tukey (1964, p. 199) introduction of additive conjoint measurement as a “revolutionary” alternative to extensive measurement. That work studied the tradeoffs of factors affecting a single attribute which Tukey and I showed was another source of quantification. In fact, although we were unaware of it at the time, Debreu (1960, p. 200) had earlier discovered a similar, but topological axiomatization. The 1964 formulation was algebraic and somewhat more general than Debreu's. Krantz (1964, p. 200) and, more transparently, Holman (1971) showed how to construct an extensive operation from the trade-off and so reduced the proof to Hölder's classical result. Later Krantz, Luce, Suppes, and Tukey (1971, p. 200) showed that bisymmetric operations, which are the source of averaging representations as in utility theory, could in turn be reduced to conjoint measurement.

I assume that Michell thinks that these historical distortions are justified by the fact that Tukey and I addressed our work to psychologists, that conjoint measurement is fairly simple, and that it, rather than the averaging representation of utility theory, is potentially more important for psychology.

Thus, it was argued that Condition I can be and, Michell and I agree, should be retained; and because Condition 2 is greatly broadened beyond extensive measurement, the truth of Condition 3 is irrelevant. Psychological measurement is not as easily ruled out as the British physicists had thought. What, of course, is called for are difficult, detailed empirical studies to justify specific cases of quantification. Michell takes most of the psychological measurement community to task for failing to do so, or for even acknowledging the need to do so. Examples of major exceptions are contributions to psychophysics by Falmagne (1985), Krantz (1975a, 1975b), and Narens (1994, 1996) and the half century of detailed axiomatizations of variants of utility. For reasons that escape me, the utility work seems to be largely dismissed by Michell and most psychologists despite being examples of measurement of what is clearly a psychological attribute, indeed, in many ways, a psychophysical one (Luce, 2000). Most psychologists seem to restrict measurement concerns to classical psychophysical and multidimensional scaling and various types of intelligence and ability testing.

The introduction of conjoint measurement had another benefit. It was shown that if a certain type of distribution law relates a conjoint structure to an extensive operation on one of its factors, then the additive measure and the conjoint one on the same factor must be related as a power function (Krantz et al., see Chapter 10; Luce, Krantz, Suppes, & Tversky, 1990, Chapter 22; Luce & Narens, 1985; Narens, 1976). This became a fundamental account of derived measurement and it led to an axiomatization of the structure of physical quantities, which plays an essential role in what is called dimensional analysis (Luce, 1971). Moreover, the concept of "meaningfulness" of propositions was formulated within a measurement structure as an invariance property. It encompassed not only the concept of dimensional invariance (Luce, 1978), but also Klein's (1872) notion of a geometric object and clarified Stevens's discussion of permissible statistics as a function of scale type (Luce et al., 1990). Although I initiated this work, it has been greatly extended in a variety of ways by Narens. A good deal of this recent work is treated relatively lightly by Michell, despite the fact it is highly relevant to psychological issues.

For example, Narens (1981a, 1981b) began addressing the question why Stevens's classification of scales into types had 0, 1,2, and infinite degrees of freedom (corresponding to absolute, ratio, interval, and ordinal scales), but nothing in between, such as 3 degrees of freedom. The key is properties of the group of symmetries (called automorphisms by mathematicians) of a structure. The work was completed by Alper (1987), who showed that on a continuum the only other cases that can possibly

arise with finite degrees of freedom greater than zero are ones lying between ratio and interval. Moreover, it soon became clear that all sorts of nonadditive measurement is possible based on the following discoveries. Many potentially important nonadditive structures have the a priori, surprising property that the symmetry group is itself additive and, therefore, quantitative. Moreover, often such a structure can be mapped isomorphically onto its own symmetry group and thus into the real numbers, therefore, providing a numerical measurement representation of the structure. This discovery has thus far led to few applications (e.g., Narens, 1996), but it means that there is remarkable potential for nonadditive measurement beyond the simpler additive cases we knew about earlier. To my regret Michell chose not to cover these developments of the past 15 years, which have been spearheaded by Narens, at the same level of detail as the earlier work (p. 208). In my view, they need to be more widely known.

Michell's analysis ends in debate with Cliff's (1992, p. 211-216) contention that axiomatic measurement theory is a revolution that never happened. He argues that Cliff and most of the field continue to miss the point that for the most part quantification, rather than tested, has only been presumed, and he casts substantial doubt on all of Cliff's attempts to explain why the revolution has not happened.

The reader should be warned that Michell, in a familiar British tradition, does not pull his punches. An example: “Once this confusion [of the process of measurement with quantification] is exposed, Stevens's definition of measurement is revealed for the charade it is” (p. 177). Some may find his bluntness disrespectful, but I would urge that readers ignore that aspect and consider carefully the substance of his message. It is an important one if psychological measurement is ever to gain a solid scientific (in contrast to applied) basis and thereby gain the grudging respect of other sciences.

References

Alper, T. M.. (1987). A classification of all order-preserving homeomorphism groups of the reals that satisfy finite uniqueness. *Journal of Mathematical Psychology*, *31*, 135-154. PsycINFO

Holman, E.. (1971). A note on additive conjoint measurement. *Journal of Mathematical Psychology*, *8*, 489-494. PsycINFO

Falmagne, J.-C.. (1985). *Elements of psychophysical theory*. New York: Oxford University Press.

Klein, F.. (1872). *Vergleichende Betrachtungen $über neuere geometrische Vorschungen* [A comparative

review of recent research in geometry]. Entrance Program for the Faculty of Philosophy and the Senate of the University of Erlangen, Erlangen, Germany.

Krantz, D. H.. (1975a). Color measurement and color theory: I. Representation for Grassman structures. *Journal of Mathematical Psychology*, *12*, 283-303. PsycINFO

Krantz, D. H.. (1975b). Color measurement and color theory: II. Opponent colors theory. *Journal of Mathematical Psychology*, *12*, 304-327. PsycINFO

Luce, R. D.. (1971). Similar systems and dimensionally invariant laws. *Philosophy of Science*, *38*, 157-169.

Luce, R. D.. (1978). Dimensionally invariant laws correspond to meaningful qualitative relations. *Philosophy of Science*, *45*, 1-16.

Luce, R. D.. (2000). *Utility of gains and losses: Measurement-theoretical and experimental approaches*. Mahwah, NJ: Erlbaum.

Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A.. (1990). *Foundations of measurement* (Vol. 3) San Diego: Academic Press.

Luce, R. D., & Narens, L.. (1985). Classification of concatenation structures by scale type. *Journal of Mathematical Psychology*, *29*, 1-72. PsycINFO

Narens, L.. (1976). Utility-uncertainty tradeoff structures. *Journal of Mathematical Psychology*, *13*, 296-322. PsycINFO

Narens, L.. (1981a). A general theory of ratio scalability with remarks about the measurement-theoretic concept of meaningfulness. *Theory and Decision*, *13*, 1-70.

Narens, L.. (1981b). On the scales of measurement. *Journal of Mathematical Psychology*, *24*, 249-275. PsycINFO

Narens, L.. (1994). The measurement theory of dense threshold structures. *Journal of Mathematical Psychology*, *38*, 301-321. PsycINFO

Narens, L.. (1996). A theory of magnitude estimation. *Journal of Mathematical Psychology*, *40*, 109-129. PsycINFO Article

Savage, L. J.. (1954). *Foundations of statistics*. New York: Wiley.

von Neumann, J., & Morgenstern, O.. (1947). *The theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

---

[1] *The only references given in the reference list are to items not listed in Michell's book. Those in his book are identified by including a relevant page number or numbers.*

JOEL MICHELL, Department of Psychology, University of Sydney, Sydney, New South Wales, Australia.

R. DUNCAN LUCE, Institute for Mathematical Behavioral Sciences, University of California, Irvine, California 92697. Electronic mail may be sent to rdluce@uci.edu.

PsycCRITIQUES

0010-7549

Previously published in *Contemporary Psychology: APA Review of Books*, October 2000, Vol. 45, No. 5, 499-502