Comment

Theory & Psychology

Conjoint measurement underdone: Comment on Günter Trendler (2019)

Theory & Psychology
2019, Vol. 29(1) 138–143

DOI: 10.1177/0959354318814962
journals.sagepub.com/home/tap
SAGE

Joel Michell
University of Sydney

Abstract

Trendler's (2019) critique of conjoint measurement fails because he neglects to distinguish standard sequences (human constructions) from series of equal magnitudes (features of quantitative structures). The latter, not the former, is presumed in conjoint measurement. Furthermore, in so far as some mental tests use humans as measuring instruments, the only questionable assumption involved is that the relevant psychological attributes are quantitative, and that assumption is potentially testable using conjoint measurement. Finally, contrary to Trendler, psychological phenomena can be captured and the structure of psychological attributes investigated using conjoint measurement.

Keywords

conjoint measurement, psychometrics, standard sequences

The theory of conjoint measurement applies when one attribute, A, relates to two others, X and Y, and it describes ordinal conditions upon levels of A necessary and sufficient for (a) all three attributes to be quantitative (i.e., to possess additive as well as ordinal structure) and (b) for A to be a non-interactive function of X and Y (i.e., $A = f(X + Y)$, where f is increasing monotonic). If data satisfy these conditions, they support (a) and (b). If (a) and (b) are accepted, the data, being ordinal, entail a system of linear inequalities and numerical solutions may be inferred and taken as estimates of magnitudes. In this way, conjoint measurement can be used to test for quantitative structure and, if that is successful, to measure attributes.

Trendler, however, sees conjoint measurement as "practically useless" (2019, p. 100), an assessment falsified by numerous counterexamples in which it is used to test theories proposed to sustain psychological measurement, for example, that of Luce and Steingrimsson (2011; Luce being the psychologist who jointly introduced conjoint

Corresponding author:
Joel Michell, University of Sydney, Sydney, NSW 2006, Australia.
Email: joelm@psych.usyd.edu.au

measurement into psychology; Luce & Tukey, 1964). Hence, Trendler's charge that the "applicability of the axioms of conjoint measurement was called into question by Luce (2011) himself" (p. 101) would, if true, be significant, but, in fact, Luce (2011) makes no mention of "conjoint measurement" and Trendler's attack on conjoint measurement is equally empty.

The error in Trendler's argument is this: his claim that as a method of measurement, solving systems of inequalities presupposes the existence of standard sequences is false and, consequently, his argument collapses. Trendler adopts the concept of a standard sequence from Krantz, Luce, Suppes, and Tversky (1971) and these authors note, "The two best known examples of standard sequences are sets of standard weights, which are part of most laboratory balances, and rulers, which are marked in multiples of some amount, such as a millimetre" (p. 106). The "weights" referred to are the engineered, metal objects that are placed on a balance pan to match the weight of some nominated object and "rulers" are likewise manufactured objects. That is, standard sequences are constructed artefacts. This is so not only for extensive measurement but also for standard sequences constructed via difference and conjoint measurement.

There are limits to our capacity to engineer standard sequences because for some psychological attributes, as Trendler indicates, we lack the technical know-how. Take social attitudes. For argument's sake, suppose that the favourability of attitudes is a quantitative attribute (as Thurstone, 1928, proposed) and that when a person, i, prefers one statement, x, over another, y (each statement expressing different attitudes towards the same issue), the absolute difference between the favourability of i's own attitude, A_i, and the favourability of statement x, A_x, is less than the absolute difference between the favourability of A_i and the favourability of statement y, A_y (as Coombs, 1950, proposed). That is, if person i prefers x to y then $|A_i - A_x| < |A_i - A_y|$, which is an inequality. It further follows that across a range of people and statements expressing attitudes of different degrees of favourability, people's preferences will entail a system of inequalities. We lack the means to construct standard sequences of attitude statements (i.e., sets of statements equally spaced in the degrees of favourability expressed). However, we do have the mathematical means to find solutions to systems of mutually consistent inequalities. Therefore, contrary to Trendler, locating and solving inequalities does not presume construction of standard sequences.

Anticipating such an argument, Trendler (2019) thinks that because solving such inequalities translates empirical into numerical relations and "the translation is legitimate only under the assumption (or hypothesis) that the investigated attributes are decomposable into standard sequences (i.e., a series of equal magnitudes of quantity) or, in short, that they are quantities" (p. 105), standard sequences are indirectly presupposed. This confuses *standard sequences* with *series of equal magnitudes*.

No matter how precisely engineered a standard sequence is, it only ever approximates a series of equal magnitudes because all engineering, no matter how precise, tolerates errors of some size. The concept of *a series of equal magnitudes* is one thing, that of *a standard sequence* another, and each plays a different role in measurement theory as explained below.

Heath (1908) noted, following De Morgan (1836), that the concept of a series of equal magnitudes underpins Definition 5, Book V of Euclid's *Elements*, which specifies what it is for two pairs of magnitudes to stand in the same ratio (anticipating Dedekind's, 1872, treatment of real numbers). Hölder (1901), drawing upon Euclid and Dedekind, employed the concept of a series of equal magnitudes in his exposition of measurement theory. Whitehead and Russell (1913) also used it.

The idea is: the concept of an unbounded, continuous quantity (Hölder, 1901) entails that corresponding to each magnitude, *a*, of an unbounded continuous quantitative attribute, there is the series of equal magnitudes *a, 2a, 3a, ..., etc.*[1] (in general, *na*, for each natural number *n*). Equally, for *b*, another magnitude of the same attribute, there is *b, 2b, 3b, ..., etc.* (in general, *mb*). The ratio of *a* to *b* is a relation between these series: each *na* is less than, equal to, or greater than each *mb* (for all natural numbers *n* and *m*); if the two series coincide and for some *n* and m, $na = mb$, the ratio of *a* to *b* is rational; otherwise it is irrational and defined, following Dedekind, as the *cut* between all $m/n < a/b$ and all $m/n > a/b$. The concept of a series of equal magnitudes equates ratios of magnitudes with real numbers. The thing to note is that *standard sequences* (because they are human constructions) fail this task, while *series of equal magnitudes* (because they are a feature of quantitative structure) succeed.

N. R. Campbell introduced "standard series" (1920, p. 280) into measurement theory and saw them as human constructions, which is why they figured in his discussion of errors of measurement. Krantz et al. note that their concept of a *standard sequence* is Campbell's *standard series* by another name:

> In the practice of measurement, it is usually convenient to agree upon some acceptable degree of precision to be achieved. This degree of precision can then be used as a basis for constructing what has been called a *standard series* (Campbell, 1920) or what in our discussion of the Archimedean axiom we have preferred to call a *standard sequence*. This allows us, using few qualitative observations, to assign to objects a numerical measure that approximates the extensive measure to within the preassigned limit of precision. (1971, p. 105, emphasis in original)

Unfortunately, the discussion of Krantz et al. (1971) lends itself to confusing these concepts. For example, regarding the Archimedean axiom, they say that such an axiom, while "trivially true in a finite structure" (p. 26) is "a necessary axiom" (p. 25) for infinite structures, but note, "it is unclear what constitutes empirical evidence against it since it may not be possible to exhibit an infinite standard sequence" (p. 26). Here the authors have glided effortlessly from the concept of a standard sequence (a human construction and therefore finite) to that of a series of equal magnitudes (part of the concept of a quantitative attribute and potentially infinite). The distinction was not germane to their concerns, but it is crucial to Trendler's, for he argues that measurement via systems of inequalities necessarily implicates standard sequences when actually all that such inequalities presume is quantitative structure, with its implication of series of equal magnitudes. This affects Trendler's (2019) thesis that conjoint measurement is useless because it means that in contexts where it is not technically possible to construct standard sequences, measurement via inequalities may still be possible. Thus, conjoint measurement remains a potential measurement strategy.

However, Trendler attempts to foreclose this option by another ploy: "psychological tests usually employ some form of a psychometric scale (e.g., the Likert scale) to collect data" (p. 114), which means, he says, "that the unit 'test subject-rating scale' is quite literally regarded as an embodiment of a measurement instrument" (p. 114) and regarding humans as measurement instruments is he asserts "an unrealistic hypothesis" (p. 116). Even when participants are asked which of two attitude statements, *x* or *y*, they prefer, he maintains, "it is implicitly assumed that the responses are determined by consulting an 'internal' scale on which the distance of *x* and *y* to the so-called point of maximum preference is estimated" (p. 115).[2]

Overlooking the fact that the vast majority of psychological tests (e.g., ability and achievement tests) neither employ Likert scales nor require participants to make metric judgments, regarding humans as measurement instruments is not itself a problem. The practice works and is as old as measurement itself (e.g., Farr, 2010): the scientific saint, Galileo often used himself as a measurement instrument to estimate elapsed time in some of his experiments;[3] and Dawes (1977) showed that ratings of height are "very good estimates of true physical height" (p. 267). The issue is not whether humans can behave as measuring instruments when required to, but whether, in psychological applications, the mental attributes they are invited to make metrical judgments about are actually quantitative. If they are not, the numerical information ratings deliver is pseudo-metrical, thereby inviting invalid inferences. It is here that the theory of conjoint measurement is relevant: the double cancellation condition of conjoint measurement tests for additive structure in the relevant attributes and because such tests can be carried out without constructing standard sequences, it is possible using conjoint measurement to test the hypothesis that psychological attributes are quantitative (e.g., Michell, 1994).

However, Trendler (2019) has a final card up his sleeve:

> the testing of the quantitative hypothesis only makes sense under circumstances where it is possible to control systematic disturbances (Trendler, 2009). The problem is invariably solved through the construction of experimental apparatus. This method is inapplicable in psychology, since—contrary to physical processes—it is not possible to capture psychological processes in experimental apparatus, devices, or machines as would be required in order to control systematic disturbances. (p. 120)

His objection is that psychologists cannot capture the mental phenomena they aspire to measure using experimental apparatus and, so, they cannot measure these phenomena. On the contrary, however, mental phenomena are captured via experimental apparatus (viz. psychological test items), not with the precision physics displays, but with a useful degree of verisimilitude nonetheless.

Take ability test items: it is possible to specify the knowledge, skills, and strategies a person must possess in order to produce the correct answer to each item and when a response is correct, *ceteris paribus*, it is not unreasonable to infer that the respondent possesses those cognitive resources. To this extent, items in a test pin down mental phenomena and these constitute the mental attribute assessed by such items.

Far from psychological attributes being, as Trendler insists, "non-observable," in ability testing they are on display in the sets of cognitive resources required for correct answers. Furthermore, the structure of the relevant mental attribute is observable in the relationships between those sets of cognitive resources. It is because these relationships involve qualitative differences as we move from easier to more difficult items that the structure of such attributes is incompatible with quantitative structure (Michell, 2014). Conjoint measurement provides a means of testing whether quantitative structure hides behind the manifest non-quantitative structure of the attributes tests assess. In the first instance, it serves the scientific task of quantification (Michell, 1997).

Psychometricians imposed "measurement" upon mental phenomena because their model of successful science, physics, made them vulnerable to the demands of the quantitative imperative (science necessitates measurement). Trendler denies measurement to psychology because his model of successful quantification, physics again, makes him vulnerable to the demands of the Procrustean imperative (one size fits all). However, rejecting imperatives and endorsing, instead, empiricism: psychometricians provide no compelling evidence that mental attributes are measurable; and Trendler, no compelling reason to think conjoint measurement is not an avenue still worth exploring to resolve the issue empirically.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Note that *2a* is not a physical construction of *a* with itself, but is a magnitude consisting entirely of two discrete parts, each identical to *a*; similarly, *3a* is a magnitude consisting entirely of two discrete parts, *2a* and *a*; and so on. That is "+" is understood not as an operation of concatenation (as in extensive measurement), but a relation of composition and, so, applies to all quantities.
2. For the record, Coombs' (1950) theory does not require that the person "consult" an internal scale but only that order relations between the favourability of the person's attitude and the favourability of the relevant attitude statements determine the person's preference.
3. According to Drake (1986), in some of his experiments, Galileo would count his "musical beats of about a half-second each" (p. 303) to measure time elapsed.

References

Campbell, N. R. (1920). *Physics the elements*. Cambridge, UK: Cambridge University Press.

Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, *57*, 145–158.

Dawes, R. M. (1977). Suppose we measured height with rating scales instead of rulers. *Applied Psychological Measurement*, *1*, 267–273.

Dedekind, R. (1872). *Stetigkeit und irrationale Zahlen* [Continuity and irrational numbers]. Braunschweig, Germany: Vieweg.

De Morgan, A. (1836). *The connexion of number and magnitude*. London, UK: Taylor & Walton.

Drake, S. (1986). Galileo's physical measurements. *American Journal of Physics*, *54*, 302–306.

Farr, H. (2010). Measurement in navigation: Conceiving distance and time in the Neolithic. In I. Morley & C. Renfrew (Eds.), *The archaeology of measurement* (pp. 19–26). Cambridge, UK: Cambridge University Press.

Heath, T. L. (1908). *The thirteen books of Euclid's elements: Vol. 2*. Cambridge, UK: Cambridge University Press.

Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Mass [Axioms of quantity and the theory of measurement]. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, *53*, 1–46.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement: Vol. 1*. New York, NY: Academic Press.

Luce, R. D. (2011, November 15). Inherent individual differences in utility. *Frontiers in Psychology*, *2*(297). doi:10.3389/fpsyg.2011.00297

Luce, R. D., & Steingrimsson, R. (2011). Theory and tests of the conjoint commutativity axiom for additive conjoint measurement. *Journal of Mathematical Psychology*, *55*, 379–385.

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

Michell, J. (1994). Measuring dimensions of belief by unidimensional unfolding. *Journal of Mathematical Psychology*, *6*, 7–24.

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

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

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*, 529–554.

Trendler, G. (2009). Measurement theory, psychology and the revolution that cannot happen. *Theory & Psychology*, *19*, 579–599. doi: 10.1177/0959354309341926

Trendler, G. (2019). Conjoint measurement undone. *Theory & Psychology*, *29*, 100–128. doi: 10.1177/0959354318788729

Whitehead, A. N., & Russell, B. (1913). *Principia mathematica: Vol. 3*. Cambridge, UK: Cambridge University Press.

Author biography

Joel Michell is an Honorary Associate Professor in the School of Psychology at the University of Sydney, where he taught psychometrics and the history and philosophy of psychology until his retirement in 2004. He continues to investigate issues relating to the history and philosophy of measurement in psychology.