

University of Wollongong
Research Online

Faculty of Arts - Papers (Archive)


Faculty of Arts, Social Sciences & Humanities

1-1-2008

I can't believe it's not measurement: the legacy of operationism in social-scientific uses of numbers

George Matheson
University of Wollongong, gmath@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/artspapers>

 Part of the [Arts and Humanities Commons](#), and the [Social and Behavioral Sciences Commons](#)

Recommended Citation

Matheson, George, I can't believe it's not measurement: the legacy of operationism in social-scientific uses of numbers 2008, [9].
<https://ro.uow.edu.au/artspapers/861>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

I Can't Believe It's Not Measurement: The Legacy of Operationism in Social-Scientific Uses of Numbers

by

George Matheson PhD

School of Social Sciences, Media and Communication
University of Wollongong

gmath@uow.edu.au
+61 2 4221 3743

Paper presented to TASA 2008 Conference, University of Melbourne

December 2008

I Can't Believe It's Not Measurement: The Legacy of Operationism in Social-Scientific Uses of Numbers

Abstract

What is called measurement in human sciences such as sociology is different from other uses of the term, embracing not only quantification in the strict sense, but also all kinds of scaling, ranking and even classification *per se*. This paper considers such habits as a legacy of the 'Operational' measurement theory of S. S. Stevens, wherein science meant measurement, but concepts (e.g., measurement) meant whatever we all agreed they did. Coupled with a broader cultural tendency to privilege mind over matter, this has led to great efforts to quantify the intangible, possibly at the expense of sociologically-relevant material factors which better lend themselves to it.

The Meanings of Measurement

The human sciences, sociology included, have a conception of measurement that is all their own. Simply put, we call things measures that others would not. To understand why this is so requires an excavation in the archaeology of knowledge, in the process revealing something about how social scientists and related professionals conceive of their objects of inquiry and how to study them. First things first, though: what is measurement?

In the natural sciences and everyday life alike, to measure is to 'ascertain the size, amount, or degree of (something) by comparison with a standard unit or an object of known size' (*Concise Oxford*). A number is a measurement if it is expressing a quantity relative to some unit and we can perform ordinary arithmetic operations upon it. We measure our body weight in kilograms, the length of our room for furnishing purposes in paces, the electromotive force of our circuit in volts. In sociology and cognate fields, on the other hand, almost any application of numbers seems to be regarded as measurement: index numbers with an arbitrary zero, rank-orderings, even classifications with unordered but numbered categories. One simple test might be whether one can speak meaningfully of half or twice as much. Someone could have twice (half) as many houses or children or twice (half) as much income in dollars or schooling in years, a village or nation-state could have twice or half as many people. Can one have twice or half as much occupational status or job satisfaction? What about twice or half as much social class? In the former case, we are presumably dealing with

not a measure in the strict sense, but an indicator, in the latter case with a classification. Yet in the textbooks of social science research methods, and in the practice of the trade, it is remarkable that all these things are discussed under the rubric of measurement.

The first of a convenience sample of social science methods texts (i.e., those on my bookshelf) has a chapter on ‘Qualitative and Quantitative Measurement’ (Neuman 2007: 109-139). Another explains that measurement ‘involves categorising and/or assigning values to concepts, and is diverse in nature and level of operation’, and again goes on to discuss qualitative as well as quantitative measurement, while recognizing the potentially controversial nature of this usage (Sarantakos 2005: 73-74). The seventh edition of one classic textbook explains that ‘by measurement we mean rules for assigning symbols, usually numbers, to objects, usually people’ (Hoyle *et al* 2002: 96). Yet another includes a half-page text box setting out the difference between measures *per se* and indicators, which is thereafter largely forgotten (Bryman 2004: 67). Assigning numbers or other symbols to things, classifying qualitative attributes: all of this seems far removed from the determination of quantities, of how much or how many. So how did the practice of including everything under ‘measurement’ originate?

Operationism and the Stevnsian Theory of Scales

The expansion of the idea of measurement from what it means in physics to what it means in social research was associated historically with the early twentieth-century movement for the ‘unification of science’, with its professedly anti-metaphysical philosophy of Logical Positivism, also known as Operationism, especially in its social-scientific guise.

‘Operationism’ was defined by Stanley Smith (‘Smitty’) Stevens, in his entry in Dagobert D. Runes’ 1942 *Dictionary of Philosophy*, as ‘the doctrine that the meaning of a concept is given by a set of operations’. On this basis, something simply *is* whatever we have decided to operationally define it as, e.g., E. G. Boring’s famous definition of ‘intelligence’ as ‘what intelligence tests test’.

S. S. Stevens (1906-1973) was the person principally responsible for the human sciences’ wide-ranging concept of measurement, as in the introductory texts mentioned above. His (in)famous definition of ‘measurement’ as ‘the assignment of numerals to objects or events according to rules’ (Stevens 1946, 1951) has been echoed in countless textbooks and classrooms over the past half-century or so. Many will recall his model of ‘levels of measurement,’ in which measuring is not all of a piece, but takes place to varying degrees.

We go from nominal to ordinal to interval to ratio scales as we acquire, cumulatively, operations for the determination of identity and difference, rank order, equality of intervals and equality of ratios. Ascending the hierarchy, each scale type has a smaller range of information-preserving transformations and a wider range of statistics that can be sensibly applied, as well as its own formal definition in terms of group theory. For now, the important points are two. Firstly, *all* of these scale types constitute measurement, as Stevens understood the term. Secondly, that does not mean he thought all were as good as each other. Stevens, for all the apparent ecumenism of his approach, retains in his distinction between ratio scales and the rest the same distinction that others would have made between ‘true’ physical measurement and other uses of numbers for indicating or labelling (see Matheson 2006: 67).

Smitty Stevens was Harvard’s and the world’s first Professor of Psychophysics (see Miller 1975 for a biography). The director of the Psycho-Acoustic Laboratory during World War II and the formulator of the psycho-physical power law, he liked his science ‘hard’, his world invariant and deterministic. Stevens’ primary inspiration in the philosophy of science was his Harvard colleague, the physicist Percy Bridgman (1927), who sought to rethink things in the wake of Einstein, placed primary emphasis on operations and disdained metaphysics in all its forms.

The theory of measurement only appeared on Stevens’ radar in the late 1930s, when a committee of the British Association for the Advancement of Science was debating whether or not it was, in principle, possible to measure sensation: could one say that one light was subjectively twice as bright as another, a weight twice as heavy, a noise twice as loud? In response to these inconclusive debates, and with the war intervening, Stevens published ‘On the Theory of Scales of Measurement’, setting out both his operational definition and his four ‘levels’, as lead article in *Science* on 7 June 1946. Essentially the same case was made in his introductory chapter to the *Handbook of Experimental Psychology* (Stevens 1951). Given that Stevens had already developed a unit of subjective loudness, the sone, it comes as no surprise that he thought measurement of such phenomena to be perfectly doable, and rejected out of hand any suggestion that the lack of a defined operation for the addition of sensory intensities made the whole business a sham. Interestingly, he seems to have viewed his own measurements of psycho-physical magnitudes as the only ratio-level, physical measures in the psychological or sociological literature, everything else being at a lower level in the hierarchy of scales.

It is not to be thought that the theory of measurement in sociology, psychology, anthropology and the like has remained frozen in the 'Operational' perspective of the 1940s. If anything, that theory has given way, at least in principle, to something more akin to a representational approach: there is a real world that our measures are imperfectly trying to capture, not an empirical world co-extensive with our operations upon it (see e.g., Sarle 1997). Yet in practice, the song remains the same: measurement is about putting numbers on things according to some sort of rules, and it is not limited to physical or ratio measurement, but includes rankings and interval scales as well, if not nominal or 'qualitative measurement'. The practical if not theoretical aspects of the Stevens approach are alive and well, and living in the social sciences.

One thing that has not happened in discussions of counting and measuring in social research is any move towards what Joel Michell (1999, 2005) calls the 'Classical' theory of measurement. By that is meant the ideas, common outside the human sciences, that an attribute can be measured only if it is quantitative in character, that whether it is or not is an empirical question, and that the role of scientific inquiry of whatever kind is to discover its magnitude.

In fact, Smitty Stevens' definition of measurement was controversial enough in its own time. In physics, Campbell (1920: esp. chap. 10) concurred that measurement meant the assignment of numerals to represent properties, but also insisted that such assignments had to reflect both rank order and either a physically defined operation of addition or indirect derivation from something that had one. In psychology, Johnson (1936) identified nominal, ordinal and cardinal uses of numbers, in the course of arguing that only the last of these is measurement and that psychology did not do it.

The Bigger Picture: From Mind to Matter?

The question remains of how any of this matters. If social researchers wish to use the word 'measurement' differently from other specialists, is it a problem? Some would have thought otherwise, S.S. Stevens among them. To Stevens, 'the naming of things is an arbitrary business' (1951: 26). From his viewpoint, firmly anchored in 1940s logical positivism, definitions of terms were matters of social convention among relevant professionals, so measurement was whatever scientists deigned to call such. If psychologists and those in allied trades were to apply the word to things that no self-respecting physicist or chemist would so describe, it was practically necessary as well as simply polite to include the social scientists'

usages as well. Even if one does not buy into the anti-metaphysical metaphysics of those times, however, a dispute over the legitimacy of calling it measurement whenever categories are numbered or finishing places ranked may still seem a storm in a teacup. Does a certain conceptual looseness in what we think of as measuring people and society make any practical difference?

Perhaps a more constructive question would be whether there are any important differences between the sociologically relevant variables we can measure – i.e., in the classical, physical sense – and those for which we settle for quasi-quantitative indicators or indices. Adopting the simple criterion of whether one can sensibly determine ratios in a given numeric attribute makes it (usually) easy enough to judge what falls into one category or the other. And even allowing for a degree of unconscious confirmation bias in selection of examples, a pattern emerges in the sorts of attributes amenable to measurement in the strict sense versus classifying, rank-ordering or using index numbers from an arbitrary origin-point.

When the style of work is social-psychological and the chosen indicators meant to represent but the outward visible signs of a putative tendency or disposition internal to the individual, the measurement goal tends to be ‘ordinal’ or ‘interval’ scaling. Often, as in the treatment of five- or seven-point Likert-type questionnaire items, data not obviously more than ordinal acquire equal intervals by postulation. It makes the sums easier, the standard statistical repertoire of the social scientist coming into play at that point. Thus we encounter ordinal-tending-to-interval ‘measures’ of occupational prestige, social distance, attitudes, beliefs, values, opinions and other inner conditions assumed to be manifesting themselves in the answers people give to survey questions. Factor or principal components analysis is used to seek the traces of an underlying attitude or value in a multitude of individual agree-disagree items, just as the psychometricians traditionally hunted for the ‘g’ factor of so-called general intelligence. No-one particularly concerns themselves with what might be meant by twice or half as much scholastic aptitude, political conservatism or religiosity.

In contrast, those measures employed in the human sciences which would fit the description in the world outside our field, i.e., constitute answers to questions of how much or how many relative to a specified unit, take a different character. Time use surveys and family budget studies seek, however imperfectly, to put figures to quantities of time and money spent in various ways. Research in demography and population ecology likewise draws upon observations of real, countable units in the form of the individual humans found in vital records, surveys and censuses. In each of these areas of study, the attributes being measured

(however crudely) are properly quantitative; one can perform any arithmetical operations upon them, and must be wary of applying any transformations to the raw numbers. Moreover, the interesting part is that these quantitative characteristics are all about matter in motion, patterns of activity, behaviour as opposed to attitudes, doings rather than stated dispositions. They refer to an inter-subjective material reality in which human beings are living things in a physical world, living and dying and reproducing, expending and exchanging money, time and things.

So, if you like, the attributes of people or things to which we sociologists seek to give a numeric expression can be arrayed along the same dimension of mind to matter, spirit to flesh, ideal to material, which characterizes Western (or even Indo-European) thought in general. Thus the properly quantitative would be found closer to the 'material' end of things, and the substitutes for it lower down in the hierarchy of so-called measures (i.e., Stevens' interval and ordinal scales) might be expected as we come nearer to the purely 'mental'.

This hypothesised correspondence in social science methods between matter and measurement on the one hand and mind and not-quite-measurement on the other might be thought a little too neat, too convenient. Surely any qualitative characteristic, anything either present or absent, can turn quantitative when we know its incidence by category of the relevant population or its frequency over time? Yet by the very act of aggregation we are counting heads or measuring time among events, and back in the realm of the physical. More controversial is the treatment of socially-relevant attributes of individual subjects, respondents or informants: are attitudes and beliefs directly quantifiable like attributes and behaviours, each expressible as a ratio of some unit? Should we attempt this? Why or why not?

Ultimately, whether one seeks quantitative expressions of internal mental states on all fours with counts and measures of the inter-subjectively observable outside is less a methodological matter than a question of how one understands sociology and the other human sciences, their nature and role(s). If one assumes, as did Smitty Stevens and the social-scientific logical positivists of the twentieth century, that who says science says measurement, then all observation and experiment must lead to such, and we need to put numbers on anything we think we know something about, *pace* Lord Kelvin. Add to this a Cartesian dichotomy between body and mind, together with further, possibly religion-derived, assumptions that place the latter of the two in the driver's seat, and we have to measure attitudes, values *et cetera* in whatever way we can, in order to explain actual behaviour.

Conversely, if humans are symbolically-communicating living organisms surviving in a material universe, then what we say is only part of what we do, and privileging the current contents of consciousness gets us nowhere. When it comes to the study of symbolic culture, Ann Swidler (2001) has written of the limitations of research that focuses on the imputed deep attitudes, values or beliefs for which fixed-choice survey responses or similar indicators are taken as proxies. Instead, she advocates focussing on the overt patterns of rituals and relations which structure the symbolic forms in social life. Going beyond the symbol sphere as such, every person and society is also operating in a world of practical material problems of production, reproduction and destruction, which cannot always be discursively spirited away. At the level of demography, geography and socio-economics, measurement in the strict sense of the word remains an issue, with the most cursory inspection of human ecology at present indicating that ‘how much?’ and ‘how many?’ are crucial questions when it comes to the sociology (and politics) of population, land, energy and resources.

References

- Bridgman, P. W. (1927) *The Logic of Modern Physics*. New York: Macmillan.
- Bryman, A. (2004) *Social Research Methods*. Second Edition. Oxford: Oxford University Press.
- Campbell, N. R. (1920) *Physics: The Elements*. Cambridge: Cambridge University Press.
- Hoyle, R. H., M. J. Harris and C. M. Judd (2002) *Research Methods in Social Relations*. Seventh Edition. Fort Worth, TX : Wadsworth
- Johnson, D. (1936) ‘Pseudo-Mathematics in the Social Sciences’, *American Journal of Psychology* 48: 342–51.
- Matheson , G. (2006) ‘Intervals and Ratios: The Invariantive Transformations of Stanley Smith Stevens,’ *History of the Human Sciences* 19(3): 65-81.
- Michell, J. (1999) *Measurement in Psychology: Critical History of a Methodological Concept*. Cambridge: Cambridge University Press.
- Michell, J. (2005) ‘Measurement Theory’, *Encyclopedia of Social Measurement*, vol. 2. Amsterdam: Elsevier.
- Miller, G. A. (1975) ‘Stanley Smith Stevens’, *Biographical Memoirs*, vol. 47. Washington, DC: National Academy of Sciences.

Neuman, W. L. (2007) *Basics of Social Research: Qualitative and Quantitative Approaches*. Second Edition. Boston: Pearson.

Sarantakos, S. (2005) *Social Research*. Third Edition. New York : Palgrave Macmillan.

Sarle, W. S. (1997) *Measurement Theory: Frequently Asked Questions*, version 3, 14 September 1997. URL: <ftp://ftp.sas.com/pub/neural/measurement.html>.

Stevens, S. S. (1942) 'Operationism', in D. D. Runes (ed.) *Dictionary of Philosophy*. New York: Philosophical Library.

Stevens, S. S. (1946) 'On the Theory of Scales of Measurement', *Science* 103: 677–80.

Stevens, S. S. (1951) 'Mathematics, Measurement and Psychophysics', in S. S. Stevens (ed.), *Handbook of Experimental Psychology*. New York: Wiley.

Swidler, A. (2001) 'What Anchors Cultural Practices', in T. R. Schatzki, K. Knorr Cetina and E. von Savigny (eds) *The Practice Turn in Contemporary Theory*. London: Routledge.