

This article was downloaded by: [University of Canterbury Library]

On: 4 April 2011

Access details: Access Details: [subscription number 917001820]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Measurement: Interdisciplinary Research & Perspective

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t775653679>

Invoking Arbitrary Units Is Not a Solution to the Problem of Quantification in the Social Sciences

Paul Barrett^{ab}

^a Advanced Projects R&D, New Zealand ^b Department of Psychology, University of Auckland,

Online publication date: 30 March 2011

To cite this Article Barrett, Paul(2011) 'Invoking Arbitrary Units Is Not a Solution to the Problem of Quantification in the Social Sciences', *Measurement: Interdisciplinary Research & Perspective*, 9: 1, 28 – 31

To link to this Article: DOI: 10.1080/15366367.2011.558783

URL: <http://dx.doi.org/10.1080/15366367.2011.558783>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Invoking Arbitrary Units Is Not a Solution to the Problem of Quantification in the Social Sciences

Paul Barrett

*Advanced Projects R&D, New Zealand
Department of Psychology, University of Auckland*

The article by Stephen Humphry (this issue) is a technical tour de force. At one level, I marvel at the ingenuity and sophisticated logic and argument on display. This is impressive work and thinking whichever way you look at it. However, after twice re-reading the manuscript, the same question arises in my mind: What exactly has any of this to do with the measurement of psychological attributes? Humphry has produced an interesting line of argument and a mathematical derivation that augments the Rasch IRT model so that it can incorporate differential discrimination parameters for sets of questionnaire items composing a “measure” of a latent variable. As part of this derivation, a standard unit of measurement is apparently created for an arbitrary, statistically constructed, latent trait. Within the simulation example he provides, the unit is never even defined; it’s presumably invoked somewhere but apparently cannot be specified as unambiguously as one might specify a physical unit. Even if the data are artificial, surely it is not impossible to show how the data were generated as magnitudes of a particular ability, where those magnitudes were constructed as ratios of a specific standard ability unit.

This is why my reaction to the article is one of puzzlement. If there is one thing we have learned from the work of Schonemann (1994), Michell (1997), and Trendler (2008), it is that empirical experimental manipulations of attributes are required to be undertaken prior to the representation of attribute magnitudes by the real number system, let alone the instantiation of a unit of measurement. Yet, we know such manipulations are all but impossible to imagine for many psychological attributes, including abilities. The idea that a statistical model of questionnaire item responses can somehow reveal quantitative structure for an attribute constructed by that very same statistical model seems unreal to me. Describing the constructed attribute as a “latent variable” is an attempt to avoid admitting the variable is just something constructed from data observations using a particular form of data analysis. There is nothing “latent” about such variables, as Maraun (2007) and Maraun and Halpin (2008) have been stating for years now. Yet, Maraun and colleagues’ work on latent variables is clearly considered irrelevant by Humphry, as the publications are not even acknowledged in the references.

If one looks beyond the restrictive world of the psychometrician working with educational test items and peers for a moment into the world of psychological science, empirical evidence shows that some “latent” traits (variables) really are artifacts of the particular linear methods applied to describe data patterns. Maraun (1997) had already demonstrated to the psychometric community that the Big Five personality trait model is quite arbitrary, merely a function of the linear methodology one used to analyze data. The problem was, nobody wanted to be reminded of that inconvenient fact. For example, within the domain of personality and the Big Five, it has now become apparent that one can make reasonable predictions of behavioral outcomes without invoking a single “Big Five” latent variable measure, using a theory of personality that is tied to biological and cognitive functioning and with no traits involved (Read et al., 2010). Indeed, as they state:

This model is intended as a potential model of the structure of human personality. But by that we do not mean that any specific instantiation of the model will provide a replication of personality structure, such as the Big Five. Instead, we view any specific set of parameters and learning experiences as representing a particular individual or type of individual. The Big Five is a representation of the structure of human personality across a group of people. This structure is not seen for a single person but is rather the result of the covariation among characteristics within a large sample of people. Therefore, if we created a large number of virtual individuals, each with a different random set of parameters, we would expect the resulting patterns of behavior across individuals to give us something like the Big Five. (p. 87)

One also has to question why Humphry would even bother to derive the link between discrimination and a unit of measurement when, as Michell (2004) has stated:

If a person’s correct response to an item depended solely on ability, with no random “error” component involved, one would only learn the ordinal fact that that person’s ability at least matches the difficulty level of the item. Item response modelers derive all quantitative information (as distinct from merely ordinal) from the distributional properties of the random “error” component . . . Here, as elsewhere, psychometricians derive what they want most (measures) from what they know least (the shape of “error”) by presuming to already know it . . . *By any fair-minded reckoning, our state of knowledge at present does not support claims to be able to measure psychological attributes using item response models.* (p. 126)

What is the point of developing a derivation for an arbitrary unit of measurement within a statistical model that cannot, in and of itself, justify a claim for the quantitative measurement of any psychological attributes?

The very real problem we face as psychological scientists is how to conceive of defining any psychological attribute in a clear, technical manner, such that we can propose experimental manipulations that might test our expectations about magnitude relations. We can’t even define the meaning of personality attributes uniquely, as has become increasingly apparent recently within the meta-analysis literature (Barrett & Rolland, 2009; Pace & Brannick, 2010; but see also Woods, 2009). The response may be that abilities are different; here we possess clear technical definitions of abilities such as numerical reasoning, verbal fluency, speed of closure, etc. But do we, to the extent required by the invocation of a standard unit of measurement of each?

To some, the kinds of criticism stated above look nihilistic—a virtual admission that the quantitative measurement of psychological attributes may be illusory for many attributes. However, we can and do approximate a fuzzy-order measurement pretty well, especially with regard to the

assessment of performance. But you soon realize that in our day-to-day interpretation and practical use of psychological measures, most of this “measurement” is expressed in terms of orders and not real numbered continuous magnitudes, largely because we don’t have a clue what the difference means between individuals who score, say, 2.75 and 2.76 on a “latent” IRT variable. Until we can develop theory and experimental manipulations that allow us to test whether such apparent quantitative precision is empirically justified, we have to accept that fitting a certain type of sophisticated statistical model to item response data will remain relevant only to those who are more concerned with fitting data models than with the measurement of substantive psychological attributes.

Seeking to establish quantitative measurement is, I think, primarily a theory-driven exercise. In addition, there has to be a body of empirical evidence that provides unambiguous results that an attribute is varying quantitatively. That evidence will be generated from theories about the attribute and how and why it varies linearly in magnitudes. Once that body of evidence is established, then quantification can proceed.

New methods of data analysis, like Grice’s Observation Oriented Modeling (in press) permit the causal analysis of data without invoking any metric. Deterministic explanatory and actuarial accuracy drives this modeling, not statistical model fit or aggregate probability models of phenomena. But this methodology is one suited to scientific exploration, in contrast to psychometric IRT models that can only be used for describing item response data.

In conclusion, I would have been more sympathetic with Humphry’s work if he had just presented his derivation as something of relevance solely to atheoretical psychometricians/statisticians working with item responses. But, in his abstract, he claims the work is of relevance to measurement in the social sciences. On the contrary. I don’t think it has any relevance.

As I see it, the problem remaining for any social scientist is, not one of developing yet more derivations of existing statistical item response models or even new such models, but one of creating bodies of evidence that demonstrate that a psychological attribute does indeed vary additively. If these bodies of evidence are missing, then we must continue to explore and make careful observations and, where possible, manipulate features of phenomena and attributes, but without this continuing pretence of an artificial precision accorded by so-called “measurement models” within “quantitative psychology.” And we continue like this until such time as the body of observational evidence either invites obvious and unambiguous quantification or theory-related causal explanations of our observations show it is simply not possible in principle.

REFERENCES

- Barrett, P. T., & Rolland, J.-P. (2009). *The meta-analytic correlation between the big five personality constructs of emotional stability and conscientiousness: Something is not quite right in the woodshed*. Retrieved from <http://www.pbarrett.net/stratpapers/metacorr.pdf>
- Grice, J. (in press). *Observation oriented modeling: An analysis of cause in the behavioral sciences*. New York: Elsevier.
- Maraun, M. (2007). *Myths and confusions: Psychometrics and the latent variable model*. Retrieved from <http://www.sfu.ca/~maraun/Mikes%20page-%20Myths%20and%20Confusions.html>
- Maraun, M. D. (1997). Appearance and reality: Is the big five the structure of trait descriptors? *Personality and Individual Differences*, 22(5) 629–647.
- Maraun, M. D., & Halpin, P. F. (2008). Manifest and latent variables. *Measurement: Interdisciplinary Research & Perspective*, 6(1 & 2), 113–117.

- Michell, J. (1997). Quantitative science and the definition of measurement in Psychology. *British Journal of Psychology*, 88(3) 355–383.
- Michell, J. (2004). Item response models, pathological science, and the shape of error. *Theory and Psychology*, 14(1) 121–129.
- Pace, V. L., & Brannick, M. T. (2010). How similar are personality scales of the “same” construct? A meta-analytic investigation. *Personality and Individual Differences*, 49(7) 669–676.
- Read, S. J., Monroe, B. M., Brownstein, A. L., Yang, Y., Chopra, G., & Miller, L. C. (2010). A neural network model of the structure and dynamics of human personality. *Psychological Review*, 117(1), 61–92.
- Schonemann, P. (1994). Measurement: The reasonable ineffectiveness of mathematics in the social sciences. In I. Borg & P. Mohler (Eds.). *Trends and perspectives in empirical social research* (pp. 149–160). Berlin, Germany: Walter de Gruyter.
- Trendler, G. (2009). Measurement theory, psychology and the revolution that cannot happen. *Theory and Psychology*, 19(5), 579–599.
- Woods, S. A. (2009). The structure and comparative validities of six personality inventories. Paper presented at the British Psychological Society, Division of Occupational Psychology, Annual Conference. Retrieved from [http://www.bps.org.uk/document-download-area/document-download\\$.cfm?file_uid=6731AAC2-E2A5-EF8E-94E2-FE12DB903247&ext=pdf](http://www.bps.org.uk/document-download-area/document-download$.cfm?file_uid=6731AAC2-E2A5-EF8E-94E2-FE12DB903247&ext=pdf)