

The Fault is Not in Ourselves, but in Our Methods: Comment on Schwarz

Henderikus J. Stam

Published online: 4 September 2010
© Springer Science+Business Media, LLC 2010

Abstract Schwarz is right to question the methodological foundations of much of contemporary personality research. I argue that he does not go far enough, opting instead to salvage the psychometric tradition for research it cannot possibly accomplish, namely the understanding of persons in an evolutionary and historical context. Furthermore he does not address the question of measurement that has bedeviled the discipline. For all its historical tenacity, the psychometric tradition has been good at classification but weak at understanding, explanation, or description of the phenomena that most interest psychologists.

Keywords Methodology · Psychometrics · Theory · Classification · Explanation

In his original and theoretically innovative overview of the psychology of individual differences, Michael Schwarz argues that the discrepancies between normal distributions and obtained distributions are more than just errors. They reflect real differences in the manner in which his research participants understood the questions asked which is appropriately reflected in their responses. In an effort to understand the resulting data and its effect on the psychometric enterprise, Schwarz (2009, p. 195) argues that the “fluctuation of validity” can be understood as a problem of “judgmental categories.” There are three homologous judgmental patterns to be found in his results and these are what he calls “affiliation,” “approach vs. avoidance” and “life history-present state-anticipation” (p. 195). Arguing that the “human mind is still imprisoned within these behavioral categories,” the distribution anomalies of the discrete responses shown on rating scales are in effect the outcome of these basic principles. Schwarz then goes on to provide a theoretical justification for these claims.

H. J. Stam (✉)
University of Calgary, Calgary, AB, Canada
e-mail: stam@ucalgary.ca

What Do We Want to Know?

In this brief comment there are only two questions I would like to raise, neither of which I take to be final in any sense for the issues are obdurate and of long standing. The first concerns the relation between what we want to know (the phenomenon) and how we come to know this (the method). The two are not separate, for our knowledge claims are dependent on our methods while our methods will entangle us in what there is to know. I should like to begin however by taking as my point of departure the quote from Köhler that Schwarz has included in this paper. In his important book, *The place of value in a world of facts*, Köhler (1938) argued that although the aim of science is explanation, “it cannot be admitted, however, that for this reason psychologists should not be interested in phenomenal data. It seems natural to acquire at least some knowledge of those data which we intend to explain, which our constructs are expected to fit, before we begin the construction. Otherwise, why should the constructs fit?... If I were a Positivist I should, for this reason, insist upon phenomenology as the genuine basis of all explanatory construction” (p. 68). In this statement is summarized the current conundrum. The phenomena of psychology have, for the better part of the last century, been squeezed into just those methodological tricks available to us. The consequence is that the phenomena that are the focus of the question at hand are often lost to the psychologist and the discipline in favor of a Procrustean adherence to a sedulously conceived methodology, an adherence to which was once referred to as “methodolatry” by Gordon Allport who had borrowed the term from the philosophical literature.

Schwarz attempts to correct this problem through an elaborate and abstract conceptual system that covers the biological, psychological and social domains, what he calls a “bio-psycho-social model.” Not unlike the biopsychosocial model in Health Psychology (cf. Stam 2000), such phrasings signal an attempt to be comprehensive but beyond a rhetorical value, their utility is limited. While proclaiming breadth in the research enterprise the biopsychosocial model in health psychology had but a scant payoff. Such abstract and all-encompassing frameworks do not give us a way to understand problems that are ultimately solved in the minutiae of research or practice. For the question is ultimately not to solve the kinds of problems that psychometrics have bequeathed us but to understand how psychometrics frequently miss the very phenomena they might have taken on board in the first place. Schwarz’s intuitions are correct here, we have been misled by method, but more of the same method will not solve our problems.

What is a psychometric evaluation meant to do in the first instance? The invention of the correlation coefficient by Galton and Pearson was crucial for purposes of measuring the covariation of a set of variables. Galton and Pearson were primarily interested in demographics and intelligence was of course of prime interest to early psychologists along with other measures of individual difference. The relevant part of this history is the fact that the focus on the covariation of measures led to the gradual elaboration and extension of the notion of an *interindividual difference* rather than the elaboration of features of the person (for more extensive discussions see Danziger 1990; Lamiell 2003). The practical import of this notion is obvious; one can carefully delineate the position of an individual with respect to a defined

group or sub-group of scores. What is relevant in such an approach is the relative standing of the individual vis-à-vis a population as a whole and not the individual per se. Deviations of each individual's score from the group's mean is a useful indicator of the relationship between an individual's level of some attribute and the group defined by the attribute. The development of correlational techniques allowed for the creation of the notion of the normative, witness the explosion of research on intelligence early in the 20th century. The norms for the tests were aggregates of all such tests given to a normative sample. Furthermore it began the process of placing the notion of an aggregate in the center of the psychological enterprise. Hence, the correlation coefficient would be among the first of many attempts to create aggregate measures that allowed for interindividual comparisons (Danziger 1990; Lamiell 2003).

There would be a subsequent move involved in the creation of psychological methods in the first half of the 20th century. The model described above, the Galtonian model, gradually developed into the “neo-Galtonian” model, a label that Danziger (1990) has used to describe the changing models of methodology. This model is the extension of the examination of naturally constituted groups, such as children of different ages, to the comparison of artificially created groups such as those encountered in experimental settings. This was a crucial extension for it solved the problem of attributing an aggregate score to an internal quality of persons. A memory experiment could take a mean score and attribute the differences between two groups to the memory process under study. This while the experimenter was prohibited from using the mean score to point back to a single, real person taking part in the experiments. Hence aggregate scores in the research subsequently constituted the theoretical entity under investigation. For example, a measure of memory recall used to test an intervention to aid recall in two groups would be described by the aggregate score (e.g., as in ‘the treatment group demonstrated greater recall...’). No single participant in such groups would be of genuine interest, individuals ceased to matter (cf. Halpin and Stam 2006; Stam 2010). The development of experimental design in the wake of Fisher's (1966) publications followed these earlier developments and codified the procedures that would take on their contemporary character as ‘research methods’ in psychology.

I have not done justice to the historical developments here; suffice it to say that a particular form of methodology came to represent psychological phenomena in such a way that it precluded serious questions about the nature of the phenomena itself (Stam 2004). The practice of using aggregate scores in psychometrics and in experimental research meant that the search for the phenomena of interest was limited in advance to questions of how such phenomena could be characterized by the available aggregates. Alternative methods, be they psychoanalytic, gestalt, or such novel procedures as the one introduced by Egon Brunswik's “lens model” were ignored and gradually disappeared within psychology.

The novel solution proposed by Schwarz does not question this methodology so much as realign it. The fundamental use of the aggregate is maintained in his revision but replaced by a more complex, presumably inclusive model. Whatever its positive features, it does not address the fundamental issues associated with the phenomenon to be explained so much as it seeks further variation in new variables that retain the fundamental neo-Galtonian nature of the enterprise, namely the

relative comparison of individuals on psychometric scales. The subsequent attempt to explain anomalies in distributions of scores through the use of highly abstract notions does not, unfortunately, tell us about evolutionary properties such as ‘affiliation.’ It is idealized and far removed from the individual phenomenon it is meant to explain, in this case, “subjective health.” I would like to suggest instead that the original anomaly, described by Schwarz as one that became salient in the difference between psychometric measures and psychotherapeutic interviews, deserves to be explored using alternate methods. Indeed, the original aim of the psychometric component of the study was to “improve the diagnostic process; first, by identifying patients with psychological co-morbidity and place them to appropriate treatments; and second, to visualize the transition between healthy and sick as a performance benchmark for quality assurance purposes” (p. 187). That is, rather than understand “subjective health” the aim of the study was the institutional one of the classification and treatment of patients. If classification was primary, then the measure of “subjective health” was indeed not a measure of health, subjective or otherwise, but a measure of “relative comparative reported health.” Its failure to function as a device through which one might understand “subjective health” is obvious, psychometric devices are ultimately classificatory and don’t tell us about the psychological life of individuals. Like many such psychometric evaluations the primary purpose of this research was the appropriate categorization of patients for institutional functions. The attempt to understand, and even explain, is lost to an enterprise that is limited by the methodology of classification.

Whatever could be meant by ‘subjective health’ would ultimately underscore the reflexive nature of psychological constructs. For only as a member of a particular culture where such a phrase even makes sense can an investigator begin the process of attempting to understand how this is taken up by other members of that culture. Perhaps this is a trivial observation, for this holds ipso facto for all psychological constructs. In the case of subjective health however one must consider that the preoccupation with health in the contemporary world is one that has taken on new dimensions in the past 125 years as we continue to improve mortality rates, live longer with chronic diseases and have come to expect that certain acute illnesses have ceased to matter in the industrialized world (such as smallpox, poliomyelitis or dracunculiasis). Whatever could be meant by subjective health is neither global nor ahistorical and requires at the very least a historical and cultural sensitivity that ultimately reflects the preoccupations of the researchers as much as her participants. Once we abandon the project of classification we are back to the beginning or rather we start with a history of the phenomenon that is ours as well as those we wish to understand (assuming such differences even matter).

Whither Psychometrics?

The question of just what constitutes methods in modern psychology is not a single question since the discipline is home to multiple methods, tools, topics and research practices. The complexity and variety of methods currently on offer is substantial. Not even considering the full range of qualitative methods, the mathematical and quantitative issues in measurement and evaluation are

substantial. One key question of relevance to Schwarz's dilemma however is just this question of measurement.

Joel Michell, in one of the most searing critiques in recent years, has argued that psychometrics are pathological insofar as they accept a hypothesis without critical evaluation and subsequently ignore the fact that they have failed in evaluating the hypothesis (Michell 1999, 2000). The hypothesis in question is, according to Michell, simply that some psychological attributes are quantitative. This is an empirical question argues Michell and the failure to test this hypothesis, along with the subsequent concealment of this fact, is responsible for the pathology. There are ways in which one may proceed to test for the quantitative structure of an attribute that is analogous to the established sciences where, for example, we take for granted that such attributes as temperature do in fact have a quantitative structure because it can be demonstrated to be the case. Conjoint measurement has in recent years been proposed as one of the few techniques with which to determine hidden quantitative structure (see Cliff 1992; Michell 1999).

More recently Trendler (2009) has extended this argument against psychometrics to include all psychological attributes. Even conjoint measurement cannot save psychological measurement from the lack of progress on his account. In effect, he has argued that since psychological phenomena are not inherently manipulable nor controllable they are also not measurable. This is not to say that psychological phenomena cannot be manipulated or controlled at all, but rather that they are not capable of being manipulated and controlled to the degree required for measurement. They do not yield quantities. In a rebuttal, Markus and Borsboom (*in press*) challenge Trendler's thesis as flawed because it identifies a necessary condition for measurement that it cannot meet. Without rehearsing the argument here, I merely note that the disputes in measurement hinge on the degree to which obstacles to measurement make true measurement possible or impossible. However, if the quantity argument holds, and Michel is right, then it is still the case that measurement ultimately depends on a demonstration that psychological attributes are quantitative. Although unable to address the extent and depth of this question here, I raise it to note that in addition to the issue of just what phenomenon is under investigation, psychometric research of the sort advocated by Schwarz must at some point address the problem of measurement. Instead, by working backwards from the data, Schwarz demonstrates that there are anomalies in the measures he has used. He then uses these anomalies to posit abstract processes that presumably are responsible for the anomalies. However, the measurement question becomes even more difficult under such circumstances since we are now no longer pointing to the thing presumably being measured, but merely that the measures demonstrate some underlying regularities.

Schwarz argues "The validity of a psychometric scale is defined as the degree to which it measures what it was designed to measure. This requires evidence that, first, the axiomatic of measurement is 'true', and second, the operationalized construct is based on a theoretical model" (p. 203). However, Schwarz is not particularly interested in defending what he calls the "test theoretical axioms" (p. 186) since he views these as essentially conventional. If this were the case, we could merely adjust our conventions and solve the measurement problem. However, if the critics of psychometrics are correct, it is not so much the conventions that are at stake but the

fact that there is no evidence that the things being measured are in fact quantities in the first place.

Without defending psychometrics at this point or explaining what ‘true’ might refer to, Schwarz notes “there exists an inherent judgment forming mechanism that thwarts measurement by the nature of introspection and self-evaluation” (p. 203). That is, human beings are poor measuring instruments of their own subjective states since these are plagued by “inherent judgment biases.” What Schwarz calls the “fluctuation of validity” is the “mismatch between inherent categorical references to homogenous linear data” (p. 204). But the knowledge of these inherent categorical references is also determined by psychometric data and relies on an assumption that “measurement is true” while bypassing the quantity question. For what if the problem is not an inherent judgment bias but a foundational measurement problem? I agree there is no obvious way to test this difference, except to say that it is incumbent on Schwarz to demonstrate the quantitative nature of the new variables he introduces to replace the original model. I expect this will be difficult.

Let us assume however that the “axiomatic of measurement is ‘true’” on Schwarz’s account. We are not still much further along since we do not know if the “operationalized construct” is in fact the right one. There are many ways to operationalize a construct, as the argument goes, and of course construct validity was invented in the 1950s to help us overcome the possibility that we might have many operational definitions for the same construct and not much by way of choosing between them. I appreciate that we are better off having the operationalized construct based on a theoretical model, as Schwarz suggests, but such models are no guarantee that the operationalized construct is in fact representative of anything. We are now at the point where we have to confront the entire conventionalist approach of psychological methods whereby a belief in conventions surpasses the difficult work of slowly and carefully evaluating phenomena without the aid of routine and formulaic methods. We have witnessed over 100 years of methodologically sophisticated research that has yet to demonstrate the kind of serious advances we might associate with advances in mathematics, biology or neuroscience. One reason for this continues to be (and I am hardly original in this claim) the adherence to methodological practices whose aim is the refinement of theoretical models whose relationship to the life of human beings is only vaguely realized.

In conclusion, I applaud Schwarz for his careful analysis of his research findings that led him to reject the methodological dictates of the psychometric approach he was using. I fear however that its refinement will only lead to more of the same, a gradual closing of the phenomenon at hand in favor of a technological sophistication that will only lead us further astray.

References

- Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science*, 3, 186–190.
- Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. New York: Cambridge University Press.
- Fisher, R. A. (1966). *The design of experiments* (8th ed.). London: Oliver and Boyd. Original work published 1935.

- Halpin, P. F., & Stam, H. J. (2006). Inductive inference or inductive behavior: Fisher and Neyman–Pearson approaches to statistical testing in psychological research (1940–1960). *The American Journal of Psychology*, *119*, 625–653.
- Köhler, W. (1938). *The place of value in a world of facts*. New York: Liveright Publishing Corporation.
- Lamiell, J. T. (2003). *Beyond individual and group differences*. Thousand Oaks: Sage.
- Markus, K. A. & Borsboom, D. (in press). The cat came back: Evaluating arguments against psychological measurement. *Theory & Psychology*.
- Michell, J. (1999). *Measurement in psychology: Critical history of a methodological concept*. Cambridge: Cambridge University Press.
- Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory & Psychology*, *10*, 639–667.
- Schwarz, M. (2009). Is psychology based on a methodological error? *Integrative Psychological and Behavioral Science*, *43*, 185–213.
- Stam, H. J. (2000). Theorizing health and illness: functionalism, subjectivity and reflexivity. *Journal of Health Psychology*, *5*, 273–283.
- Stam, H. J. (2004). Is there (still) a place for theory in psychology? *History and Philosophy of Psychology Bulletin*, *16*(2), 3–9.
- Stam, H. J. (2010). The tradition of personalism and its relationship to contemporary indeterminate functionalism. *New Ideas in Psychology*, *28*, 143–150.
- Trendler, G. (2009). Measurement theory, psychology and the revolution that cannot happen. *Theory & Psychology*, *19*, 579–599.

Henderikus J. Stam is a Professor of Psychology at the University of Calgary and a member of the Theory as well as the Clinical Psychology programs. He is the founding and current editor of *Theory & Psychology*. Although he is the author or co-author of a number of articles on methods he is most interested in their historical creation and enduring significance for the discipline.