

## **Echoes from the Past: Meaning in Measures, Environments, and Predictions**

Beatrice J. Krauss

Hunter College, City University of New York and University of Arizona

The elder statesmen of psychology – such as Lewis Petrinovich, Donald Campbell, Samuel Messick, Kurt Lewin, and Paul Meehl – have been instructing on methodology for decades. But psychology seems to have a short memory and an aversion to becoming a cumulative science. Examples are given in measurement, health intervention, understanding of environments, applying the wisdom of each of these eminent scientists.

I learned methodology from Donald T. Campbell (1916-1996), and I learned measurement from Samuel Messick (1931-1998). I was trained as a social psychologist, but my career became that of a health psychologist, largely in HIV prevention. Thus, in my work, the measured effects of tested interventions matter, and their viability in specific environments and with specific populations matter. I will also say that Lewis Petrinovich landed squarely on some of my pet peeves about research in psychology. Reading his treatise now, I find myself asking, “Has the field no memory? When are we going to learn to build a sound science that is cumulative?” There are, however, some glimmers of hope.

Let me begin with measurement.

**What are you measuring, really?** Sam Messick had a hard and fast rule regarding questionnaires, surveys, and other forms of tests: you are measuring responses to the items or tasks. I read so many articles that conclude someone has “high self-esteem,” or “high blood pressure,” or “poor range of motion in their wrist,” or “normal body weight.” If the weight is a self-report, it is almost universally closer to a report of ideal weight than real weight—ten US states have dropped weight from their drivers’ license descriptions for that very reason. It matters that it was self-report. But so often self-report is reified as “the thing being reported on.” The “self-esteem” was likely a set of reactions to items on a specific questionnaire—and those questionnaires vary widely in their underlying theory of what contributes to, or constitutes, self-esteem (Robinson, Shaver & Wrightsman, 1991).

You think you are out of the woods with “objective” measures from fancy equipment. We all know blood pressure measurements vary if they are taken at home, or in the doctor’s office (if you are prey to the “white coat” effect). But lately I have been studying cognitive bias in medical diagnosis. Rotem-Lehrer et al. (2016) describe an anchoring effect that leads to significantly different wrist range of motion readings using a goniometer; their manipulation in that study was a past history with no prior injury, a moderate, or greater degree of injury. The patient was a volunteer who, in reality, had never had an injury and who was unaware of the condition to which the professional

therapists had been assigned. Radiologists have policed their own specialty because eye-tracking reveals diagnostic images have sections that are “missed,” or “misinterpreted.” The entire field of medicine is prey to over 50 documented biases, which include throwing out sophisticated lab results which don’t “fit” with an initial, and perhaps premature, diagnosis (Croskerry, 2020; Ryle, 2019). Do we think psychologists and psychology researchers are immune to similar errors?

The antidotes to these errors include correct reporting (e.g., self-reported self-esteem on a paper-and-pencil version of the Rosenberg scale), and Campbell’s approach of critical multiplism (Shadish, 2010), that is, applying multiple methods, multiple disciplines, and multiple viewpoints to every step of the research enterprise.

### **How can you study populations and environments you don’t understand?**

Answer—you can’t. You may have to do “discovery” research preliminarily to any attempt to design and evaluate an intervention, and to understand boundary conditions which may limit or enhance generalizability. Boundary conditions, however, cannot merely be asserted, but should be theoretically supported and testable. A Public Health student of mine gave a telling example. Injecting drug users (IDUs) often develop potentially lethal infections at the site of injections. As part of a larger HIV prevention and health promotion program, outreach workers would give an alcohol wipe with each clean syringe (in a syringe exchange program) to each IDU and instruct the IDU to clean the injection site before injecting. But, in this neighborhood, many IDUs were homeless or poorly housed. Observers noted that the homeless or poorly housed IDUs first used the wipe to clean their faces, then their hands, and then used the now thoroughly contaminated wipe at the injection site. Solution—more instruction and three alcohol wipes. Verification of the “boundary condition” of prior access to resources for hand and face washing could be readily tested by looking at the initial appearance (clean/dirty) of those who did or did not use the alcohol wipes initially for washing. Observation is a research method (see Byrne, this issue). But if there is not a strong ethic of upward communication from client to outreach worker to Principal Investigator, many important observations are missed. If there is not strong formative work, many important observations are missed. People live their daily lives in environments. And as Kurt Lewin tried so very hard to tell us those environments can have hard or soft boundaries; environments are also often perceived differently from the inside and the outside, with both perspectives necessary in many cases for effective intervention (Lewin, 1951).

**A ray of hope.** In the field of health psychology, we are beginning to talk about “Thresholds of Clinical Significance (TCS)” rather than statistical significance, that is, treatments with an effect size that makes a difference in health outcomes (Freedland, 2020). Although not cited, Freedland’s article makes a point previously made by Paul Meehl in 1990 about the utility of statistical significance testing vs point prediction in verification of theories and findings. Now, let’s polish TCS’s for specific populations in specific environments where necessary, and recognize when some effects are quite generalizable.

## References

- Croskerry, P. (2020). *The cognitive autopsy*. Oxford University Press.
- Freedland, K.E. (2020). Pilot trials in health-related behavioral intervention research: Problems, solutions, and recommendations. *Health Psychology, 39*, 851-862.
- Lewin, K. (1951). *Field theory in social science: Selected theoretical papers* (D. Cartwright, Ed.). Harpers.
- Meehl, P. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1*, 108-141.
- Robinson, J.P., Shaver, P.R., & Wrightsman, L.S. (1991). *Measures of personality and social psychological attitudes* (Vol.1). Academic Press.
- Rotem-Lehrer, N., Singer, N., Reshit, O., & Springer, S. (2016). Measuring up to expectation: cognitive bias in wrist range-of-motion measurement. *Journal of Orthopaedic and Sports Physical Therapy, 46*, 1037-1041.
- Ryle, C.A. (2019). *Risk and reason in clinical diagnosis*. Oxford University Press.
- Shadish, W.R. (2010). Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods, 15*, 3-17.