



Barrett View #5

Does it Matter if Psychological Attributes Do Not Vary as Quantities?

With respect to the recent publication [The EFPA Test-Review Model: When Good Intentions Meet a Methodological Thought Disorder](#), and its accompanying summary blog article: [The Achilles' Heel of Psychometrics](#), a colleague asked:

Does it matter if trait scores aren't interval in nature? If they predict outcomes of interest, is it problematic to assume that these lumpy ordinal data are "quantitative" in the physics sense?

Bottom line, it's all about the 'precision' of a score, and the reliance placed upon the psychological attribute varying so precisely (as a quantity with all that implies).

Here are some example contexts where the assumptions concerning the precision of a psychometric test score may have adverse consequences:

1. When a statistically significant mean difference of scores of say 2 or 3 points on a 0-24 item scale is used to make a claim of substantive magnitude difference on the psychological attribute, you will now have to justify that claim. If you have no evidence the attribute varies as a quantity, and you cannot explain how you would observe the psychological consequence of a 2 or 3-point difference in score magnitude, then a judge or critic would ask on what basis are you making the claim that the difference is important? The specific line of attack would be from [Michell, J. \(2012\). Alfred Binet and the concept of heterogeneous orders. *Frontiers in Quantitative Psychology and Measurement*, 3, 261, 1-8.](#)

2. My published example of the 'threshold' definition of learning disability both here in NZ and in the US (where the death penalty in US homicide cases is assigned or not, based on the assessed IQ score of a felon being below or above 70). All of the calculations, the confidence intervals etc, true-scores, latent variable and usual psychometrics - all assume "Intelligence" varies as an equal-interval continuously varying quantity. I'm now saying "show me and the judge the evidence that it does so" - given the solid background of referenced publications which show what is required to be presented as 'evidence'. The reality is, what's likely to happen after skilled, evidence-based cross-examination is the psychologist saying:

"I believe it to be so but I can't offer any evidence that it is so".

The point being that we can clearly observe that there are relations between IQ scores and outcomes, but these can be found using order-based statistics which match how everyone but psychometricians actually use and interpret the scores. Again, a simple questions to ask in court of your "measurement expert is:

"Given a numerical difference in intelligence between someone with an IQ score of 65 and 75, or 80 and 90, or 120 and 130, can you can explain to the court how we can observe that difference in this person's

intelligence, describing to us how it varies over each of the 10 points?"

I'm asking for psychologists to be honest about what can reasonably be concluded based upon a range of empirical evidence, and what cannot be concluded.

3. When an organization forces incumbent employees to re-apply for their jobs using psychometric assessment, then the 'precision' of the scores used as 'filters' comes into question. If there is no evidence that the attribute indexed by a score magnitude varies as a quantity, and the test publisher cannot provide evidence for what simple score differences meaningfully convey in terms of the differences between individuals or the likely outcome that would occur as a result of rehiring or retiring a person with a particular score, then such activity could be challenged in an employment court. Clearly, not all organizations rely upon such precision in the test scores. But, it now means any organization contemplating using psychometrics in such contexts need to be very careful about the assumed precision of a test score.

4. When a psychologist seeks to *causally* explain variation in a psychological attribute, they need to explain how any biological system can sustain an equal-interval, continuously varying quantity of a psychological attribute (for which we can't even agree upon as to its precise meaning); for example:

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.

Borsboom, D., & Mellenbergh, G.J. (2002). [True scores, latent variables, and constructs: a comment on Schmidt and Hunter.](#) *Intelligence*, 30, 505-514, dealt with the *true-score-as-a-psychological-quantity* assertion years ago - completely ignored as usual by those who simply could not countenance any threat to their personal beliefs.

So, What Can We Do Better?

[The EFPA Test-Review Model: When Good Intentions Meet a Methodological Thought Disorder](#) article explains in some detail out how to proceed in future, avoiding the legal risk of claims of quantitative precision. And, of course, there are many ways of providing evidence of the pragmatic consequences of any test score magnitude - but these are mostly actuarial (linking test score magnitudes to empirical (*not hypothetical*) counts or probabilities of criterion outcomes). Psychologists have to do the hard yards of collecting such data. Forensic psychology and Psychiatry had to do these post-Ziskin because judges no longer took psychologists' beliefs or assertions as 'evidence' for any prediction they made.

Alternatively, we accept that we can make reasonable claims about fuzzy-boundary orders of magnitudes that do possess pragmatic utility ... but to go further in terms of being 'precise' requires the kind of evidence mostly lacking - because psychologists have refused for so long to accept that they cannot show anyone that they are making quantitative measurement of any psychological attribute. So, the hard yards have not been done; and yes, they are very hard.

posted 21st January, 2018