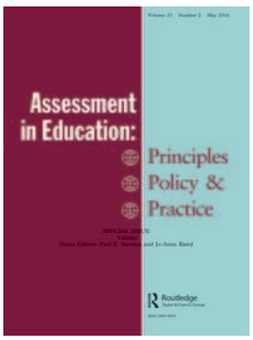

Routledge
Taylor & Francis Group

Assessment in Education: Principles, Policy & Practice

ISSN: 0969-594X (Print) 1465-329X (Online) Journal homepage: http://www.tandfonline.com/loi/caie20

Frankenstein's validity monster: the value of keeping politics and science separated

Denny Borsboom & Lisa D. Wijsen

To cite this article: Denny Borsboom & Lisa D. Wijsen (2016) Frankenstein's validity monster: the value of keeping politics and science separated, Assessment in Education: Principles, Policy & Practice, 23:2, 281-283

To link to this article: http://dx.doi.org/10.1080/0969594X.2016.1141750

Published online: 20 Jun 2016.

Assessment in Education: Principles, Policy & Practice, 2016
Vol. 23, No. 2, 281–283, http://dx.doi.org/10.1080/0969594X.2016.1141750

COMMENTARY

Frankenstein's validity monster: the value of keeping politics and science separated

Denny Borsboom* and Lisa D. Wijsen

Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

(*Received 11 January 2016; accepted 11 January 2016*)

Keywords: Validity; assessment; testing

> [S]cience can stand on its own feet and does not need any help from rationalists, secular humanists, Marxists and similar religious movements; and … non-scientific cultures, procedures and assumptions can also stand on their own feet and should be allowed to do so … Science must be protected from ideologies; and societies, especially democratic societies, must be protected from science.
>
> – Paul Feyerabend, *Against Method*, p. viii.

The original definition of validity (Newton & Shaw, 2015), which is arguably still the textbook definition in methodology (Hood, 2009), is concerned with a matter of *truth*: does a test measure what we intend to measure? This definition, which can be fleshed out in terms of a causal relation between the intended attribute and the test score (Borsboom, Mellenbergh, & Van Heerden, 2004), defines validity as a question about the test, which has an empirical answer and may thus be scientifically investigated. Later definitions (e.g. Messick, 1989), however, defined validity in terms of *evidence*: do we have sufficient empirical and theoretical justification for a given score interpretation? Unless one conflates evidence and truth, which is highly inadvisable, these questions cannot be answered with the same resources (Borsboom & Markus, 2013). In the past decades, these two already incompatible questions have been complemented with a third; namely, whether test *use* is justified or not (Kane, 2015; Sireci, 2015). This question transcends both truth and evidence, because it requires the specification what one wants to achieve through test use, which typically involves moral and ideological considerations (Markus & Borsboom, 2013).

These three questions mirror the three classical philosophical questions of *what there is* (ontology), *how we can know it* (epistemology), and *what we should do* (ethics). Current validity theory does not respect the distinctions between these domains, but instead advises to mingle them into a single overarching judgement. In our view, this is tantamount to an advice to create Frankenstein's monster. One cannot meaningfully combine these issues because, to use a psychometric metaphor, they are not scalable on the same dimension; or, in Kuhnian (Kuhn 1962) terms,

*Corresponding author. Email: d.borsboom@uva.nl

they are *incommensurable*. A lack of ethical justification cannot be made up for by better psychometric properties or vice versa; Nazi war crimes are not mitigated by the fact that the Nazis used really good measurement procedures, and IQ tests do not gain psychometric quality from the fact that their use has advanced society. If one cannot trade-off such properties, it is unclear how to integrate them so as to actually construct Markus (2015) 'ladder C'. It is illustrative that, as Cizek (2015, p. 15) notes 'there is no evidence that the desired outcome – an integrated evaluative summary of the evidence – has ever successfully been accomplished'.

The distinction between facts and moral values is, in our view, highly desirable: science and politics should keep to their own territories. Traditionally speaking, science can be seen as an ivory tower, which attempts to do its job in isolation of external influences. Politics does not mandate methods of scientific research or standards of justification; science is responsible for this by its own. Only by upholding this strict borderline can we aim for objective knowledge, which is independent of moral judgements. Likewise however, political considerations should not be mingled with or reduced to scientific ones, lest we degrade into scientism. Science and politics simply have different goals: whereas science is interested in questions about truth and falsity, politics is concerned with what is right or wrong. Matters of test use justification (politics), are essentially different from a matters of test validity (science), so that one concept simply cannot answer to both. The statement 'a thermometer is a valid instrument for measuring temperature' is of a different *kind* from 'it is justified to hire people based on their IQ-score'. As Cizek (2015) argues, trying to combine these issues in a single equation is like attempting to square the circle.

Now, why should we not let go of this rigid distinction? Why not break with tradition? The answer to this question is very simple: *because this distinction is in fact highly useful.* Validity, as a scientific concept that encodes the truth of a central psychometric claim, is worthy of keeping. It will always be a relevant question whether a test indeed measures what it is intended to measure; researchers should always be interested in this question, and having this knowledge will at all times contribute to better scientific theories. If validity becomes an overarching term that incorporates all sorts of test-related issues, this foundational question is relegated to a marginal position. Test validity, however, deserves a central position in the psychometric conceptual framework; and for this reason, Newton and Shaw's (2015) idea of dropping the concept altogether is inappropriate. Educational testers do not own the concept of validity, so it is not theirs to drop.

We realise that using such a traditional definition of validity is problematic for the field of educational testing. Finding out whether a test indeed measures the intended attribute is not easy, and often we cannot be sure that our test is indeed valid. Moreover, for most educational tests, it is hard to pinpoint one entity that is supposedly measured; often, tests measure a composite of knowledge of different fields. Even though this is problematic for the testing industry, that should not be a reason to adjust the definition of validity to something less restrictive, or to allow psychometric shortcomings to be compensated by good intentions.

It is important to stress that we do not intend to deny the importance of making the right decisions. These decisions can have a profound influence on a person's life and should not be taken lightly at all (Borsboom & Mellenbergh, 2007). Yet, test validity does not depend on whether decisions are right or wrong and should not be understood as such. Political motivation should remain distinct from science, and vice versa.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Erc [Grant Number 647209].

Notes on contributor

Denny Borsboom is professor of psychology at the University of Amsterdam. He has published on a range of issues in psychometrics, including validity; his current research involves the evaluation and development of psychometric models in clinical psychology.

Lisa D. Wijsen is a PhD student at the University of Amsterdam, who studies the history of psychometrics.

References

Borsboom, D., & Markus, K. A. (2013). Truth and evidence in validity theory. *Journal of Educational Measurement, 50*, 110–114.

Borsboom, D., & Mellenbergh, G. J. (2007). Test validity and cognitive assessment. In J. Leighton & M. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 85–118). Cambridge: Cambridge University Press.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061–1071.

Cizek, G. J. (2015). Validating test score meaning and defending test score use: Different aims, different methods. *Assessment in Education: Principles, Policy, & Practice (this issue).*

Feyerabend, P. (1975). *Against method*. London: New Left Books.

Hood, S. B. (2009). Validity in psychological testing and scientific realism. *Theory and Psychology, 19*, 451–473.

Kane, M. T. (2015). Explicating validity. *Assessment in Education: Principles, Policy, & Practice (this issue).*

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Markus, K. (2015). Alternative vocabularies in the test validity literature. *Assessment in Education: Principles, Policy & Practice (this issue).*

Markus, K., & Borsboom, D. (2013). *Frontiers of validity theory: Measurement, causation, and meaning*. New York, NY: Routledge.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). Washington, DC: American Council on Education and National Council on Measurement in Education.

Newton, P. E., & Shaw, S. (2014). *Validity in educational & psychological assessment*. London: Sage.

Newton, P. E., & Shaw, S. D. (2015). Disagreement over the best way to use the world 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy & Practice (this issue).*

Sireci, S. G. (2015). On the validity of useless tests. *Assessment in Education: Principles, Policy, & Practice (this issue).*