

The importance of ideas: the semantics of validity

Paul E. Newton

Date: 26 September 2013

Venue: Hughes Hall, Cambridge

www.ioe.ac.uk



Available at all good book shops, from April 2014



Leading education
and social research
Institute of Education
University of London

Validity in Educational and Psychological Assessment

Paul E. Newton *Institute of Education, University of London* and **Stuart D. Shaw** *Cambridge Assessment*

Validity is the hallmark of quality for educational and psychological measurement. But what does quality mean in this context? And to what, exactly, does the concept of validity apply? These apparently simple questions parachute the unwary inquirer into a minefield of tricky ideas. This book guides you through this minefield, investigating how the concept of validity has evolved from the nineteenth century to the present day.

Communicating complicated concepts straightforwardly, the authors explore answers to questions like:

- what does 'validity' mean?
- how many different kinds of validity are there?
- does validity have a social dimension as well as a technical one?
- what are the ingredients of effective validation?
- is 'construct validity' still relevant?
- is 'validity' fit for purpose?

VALIDITY in EDUCATIONAL &
PSYCHOLOGICAL ASSESSMENT

PAUL NEWTON & STUART SHAW

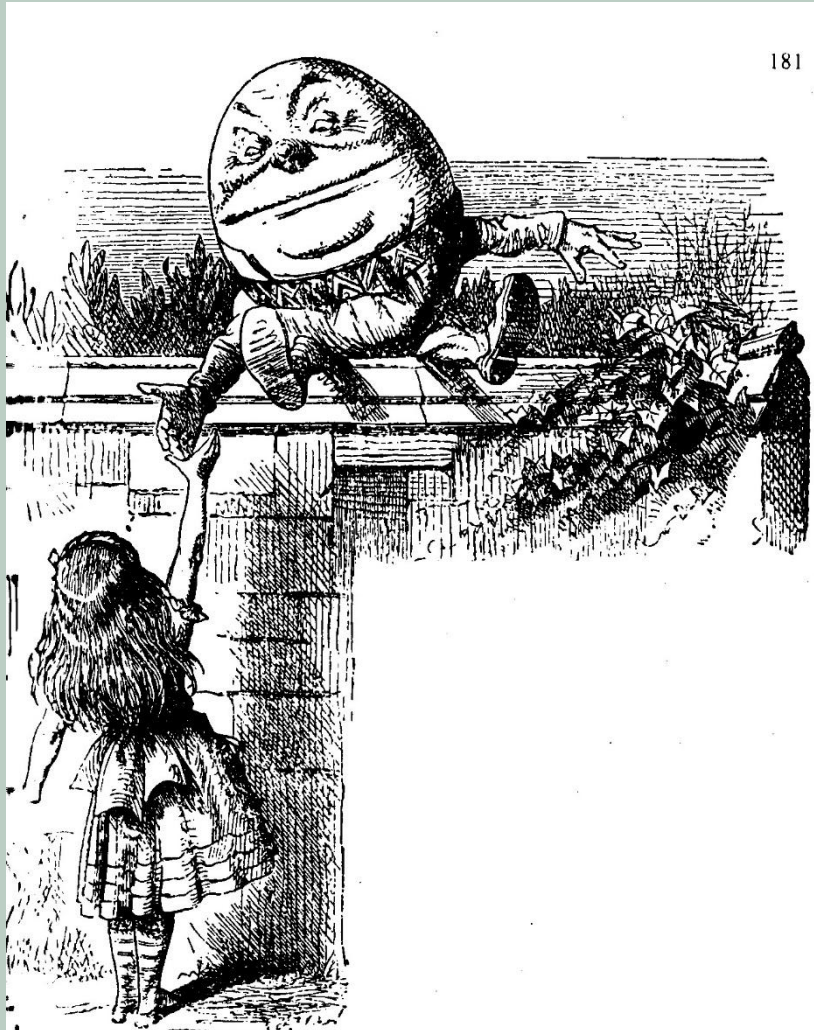


Lewis Carroll (1872)

Through the Looking-Glass



Leading education
and social research
Institute of Education
University of London



There's **glory** for you!”

“I don't know what you mean by ‘glory,’ ” Alice said.

Humpty Dumpty smiled contemptuously. “Of course you don't — till I tell you. I meant ‘there's a nice knock-down argument for you!’ ”

“But ‘glory’ doesn't mean ‘a nice knock-down argument,’ ” Alice objected.

“When *I* use a word,” Humpty Dumpty said, in rather a scornful tone, “it means just what I choose it to mean — neither more nor less.”

“**The question is,**” said Alice, “**whether you can make words mean so many different things.**”

The importance of being valid



Leading education
and social research
Institute of Education
University of London

“The key criterion driving assessment at Cambridge Assessment is validity.” (Cambridge Assessment, 2009)

“[Validity] is the most important aspect of the quality of an assessment.” (ETS, 2002)

“Validity is, therefore, the most fundamental consideration in developing and evaluating tests.”
(AERA, APA, NCME, 1999)

Claims like these are **hollow unless we can say (and others can tell) what we mean by ‘validity’**

The foundations for a semantic analysis of validity



Leading education
and social research
Institute of Education
University of London

Everyone seems to agree that:

- Validity is a **very important** concept
 - the *most* important? first among equals?
- Validity has something to do with **measurement**
 - *purely* about measurement? measurement or assessment?

Most people seem to agree that:

- Validity is a **property** of something
 - of a test? of a score? of an argument?
- Validity has something to do with **strength**
 - measurement strength? argument strength?

X has validity



Leading education
and social research
Institute of Education
University of London

If we are happy with the grammar of ‘X has validity’ then we can try to define validity by asking

- what kind(s) of **object** can X be?
- what **property** is (properties are) common to valid Xs?

We'll explore 2 very different ways of defining validity



Leading education
and social research
Institute of Education
University of London

1. Validity as **measurement**
2. Validity as **justification** (a.k.a. validity as **argument strength**)

... there are other important categories of definition, but these are the most divergent

Validity 1

Borsboom (and colleagues)



Leading education
and social research
Institute of Education
University of London

The premise from which Borsboom et al begin:

- **validity is a concept that affirms measurement**
(we can trace this idea back to classic definition)

Measurement

- “[...] there are no universal characteristics of measurement except the ontological claim involved. The only thing that all measurement procedures have in common is the either implicit or explicit assumption that **there is an attribute out there** that, somewhere in the long and complicated chain of events leading to the measurement outcome, is **playing a causal role in determining what values the measurements will take.**” (Borsboom et al, 2004, pp.1062-3)

Borsboom (and colleagues) on the semantics of validity



Leading education
and social research
Institute of Education
University of London

X has validity

- X = the **test**
- Validity = the property of **measurement**

What would make the claim “X has validity” true?

The test must be sensitive to variation in the targeted attribute, which means that

1. the attribute must **exist**
2. variations in test scores must be **caused by** the attribute.

My take on Borsboom et al and the semantics of validity



Leading education
and social research
Institute of Education
University of London

In signal processing terms

- **validity = signal acquisition**
- if the signal (from the attribute) is received (i.e. causally affects the score), then the test has validity

In addition, though, the score may also be affected by

- **noise** (random construct-irrelevance)
- **interference** (systematic construct-irrelevance)

This means (according to Borsboom et al) that a test can be both **valid *and* **useless for measuring**.**

(p.s. they are quite happy with this conclusion!)

Borsboom (and colleagues) ... in the balance



Leading education
and social research
Institute of Education
University of London

Reasons to be cheerful

- it is a tight definition of validity
- it has potential to guide validation

Reasons to be glum

- it implies that any validity argument needs to furnish empirical evidence and logical analysis to establish that “the attribute exists and causes the scores”
- tightness is achieved at the cost of lexical incongruity (a test that is valid, but that cannot be used to measure, sounds like an oxymoron)

Validity 2

The 1999 *Standards*



Leading education
and social research
Institute of Education
University of London

Definition

- “Validity refers to the **degree to which evidence and theory support the interpretations of test scores** entailed by proposed uses of tests.”

What is **X** (in ‘X has validity’)?

- the **score interpretation**
- the **claim** that scores can be interpreted in a particular way (the conclusion of the validity argument)
 - e.g. “the test scores measure [the targeted attribute]”

What is **validity** (in ‘X has validity’)?

- the degree of **justification** for the **score interpretation** claim
- the **strength** of the **validity argument** (and its conclusion)

The 1999 *Standards* and the semantics of validity



Leading education
and social research
Institute of Education
University of London

X has validity

- X = the **validity argument** (and ultimately its conclusion, which is the score interpretation claim)
- Validity = the property of **strength**

What would make the claim “X has validity” true?

The validity argument must be coherent and complete and all of its inferences and assumptions must be plausible (these are criteria for judging arguments taken from Kane, 2013).

The 1999 *Standards* ... in the balance



Leading education
and social research
Institute of Education
University of London

Reasons to be cheerful

- it is a tight definition of validity
- it emphasises the centrality of validity argument to validation

Reasons to be glum

- tightness is achieved at the cost of importing a definition of validity from a different discipline ('inductive validity' a.k.a. 'strength')
- higher validity means a stronger conclusion, not better measurement
- it begs the substantive question (e.g. how validity relates to measurement)

Validity 1 (measurement) vs. Validity 2 (justification)



Leading education
and social research
Institute of Education
University of London

Let us:

- frame the validity argument conclusion in terms of measurement *quality*: “the test scores measure the targeted attribute **adequately**”
- assume that the strength of this validity argument (and its conclusion) is **high**.

1. Define validity as measurement

- **the measurement procedure has minimal validity**
- although the argument in support of this conclusion is strong

2. Define validity as justification

- **the validity argument has high validity**
- although the measurement procedure has only minimal measurement quality

Questions that puzzle me



Leading education
and social research
Institute of Education
University of London

My question for Borsboom:

- what does your concept of validity **add to** the concept of **measurement**?

My question for the *Standards*:

- if what you mean by 'validity' is the strength of the validity argument, then why not simply refer to its **strength** (if not, then what on earth *do* you mean by validity)?

The oddness of Borsboom's narrow definition of validity



Leading education
and social research
Institute of Education
University of London

The premise from which Borsboom et al begin:

- **validity is a concept that affirms measurement**
(we can trace this idea back to classic definition)

step 1: define validity as measurement

step 2: define measurement

- but, from this perspective, what does the concept of validity *add to* the concept of measurement?
- step 1 seems to render the concept of validity **redundant** with the concept of measurement (Keith Markus makes this point in Markus & Borsboom, 2013, p.313)

My current line of thinking



Leading education
and social research
Institute of Education
University of London

1. Defining validity as an argument concept (Validity 2), rather than as a measurement concept (Validity 1), begs the important substantive definitional question.
2. However, defining validity as measurement (Validity 1) renders the concept of validity redundant with the concept of measurement.
3. Debate over ‘the proper meaning of validity’ continues to have serious negative consequences
 - it wastes time that could be spent on the substantive definitional challenge
 - it causes unhelpful rifts between measurement professionals
 - it results in widespread confusion within and beyond the field
4. Maybe we can **sidestep the need to talk about validity** (let alone to define it) by focusing directly upon the substantive concepts, e.g.
 - measurement quality (from the classical perspective)
 - testing policy value (acknowledging more recent debates)