

Improving the validity of assessments:  
It's less about the tests and more about  
what's done with them

Gordon Stobart

Institute of Education  
University of London

# The power of assessment

Unspeakably more depends *on what things are called* than on what they are...creating new names and assessments and apparent truths is enough to create new 'things'.  
*(Friedrich Nietzsche, 1887)*

The individual in contemporary society is not so much described by tests as constructed by them  
*(Allan Hanson, 1996)*

I'm no good at spelling ..and I'm hopeless at times tables so I'm frightened I'll do the SATS and I'll be a nothing.  
*(Hannah age 11, Reay & William 1996)*

# Improving the validity of assessments

Validity is ‘the extent to which the evidence supports or refutes the proposed interpretations and uses’ of an assessment.

*(Michael Kane, 2006)*

We can improve validity by both limiting what claims we make for our assessments and by encouraging the forms of assessment which may improve the quality of learning.

# Threats to validity

- Construct underrepresentation ('English')
- Construct-irrelevance (speed, predictability)
- Fallacies in validity arguments (Kane):
  - Begging the question – scoring, generalization & extrapolation
    - of consequences – 'readiness';
  - Over-generalization or spurious precision;
  - Surrogation – 'the fallacy of statistical surrogation'
  - Reification – factor analysis; 'excess meaning';
  - The Straw-Man – test developers' 'ambitious interpretations'
  - Guilding the Lily – multiple-choice alpha

# Five steps to improving validity

## Step 1.

### Avoid 'excess meaning'- focus on *achievement*

'A modest interpretation makes fewer claims and is therefore easier to justify' (Kane, p.60)

'Sometimes an expression has to be withdrawn from language and sent for cleaning – then it can be put back into circulation.' (Wittgenstein)

Is 'ability' in need of a good clean?

Hurrah for ETS: 'The SAT assesses how well you analyze and solve problems – skills you learned in school that you'll need in college'

# Ability testing – ‘the new IQism’

‘We must not say that his high score is due to his high ability, but if anything the reverse. We say he has high ability because his performance has yielded a high criterion score. His “ability” is simply a summary statement concerning his actions’ (Cureton, 1951)

‘you can’t *give* someone ability’ (Teacher, in Gillborn & Youdell, 2001)

# Ambitious interpretations and reification – the case of IQ

From Binet to Burt:

It is in this practical sense, the only one accessible to us, that we say that the intelligence of these children has been increased. We have increased what constitutes the intelligence of a pupil, the capacity to learn and assimilate instruction. (Alfred Binet, 1909,)

This general intellectual factor, central and all pervading, shows a further characteristic, also disclosed by testing and statistics. It appears to be inherited, or at least inborn. Neither knowledge nor practice, neither interest nor industry, will avail to increase it. (Cyril Burt, 1937)

# Ian Hacking's ten *Engines of Discovery* in 'making up people'

1. Count
2. Quantify
3. Create norms
4. Correlate
5. Medicalise
6. Biologise
7. Geneticise
8. Normalise
9. Bureaucratise
10. Reclaim our identity

# Other examples of 'making up people'

- Multiple Intelligences (8.5)
- Emotional Intelligence
- Learning Styles (Dunn & Dunn, VAKT)

# Five steps to improving validity

## Step 2

### Interpret results more cautiously

Avoid

- *Over-interpretation* – IQ/Ability
- *Unreliable interpretations* – EI & Learning Styles
- *Over-simplistic interpretations* – test scores = standards in accountability testing

# Five steps to improving validity

## Step 3

### **Acknowledge the importance of context**

- The problem of fair = standardised
- The elusive 'level playing field'
- The 19<sup>th</sup> century legacy – privilege becomes merit
- Culture free/reduced – Raven's matrices

## Access

## Curricular

## Assessment

<p>Who gets taught and by whom?</p>	<p>Whose knowledge is taught?</p>	<p>What knowledge is assessed and equated with achievement?</p>
<p>Are there differences in the resources available for different groups?</p>	<p>Why is it taught in a particular way to this particular group?</p>	<p>Are the form, content and mode of assessment appropriate for different groups and individuals?</p>
<p>What is incorporated from the cultures of those attending?</p>	<p>How do we enable the histories and cultures of people of color, and of women, to be taught in responsible and responsive ways?</p>	<p>Is this range of cultural knowledge reflected in definitions of achievement? How does cultural knowledge mediate individuals' responses to assessment in ways which alter the construct being assessed?</p>

# The level playing field: equal opportunities

The chances...were adjusted by burying the man to his waist, tying his left hand behind his back, and arming him with only with a mace, while his fair opponent had the free use of her limbs and was provided with a heavy stone securely fastened with a piece of stuff.

(German test of innocence, Hanson 1994)

# The 19<sup>th</sup> Century legacy

## R.H.Tawney's *Tadpole Philosophy*

It is possible that intelligent tadpoles reconcile themselves to the inconveniences of their position, by reflecting that, though most of them will live and die as tadpoles and nothing more, the more fortunate of the species will one day shed their tails, distend their mouths and stomachs, hop nimbly on to dry land, and croak addresses to their former friends on the virtues by means of which tadpoles of character and capacity can rise to be frogs. (1951)

# The Tadpole Philosophy (2)

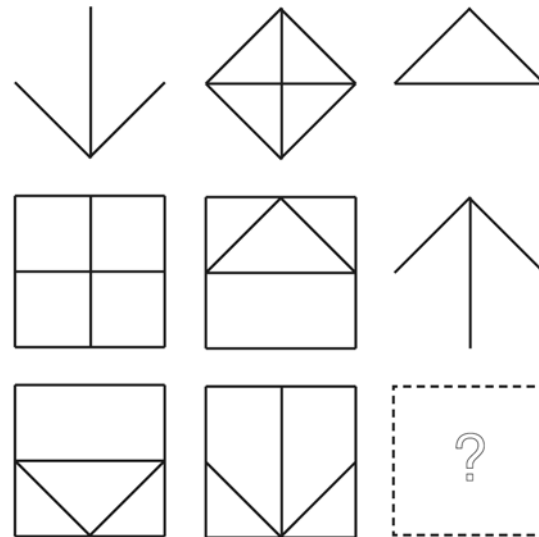
Look at every walk of life and see whether it be not true that those who attain high distinction in the world were generally men who were distinguished in their academic career....  
Whatever be the languages, whatever the sciences, which it is, in any age or country the fashion to teach, the persons who will become the greatest proficient in those languages and those sciences will generally be the flower of the youth, the most acute, the most industrious, the most ambitious of honourable distinctions.

(Lord Macaulay 1833)

Intelligence testers hard-wire merit and vice

- All feeble-minded are at least potential criminals. That every feeble-minded woman is a potential prostitute would hardly be disputed by anyone. Moral judgment, like business judgment, social judgment or any other kind of higher thought process, is a function of intelligence (L.Terman, 1916)

# Culture-free: Raven's Matrices



# Five steps to improving validity

## Step 4

### **Recognise the importance of interaction**

Resist simple additive models – ‘60% inherited’

Flynn’s *social* and *individual multipliers* – dogs, rabbits & Tiger Woods

Importance in formative assessment - feedback

# Five steps to improving validity

## Step 5

### **Create sustainable assessment**

‘Any assessment act must also contribute to learning beyond the immediate task...assessment that meets the needs of the present and prepares students to meet their own future needs’ (David Boud, 2000)

This is the ‘*double duty*’ of assessment