

# Educational Assessment Australia

[www.eaa.unsw.edu.au](http://www.eaa.unsw.edu.au)

# **Writing Good Assessments 2009**

## **Principles of Assessment**

**Dr Saw-Choo Teo  
2009**

# Requirements/Principles of good assessment

1. **Accuracy** of measurement:
  - ✓ **Validity** = pertinence of measurement
  - ✓ **Reliability** = accuracy of measurement
2. **Practicality** of test administration:
  - ✓ 'easy/not too difficult to administer'
  - ✓ has quality of 'economy'
3. **Beneficial backwash**
  - ✓ **Beneficial effect on teaching:**
  - ✓ **Teaching and testing - a symbiotic (mutually supportive) relationship**

# Validity

**Validity** relates to **relevance** or **pertinence** to the **subject/area tested**: Does the test actually test what it claims to test?

# Content validity

- 1. Content validity** – relates to the content of the test
  - ✓ The test should contain a representative sample of ALL the elements/skills of the field/subject it is meant to test
  - ✓ The greater the content validity, the more accurate/reliable the test

# Construct validity

2. **Construct validity** – relates to **what exactly** is being tested.
- ✓ A test has ‘construct validity’ when it is clear that it measures **just** the ability or skill it is supposed to measure

# Predictive validity

- 3. Predictive validity** refers to the degree to which a test can predict a candidate's future performance.

# 'Face' validity'

4. 'Face' validity – in terms of how the test 'looks': e.g. 'It *is* a Maths test.'

# Measuring validity

**In statistical terms,**

- ✓ **perfect validity**
- ✓ **zero/no validity**

# Reliability

**Reliability (= accuracy)** : How **precisely/accurately** does the test measure what the candidate is capable of doing what s/he is being assessed on?

- ✓ The smaller the **variability** in the performance/score of a candidate in multiple sittings of a test, the more reliable the test is.

# Reliability

## 2 aspects to Reliability

1. **Reliability** in terms of **student performance**

2. **Marker/scorer reliability**

# Reliability

## Measures to achieve **student performance reliability**

### 1. Length of test

# Reliability

2. Restrict **choice** – the more similar the test taken by all, the more reliable it is.
3. Clear and unambiguous **instructions**.
4. Clear and unambiguous **items**.
5. Identical test administration.
6. Familiar format of testing – provide **examples** to help.

# Reliability and cognitive load

- **Long-term and short-term/working memory**
- **Too much imposition on working memory undermines construct validity**

# Scorer/Marker reliability

## Measures to achieve **scorer/ marker reliability**

- 1. The less the qualitative judgment required, the more reliable the test.**
- 2. Facilitate marking - minimise choice for candidates.**
- 3. Provide detailed scoring key, with little chance for variability.**

# Scorer/Marker reliability

4. **Train scorers/markers well.**
5. **Enforce coordination.**
6. **Provide for multiple impression marking or analytic marking – i.e. criterion referencing.**
7. **Provide anonymity for candidates.**

# Measuring Reliability

**In statistical terms,**

- ✓ **perfect reliability**
- ✓ **zero/no reliability**

# Measuring skills and not content (1)

**Diagnostic tests target skills or abilities not content areas in a syllabus.**

**Achievement tests seek to measure what has been 'achieved' out of a program of learning/teaching**