European Board of Neurological Surgery Examination

Peter Whitfield

Plymouth, UK

THE EUROPEAN ASSOCIATION
OF NEUROSURGICAL SOCIETIES

Validity Evidence – Downing, 2003

- Validity is the evidence used to support or refute a hypothesis about the meaning of assessment results.

- Validity is appropriate to the selected population

Validity: on the meaningful interpretation of assessment data

Steven M Downing

Context All assessments in medical education require evidence of validity to be interpreted meaningfully. In contemporary usage, all validity is construct validity, which requires multiple sources of evidence; construct validity is the whole of validity, but has multiple facets. Five sources – content, response process, internal structure, relationship to other variables and consequences – are noted by the *Standards for Educational and Psychological Testing* as fruitful areas to seek validity evidence.

Purpose The purpose of this article is to discuss construct validity in the context of medical education and to summarize, through example, some typical sources of validity evidence for a written and a performance examination.

Summary Assessments are not valid or invalid; rather, the scores or outcomes of assessments have more or less evidence to support (or refute) a specific interpretation (such as passing or failing a course). Validity is approached as hypothesis and uses theory, logic and the scientific method to collect and assemble data to support or fail to support the proposed score interpretations, at a given point in time. Data and logic are assembled into arguments – pro and con – for some specific interpretation of assessment data. Examples of types of validity evidence, data and information from each source are discussed in the context of a high-stakes written and performance examination in medical education.

Conclusion All assessments require evidence of the reasonableness of the proposed interpretation, as test data in education have little or no intrinsic meaning. The constructs purported to be measured by our assessments are important to students, faculty, administrators, patients and society and require solid scientific evidence of their meaning.

Keywords Education, Medical, Undergraduate/ *standards, Educational measurement, Reproducibility of results.

Introduction

The purpose of this paper is to discuss validity in the context of assessment in medical education and to present examples of the five types of validity evidence typically sought to support or refute the valid interpretations of assessment data.[1] This essay builds on and expands the older and more traditional view of test validity expressed in the first article in this series[2] and extends the validity discussion into state-of-the-art 21st century educational measurement.

Validity refers to the evidence presented to support or refute the meaning or interpretation assigned to assessment results. All assessments require validity evidence and nearly all topics in assessment involve validity in some way. Validity is the *sine qua non* of assessment, as without evidence of validity, assessments in medical education have little or no intrinsic meaning.

Validity is always approached as hypothesis, such that the desired interpretative meaning associated with assessment data is first hypothesized and then data are collected and assembled to support or refute the validity hypothesis. In this conceptualization, assessment data are more or less valid for some very specific purpose, meaning or interpretation, at a given point in time and only for some well-defined population. The assessment itself is never said to be 'valid' or 'invalid' rather one speaks of the scientifically sound evidence presented to either support or refute the proposed interpretation of assessment scores, at a particular time period in which the validity evidence was collected.

In its contemporary conceptualization,[1,3–14] validity is a unitary concept, which looks to multiple sources of

Department of Medical Education (MC 591), College of Medicine, University of Illinois at Chicago, Chicago, Illinois, USA

Correspondence: S M Downing, University of Illinois at Chicago, College of Medicine, Department of Medical Education (MC 591), 808 South Wood Street, Chicago, Illinois 60612-7309, USA. Tel.: +1 312 996 6428; Fax: +1 312 413 2048, E-mail: sdowning@uic.edu

Validity Evidence
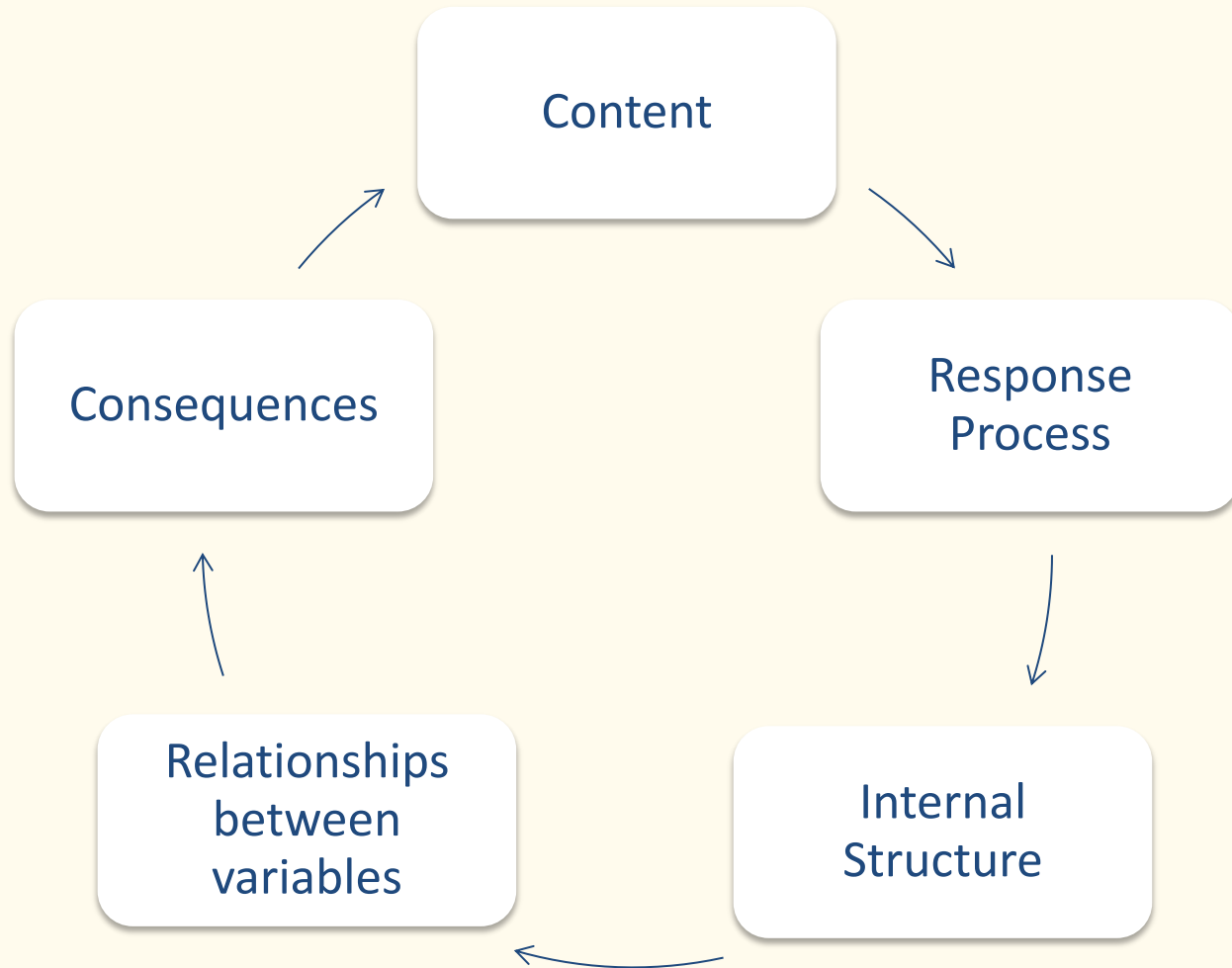
Validity: on the meaningful interpretation of assessment data

Construct

- Use of a blue print to map to curriculum
- Construct of effective questions
- Ambiguity, cultural and flaws addressed
- Sufficient questions to sample target domains
- Are clinical cases appropriate and sufficient in number?

Response Process

- Marking process – optical/ digital
- Transcription of results
- Summation of results
- QA of clinical components

Validity Evidence

Internal Structure

- Psychometric assessments of reliability (reproducibility) and generalisability
- Item analysis – difficulty, discrimination
- Reproducibility of a pass-fail score

Relationship to other variables

- To evaluate correlations between scores eg written/ clinical and overall
- Correlations with clinical cases
- Divergence provides evidence of testing different domains

Validity Evidence

Consequences of assessment

- On teaching and learning

- On practise

- Risks of harm – false positives and false negatives

- On candidates

- On wider healthcare

Validity: on the meaningful interpretation of assessment data

Content

  Theory test

  Practical test – traffic lights, crossing, roundabouts, dual carriageway, maneuvering,

Response process

  Written

  Practical

Internal structure

  Pass fail criteria

  Marking system

  Reproducibility between examiners

Relationship to other variables

  Numbers of accidents in first year of driving

  Traffic offences

  Ability to drive other vehicles

Consequences

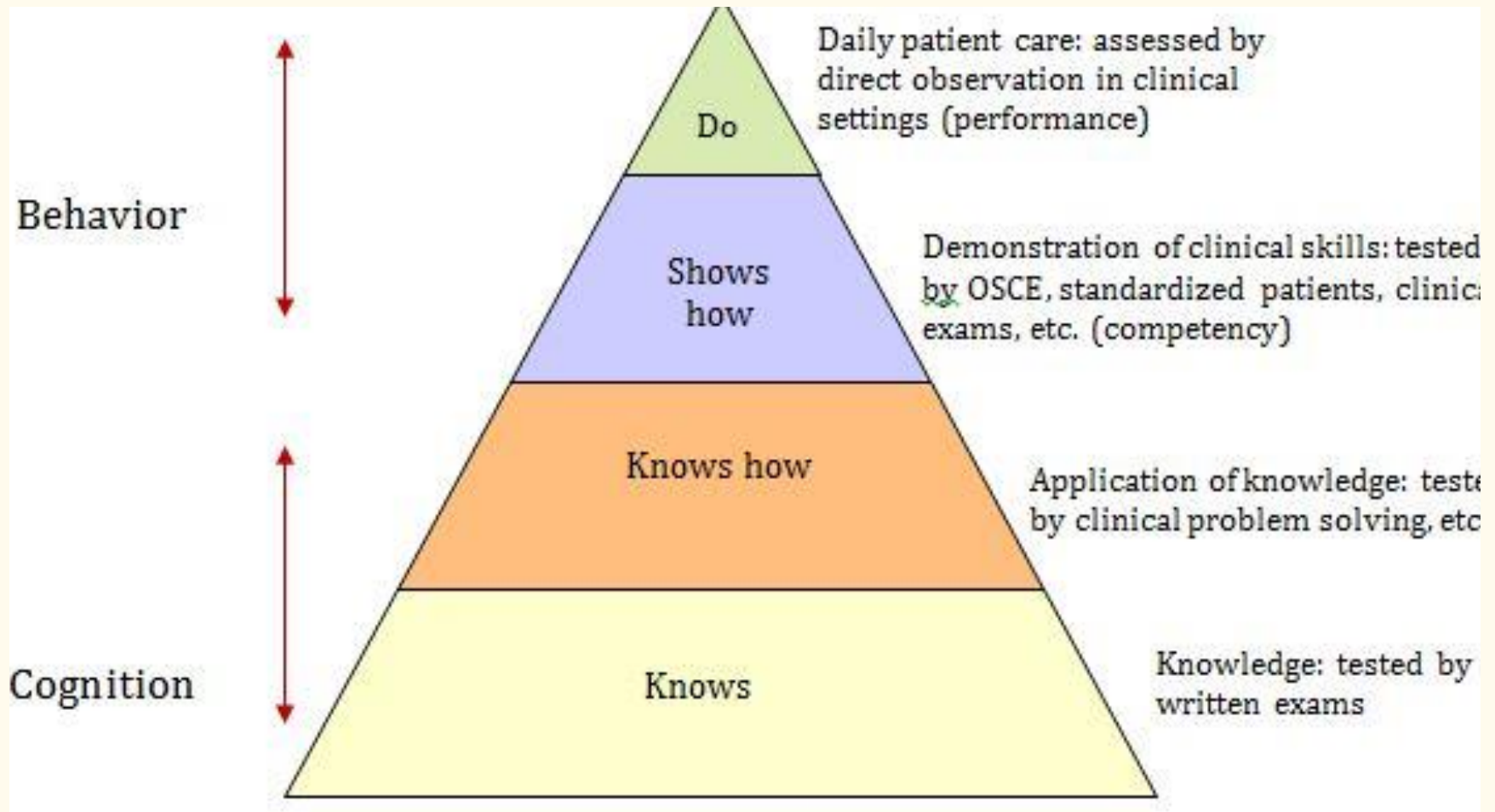 To the driver

  To passengers

  To public

16TH TIME L

Neurosurgery Part I

- 2 MCQ papers

- Single best answer

- In total: 150 questions with a focus on applied knowledge and management responses

- Content blueprint – domains include anatomy, physiology, neurotrauma, vascular, CSF, spine, infection, miscellaneous

Complexity of Assessment Tool

eg Neuroanatomy

When undertaking a trans-callosal approach for resection of a colloid cyst of the third ventricle, which of the following structures is the most useful landmark for orientation within the lateral ventricle?

A. Body of the thalamus
B. Head of the caudate nucleus
C. Septal vein
D. Septum pellucidum
E. **Thalamostriate vein**

eg Neurotrauma

A 34 year old man is assessed in the emergency department following a high speed road traffic accident. His pulse is 110, BP 105/75, respiratory rate 25/min. He has a GCS 5. Pupils are 2mm in diameter and react sluggishly to light. Following intubation and ventilation, CT imaging revealed a 1.5 cm thick acute subdural haematoma with 5mm of midline shift, a moderate left sided (3cm deep) pneumothorax,, a probable liver laceration with a small amount of intraperitoneal free fluid, a displaced fracture of left acetabulum and a left tibial plateau fracture.

What is the most appropriate next step in his management?

A. Craniotomy and evacuation of the acute subdural haematoma
B. Emergency laparotomy
C. External fixation of the pelvis
D. **Insertion of a chest drain**
E. Open reduction and internal fixation of the tibial fracture

Part I - Standards

- Psychometric assessment
- Some questions omitted based on psychometrics
- Compromise method - adjusted norm referencing- Board members determine pass mark rather than statistical parameters (eg Hofstee)
- Pass mark around 55-60%

Plans

- MCQ databank
- Given "high stakes" should criterion referencing (eg Angoff) be used?
- Should Angoff scoring be undertaken when items enter data bank?

Part II – Oral exam

Brain

Spine

Discussion

Brain and Spine
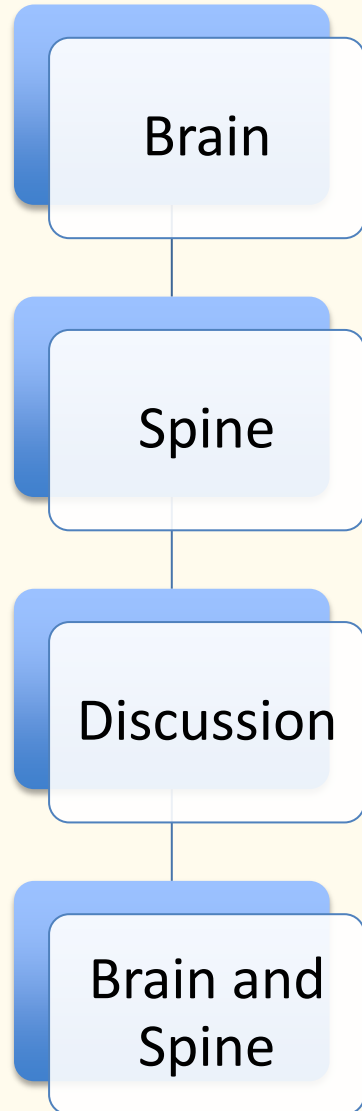
- Each oral 30 minutes. >6 cases discussed covering diagnosis/ investigation/ treatment

- Content list carried by candidates

- 2 examiners/ Independent marking at each station

- Max score 24. Passmark of 17.

Marking Criteria

Rating	Overall Professional Capability / Patient Care	Applied Knowledge & Judgment	Quality of Response
1	Candidate demonstrated incompetence in diagnosis and clinical management	Poor basic knowledge/ judgment, limited understanding	Confused, lacks insight, doesn't accept prompts.
2	Candidate failed to demonstrate competence in diagnosis and clinical management	Borderline knowledge, difficulty prioritising, poor deductive skills, sig. errors in clinical management	Hesitant, indecisive, frequent prompts needed: use of default questions
3	Candidate demonstrated competence in diagnosis and clinical management of common problems	Good knowledge & judgement of common problems, no major errors	Methodical approach: insight, minimal prompting needed.
4	Candidate demonstrated confidence and competence in diagnosis and clinical management	Good higher order thinking. Prioritises well, interpretation & judgement strong	Logical answers, good evidence for views; clear responses without prompting

Introduction	Welcoming and enabled applicant to relax
Clarity of questioning	Clear, precise. Appropriate case complexity.
Complexity of questioning	Appropriate escalation of difficulty of questioning.
Use of prompts	Used appropriately when candidate struggling.
Active listening	Good eye contact, attentive. Clarified candidate understanding of the questions appropriately. Please try and minimize saying "very good" – might allow candidate to feel they are doing well when in fact the questioning is at a very basic level.
Time management	Difficult candidate – very slow progress needing continual prompts. Although the first case >10 minutes it did allow exploration of different areas of relevance (eg interpretation of scans, pathology, operative management, implant choice). Please ensure marking reflects coverage of multiple topics to identify areas of weakness and breadth of strengths.
Marking	Entirely appropriate use of range of marks

Construct Validity

Downing SM (2003) Validity: On the meaningful interpretation of assessment data. *Medical Education* 37, 830-837.