



MIAMI 2010

OTTAWA CONFERENCE ON THE ASSESSMENT OF COMPETENCE IN MEDICINE AND THE HEALTHCARE PROFESSIONS



- AMEE
- 1200 delegates; international
- 3-day meeting (& pre-conference workshops)
 - Plenary (3)
 - Symposia (8)
 - Oral Presentations (>50)
 - Workshops (>60)
 - Poster Discussions (>20)



- Measurement of assessment
- Workplace assessment
- “Difficult student”
 - Academic
 - Professionalism



Measurement of Assessment

- Workshops
 - Overview of validity & reliability
 - Generalisability Theory
- Oral presentations



Validity

- Revised definitions (1999)
 - “degree to which evidence and theory support the interpretation of test scores entailed by the proposed use of tests”
 - Represented by a single concept – “construct validity”
 - the degree to which a score can be interpreted as representing the intended underlying construct
 - Not a property of the test instrument but of the interpretation of the instrument’s scores



Evidence for validity

- Content
 - Relationship between a test's content and its intended construct
 - Test blueprint
 - Quality of written questions
 - Items adequately represent the domain
- Response process
 - Test produces the intended response in the test-takers
 - “Think-aloud processes”; test-taker
 - Raters can correctly measure the intended responses
 - Rater training



- Internal structure
 - Scores intended to measure a single construct should yield homogenous results
 - Internal reliability
 - Factor analysis
- Relations to other variables
 - Correlates with scores testing same construct with different instrument
 - Correlation coefficients



- Consequences
 - Intended and unintended consequences of test results
 - Bias
 - Standard setting
- “Face validity”
 - Judgement based on appearance of instrument
 - Differs from content validity
 - No longer considered appropriate evidence



Reliability

- Refers to the reproducibility or consistency of scores from one assessment to another
- Necessary but not sufficient for validity



MIAMI 2010

Evidence of reliability

- Internal consistency
 - All the items measure the same construct
 - Cronbach-alpha
 - Kuder-Richardson (dichotomous responses only)
- Temporal stability
 - Instrument produces a similar result when retested
 - Test-retest reliability; Pearson correlation coefficient
- Agreement (inter-rater reliability)
 - Different raters produce similar results
 - Kappa
 - Intraclass correlation coefficient



- Generalisability theory
 - Attributes the unreliability to possible sources of error (variance) e.g. item, rater, case.
 - Uses analysis of variance to quantify the contribution of each source of error (G study)
 - Apply data to determine the most reliable test format e.g. number of stations, number of examiners per station, number of cases of similar type (D study)



Workplace Assessment

- Symposium
 - Workplace assessment
- Workshops
 - Calibrating workplace assessment
 - Subjectivity in assessment
 - Unintended consequences



MIAMI 2010

Workplace assessment

- Assessment of or for learning
 - Early, frequent, feedback
 - Lower reliability versus validity/educational impact
 - Improved by repeated measures
 - Formative versus summative
 - Single test – formative
 - Multiple tests – summative (demonstrate improvement); problematic given clinical examiners tendency to high grades



Measuring workplace performance

- Instruments
 - Global ratings preferable to checklists
 - Anchored on relevant expected standard; educationally defined standards are not familiar to clinical raters; prefer clinically anchored scales e.g. RIME
- Raters
 - Stringency (hawk/dove)
 - Subjectivity
 - Account for 60-70% of variance



- Subjectivity in assessment
 - Unavoidable, and even preferable, in some assessments.
 - Value of subjectivity in assessment increases by
 - Number of judgements
 - Independence of judgements
 - Diversity of perspectives
 - Aim to acknowledge diversity/independence and not try to reach consensus



- Unintended consequences
 - Changes in student behaviour to obtain better grades
 - Mini-CEX
 - Avoid difficult cases/domains/raters
 - Develop “grade-orientated” skills
 - RIME
 - Skip lower level standards of performance
 - Try to show higher levels (e.g. educator)



Difficult Student

- Workshops
 - Student in academic difficulty
 - Unprofessionalism
- Described local approaches illustrated by case studies



Academic difficulty

- Cognitive
- Affective; problem with adjustment (academic, personal)
- Structural; poor study skills
- Interpersonal; isolated, unresponsive
- Psychomotor; language, technical skills
- Self-awareness
- Professional



- Professionalism
 - Informing students and staff of expectations
 - Identifying and reporting unprofessional behaviour
 - Systems for reporting – simple, low threshold, transparent, educative consequences
 - Investigating allegations
 - Triangulation
 - Use of retrospective information
 - Outcomes
 - Educative (remedial)
 - “Forward feeding”
 - Punitive



- Abstracts online
- Consensus statements
 - Criteria for a good assessment
 - Technology-based assessment
 - Performance assessment
 - Assessment for selection
 - Research in assessment