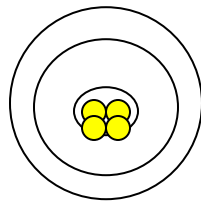


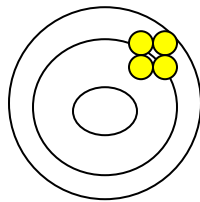
CHARACTERISTIC OF ASSESSMENT – RELIABILITY

1. Meaning of reliability

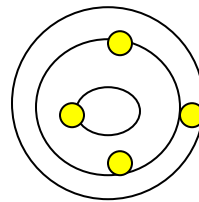
- Reproducibility/consistency of scores/measurement over time/occasion.
- Characteristic of result/outcome of the assessment, not the measurement instrument.
- Necessary but not sufficient condition for validity



Valid and reliable



Not valid but reliable



Not valid and not reliable

2. Relationship with validity, reliability and objectivity:

- Validity:
 - Content
 - Blueprinting
 - Item quality
 - Content expertise
 - Response process
 - Student briefing
 - Quality control at all steps
 - Internal structure
 - Item analysis
 - Reliability
 - Relationship to other variables
 - Correlations with other relevant variables
 - Convergent correlations
 - Divergent correlations

- Consequences
 - On the student
- Objectivity and reliability
 - Increased objectivity result lower reliability
 - E.g. essays, viva, rating
 - Strategies to reduce subjectivity → MCQs, checklist, etc: do they really improve reliability?
 - Not much
 - Lessons
 - More objective method is not necessarily more reliable or more superior
 - Choice of measurement method is firstly determined by educational context/purpose of the testing situation.

3. Types and determination of reliability:

- Reliability of achievement
 - Stability
 - Test – retest
 - Test – retest with equivalent forms
 - Internal consistency
 - Split-half
 - Coefficient alpha
 - Kuder-Richardson
- Reliability of rater data
 - Inter-rater consistency
 - Intraclass correlation coefficient → analysis of variance (ANOVA)
- Reliability of performance examinations
 - Inter-rater consistency

4. Factors affecting reliability

- Number of assessment task
 - The larger the number of tasks, the higher the reliability
 - Overcoming ‘content specificity’

- Spread of scores
 - The larger the spread, the higher the reliability
- Objectivity
 - Select procedures according to purpose, then make the procedures as objective as possible

5. Uses of reliability information

- Consequences of decision based on assessment
 - Very high stakes e.g licensure/certification: > 0.90
 - End of course/year summative exam: $0.80 - 0.89$
- Standard deviation (1-reliability)

6. Rigour vs. Practicality

- Rigour
 - Validity
 - Reliability
- Practicality
 - Feasibility
 - Ease of administration
 - Time required
 - Cost
 - Ease of interpretation and application
 - Acceptability
- A mixture of methods is preferable to relying on a single method
- Don't forget educational impact