

VALIDITY AND RELIABILITY OF ACHIEVEMENT ASSESSMENT

Presented

By

Dr / Said Said Elshama

Professor of Forensic Medicine and Clinical Toxicology

Medical Education (DHPE)

College of Medicine - Taif & Suez Canal University

LEARNING OBJECTIVES

- ✘ Distinguish between validity and reliability.
- ✘ Describe different evidences of validity.
- ✘ Describe methods of estimating test reliability.
- ✘ Mention factors that decrease exam validity and reliability and how to avoid them.



VALIDITY

- ✘ Appropriateness, meaningful and usefulness of results interpretations .
- ✘ It is unitary concept.
- ✘ It is expressed by degree
- ✘ It is specific use
- ✘ It is related to assessment using consequences
- ✘ It refers to inferences only , not the instrument
- ✘ It depends on different types of evidence.



TYPES OF VALIDITY EVIDENCE

- ✘ Content representativeness
- ✘ Criterion relationships
- ✘ Construct evidence
- ✘ Consequences of assessment using



CONTENT REPRESENTATIVENESS

- ✘ Adequate sampling
- ✘ Representative broad range of content
- ✘ Relevance content



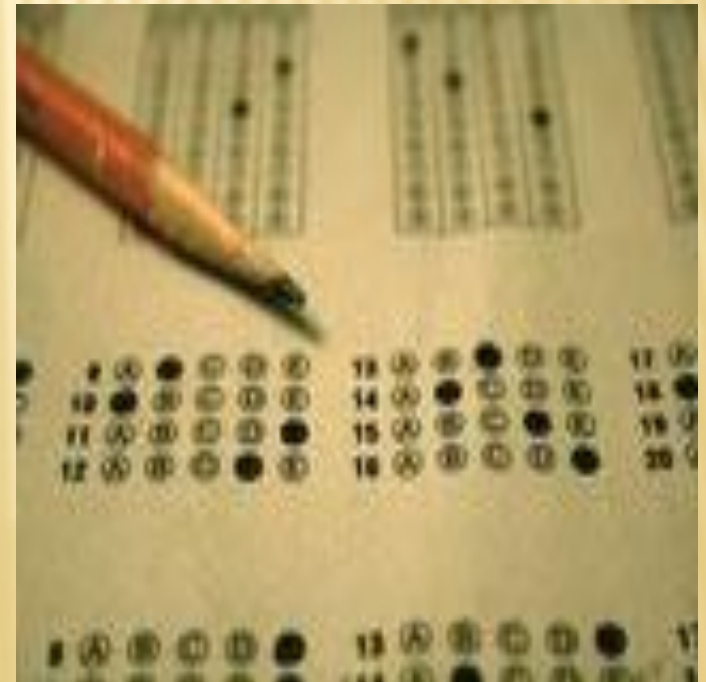
Application

We prepared the valid test content by :-

1. Assessed learning outcomes identification
2. Specification the sample of used tasks
3. Suitable assessment procedure for specifications

CRITERION RELATIONSHIPS

- ✘ Relationship degree between test scores and predicted or estimated criterion
- ✘ It is expressed by correlation coefficients or expectancy table
- ✘ Positive relationship
- ✘ Negative relationship



CONSTRUCT EVIDENCE

- ✘ It measures all possible influences on the scores
- ✘ It clarifies and verifies the inferences of results
- ✘ It depends on other different types of evidence (content, criterion)

Relevant and Representative



CONSEQUENCES OF ASSESSMENT RESULTS USING

- ✘ It is using the extracted inferences from assessment results

✘ Self assessment skills were improved by assessment results using?

✘ Independent learning was encouraged by assessment results using?

✘ Motivation was improved by assessment results using?

✘ Good study habits were encouraged by assessment results using?

✘ Performance was improved by assessment results using?



Application

TYPES OF VALIDITY

- ✗ Face validity
- ✗ Construct validity
- ✗ Content validity
- ✗ Criterion validity
 1. Predictive
 2. Concurrent
- ✗ Consequences validity



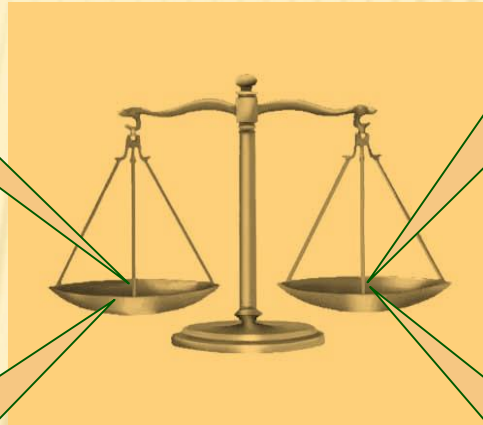
FACTORS LOWER VALIDITY

- ✘ Insufficient of assessed tasks sample
- ✘ Using of improper tasks types
- ✘ Using of irrelevant or ambiguity tasks
- ✘ Improper tasks arrangement
- ✘ Using of unclear directions
- ✘ Few tasks (test items)
- ✘ Improper administration (inadequacy of allowed time)
- ✘ Objectivity of scoring
- ✘ Inadequacy of scoring guides

WE **AVOID** FACTORS WHICH LOWER VALIDITY BY:-

- ✘ Relevance
- ✘ Balance
- ✘ Efficiency

- ✘ Specificity
- ✘ Difficulty
- ✘ Discrimination



- ✘ Proper administration
- ✘ Adequacy of scoring guides

Application

- ✘ Variability
- ✘ Clear directions

RELIABILITY

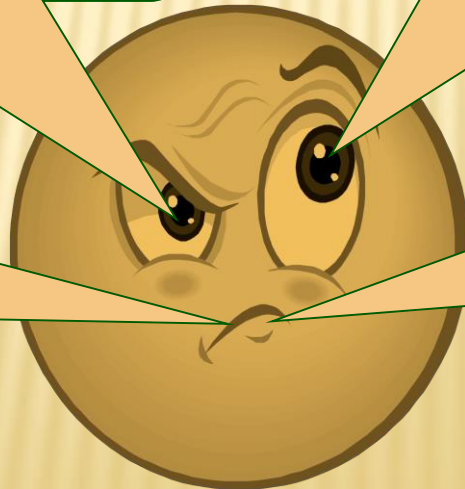
- ✘ It is consistency of results
- ✘ The scores free from measurement errors
- ✘ It is determined by reliability coefficient or standard error of measurement

✘ Generalizable results over similar samples of tasks, time periods and raters

✘ Using of different sample of the same type of task leads to the same results

✘ Different evaluators rate the performance assessment in the same way

✘ Using of assessment in the different time leads to the same results



ESTIMATION TEST SCORES RELIABILITY

× **Test –retest method**

Twice administrations of the same test to the same group with time interval in between

× **Equivalent –forms method**

Administration of two equivalent forms of the test to the same group in close succession during the same session

× **Test-retest with equivalent forms (combination)**

Administration of two equivalent forms of the same test with time interval in between

× **Internal consistency method**

One single administration of the test and account the response consistency within the test

FACTORS LOWER RELIABILITY HOW TO AVOID IT ?

Factors lower test scores reliability

- ✗ Few items of test scores
- ✗ Limited scores range
- ✗ Inadequacy of testing conditions
- ✗ Subjectivity of scoring

Avoid the factors which lower test scores reliability by the following:-

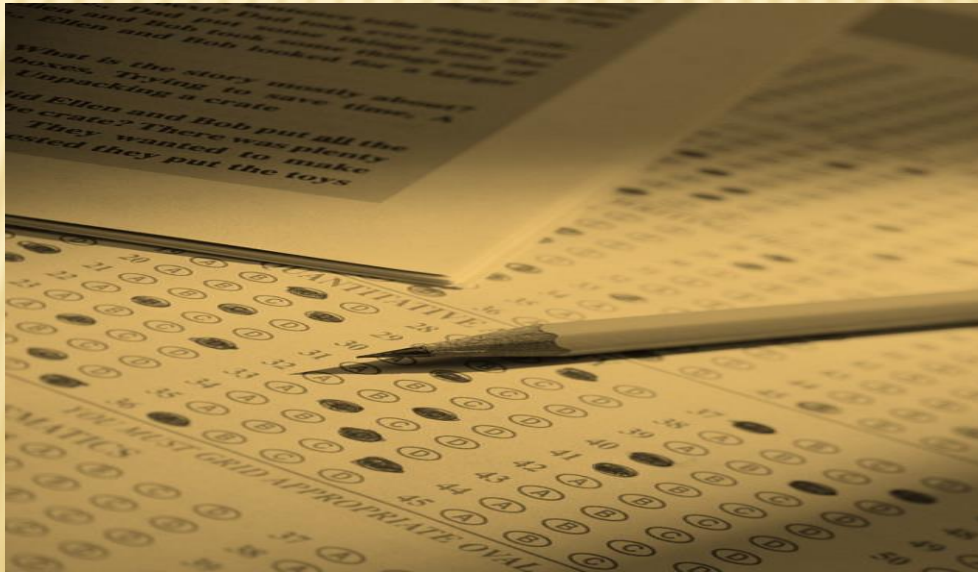
- ✓ Use longer tests
- ✓ Gather scores from many short tests
- ✓ Item difficulty adjustment for larger scores spread
- ✓ Arrange test administration time
- ✓ Interrupted factors elimination
- ✓ Scoring key preparation



✗ **Application**

ESTIMATION OF PERFORMANCE ASSESSMENT RELIABILITY

- 1) Comparison between two obtaining scores of two judges who scored the performance independently
- 2) Assessment performance reliability is percentage of agreement between two judges scores



Factors lowers performance assessments reliability

- ❖ Inadequacy of number of tasks
- ❖ Poor structured procedures of assessment
- ❖ Specific performance dimensions to tasks
- ❖ Insufficient scoring guides
- ❖ Personal bias affecting scoring judgments

Avoid the factors which lower performance assessments reliability by :-

- ✓ Gather results from many assessments
- ✓ Identify nature of tasks
- ✓ Identify scoring criteria
- ✓ Increase performance generalizability
- ✓ Use specific rating scales according to levels and criteria of quality
- ✓ Train for judging and rating
- ✓ Check scores with independent judge scores

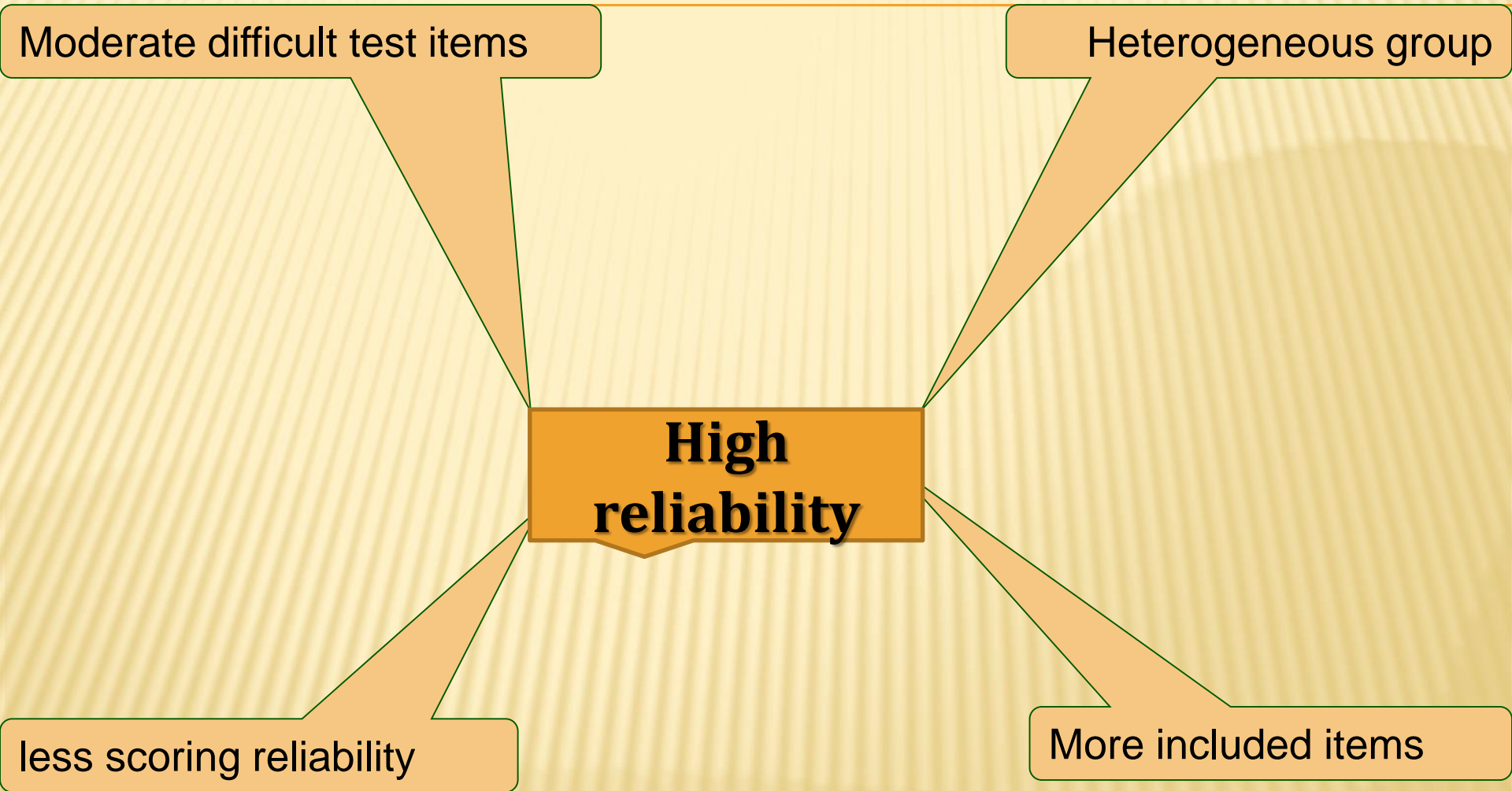


Performance Assessments Reliability

CONCLUSION

- ✘ Reliability refers to obtained results and instrument evaluation and not to instrument itself.
- ✘ Reliability refers to type of consistency.
- ✘ Reliability is a necessary but not sufficient for validity.
- ✘ Reliability provides the consistency which makes validity.

Conclusion



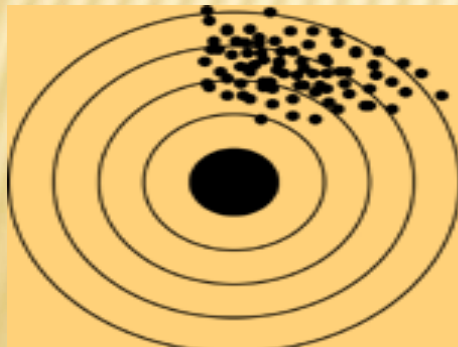
CONCLUSION

- ❖ Accuracy and appropriateness of results interpretations.
- ❖ High, moderate or low degree .
- ❖ Does the instrument measure what it tells it does?

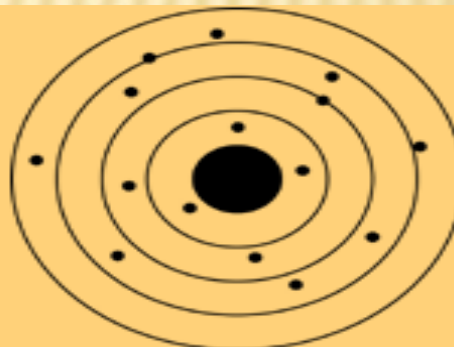
Factors lower validity

- Unclear directions
- Inadequate time limits
- Inappropriate difficulty level
- Poor constructed test items
- Test items inappropriate for measured outcomes
- Too short tests
- Improper arrangement of items (complex to easy)
- Administration and scoring
- Nature of criterion

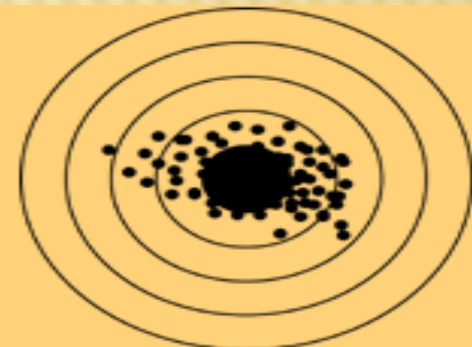
Validity



Reliable but Not Valid



Valid but Not Reliable



Valid and Reliable

REFERENCES

- 1- Gronlund, NE. (2006).** Assessment of Student Achievement. 8th Edition. Pearson, USA. Chapters 13.
- 2- Carter, DE, & Porter, S. (2000)** Validity and reliability. In Cormack D (ed.) The Research Process in Nursing. Fourth edition, Oxford, Blackwell Science. 29-42.
- 3- Knapp, TR. (1998).** Quantitative Nursing Research. Thousand Oaks, Sage.
- 4- Peat, J. (2002).** Health Services Research: A Handbook of Quantitative Methods. London, Sage.
- 5- Moskal , BM, & Leydens , JA. (2000).** Scoring rubric development: Validity and reliability. Practical Assessment, Research & Evaluation, 7(10). [Available online: <http://pareonline.net/getvn.asp?v=7&n=10>].

Thank you

Thank you



Thank you

Thank you