

# ***CONCEPTS OF VALIDITY & RELIABILITY***

***16<sup>TH</sup> June 2015  
UTeM Melaka***

# ***Good Assessments***

- Regardless of its other characteristics, the most important characteristics in determining the usefulness of assessment information are its validity and reliability.

# ***Good Assessments***

- Regardless of its other characteristics, the most important characteristics in determining the usefulness of assessment information are its validity and reliability.

# ***Validity***

Validity has been defined as referring to the appropriateness, correctness, meaningfulness, and usefulness of the specific inferences researchers make based on the data they collect.

# ***Validity***

- Validity is an evaluation of the adequacy and appropriateness of the interpretations and uses of assessment results.
- Validity is always determined by a judgment made by the test user.

# ***Nature of Validity***

- Appropriateness of the interpretation and use made of the results of an assessment procedure for a given group of individuals
- A matter of degree
- Specific to some particular use or interpretation
- A unitary concept
- Overall evaluate judgment

# *Sources of Validity*

- Test content
- Response process
- Internal structure

# Evidence of Validity

There are 3 types of evidence a researcher might collect:

## Content-related evidence of validity

Content and format of the instrument

## Criterion-related evidence of validity

Relationship between scores obtained using the instrument and scores obtained

## Construct-related evidence of validity

Psychological construct being measured by the instrument



# ***Content Validity***

- Compare the assessment tasks to the specifications describing the task domain under consideration
- The extent to which an assessment procedure adequately represents the content of the assessment domain being sampled

# ***Content Validity***

## Example:

- The content validity of the instrument has been validated by four HR managers to identify the relevancy of the human resource management practices in automobile industry.
- Their CV were attached together in Appendices 2 for reference.

# *Example*

- After the pilot studies, the PES instrument was subjected to content validation by a panel of senior teachers and educationists (Credentials enclosed in Appendix G).
- All members of the panel worked independently. Each was presented with a set of the instrument and informed of the purpose of the instrument.
- They were then requested to study the items and decide on the suitability of the items. They were also asked if any other items should be included to fulfil the purpose of the instrument, and to comment on any part of the scale's items that they felt needed amending or clarification.

# *Example*

## **Appendix G: Credentials of Panel for Instrument Validation Name, Qualification and Experience**

### **1. DR. FOO SAY FOOI**

He is a lecturer in Educational Administration, Faculty of Educational Studies, Universiti Putra Malaysia.

### **2. ASSOC. PROF. DR. TURIMAN SUANDI**

A lecturer in the Faculty of Educational Studies, Universiti Putra Malaysia. Among others, he lectures on Research Design. Presently Deputy Dean for Development and Students.

### **3. ASSOC. PROF. DR. ZAIDATOL AKMALIAH L. PIHIE**

A lecturer in the field of Educational Administration at the Faculty of Educational Studies, Universiti Putra Malaysia. She had written articles and books related to educational management and administration. She is the Deputy Dean in the FPPUPM.

# ***Criterion Validity***

- Compare assessment results with another performance obtained at a later date or with another measure of performance obtained concurrently (for estimating present status).
- The degree to which performance on an assessment procedure accurately predicts a student's performance on an external criterion

# ***Construct Validity***

- Establish the meaning of assessment results by controlling the development of the assessment, evaluate the cognitive procession used by students to perform tasks, evaluate the relationships of the scores with other relevant measures, and experimentally determine what factors influence performance.
- The extent to which empirical evidence confirms that an inferred construct exists and that a given assessment procedure is measuring the inferred construct accurately

# ***Factors Influence Validity***

- Unclear directions
- Reading vocabulary and sentence structure too difficult
- Ambiguity
- Inadequate time limits
- Overemphasis of easy-to-assess aspects of domain at the expense of important but difficult-to-assess aspects

# ***Factors Influence Validity***

- Test items inappropriate for the outcomes being measured
- Poorly constructed test items
- Test too short
- Improper arrangement of items
- Identifiable pattern of answers
- First impression
- Personal theories



# ***Reliability***



- Refers to the consistency of measurement, that is, how consistent test scores or other assessment results are from one measurement to another.

# ***Nature of Reliability***

- Refers to the results obtained with an instrument and not to the instrument itself
- An estimate of reliability always refers to a particular type of consistency (time, task, students, rater)
- Reliability is a necessary but not sufficient condition for validity
- Reliability is assessed primarily with statistical indices

# ***Terminology***

- Correlation Coefficient: a statistic that indicates the degree of relationship between any two sets of scores obtained from the same group of individuals

# *Methods of estimating Reliability*

- Test-retest
- Equivalent-forms
- Internal Consistency
  - Split-half
  - Kuder-Richardson
  - Alpha Coefficient
- Interrater



What method?

# ***Test-Retest***

- Give the same test twice to the same group with some time interval between tests, from several minutes to several days

# ***Equivalent Forms***

- Give two forms of the test to the same group in close succession



# ***Test-Retest with equivalent forms***

- Give two forms of the test to the same group with an increased time interval between forms

# ***Internal-Consistency Methods***

There are several internal-consistency methods that require only one administration of an instrument.

Split-half Procedure: involves scoring two halves of a test separately for each subject and calculating the correlation coefficient between the two scores.

Kuder-Richardson Approaches: (KR20 and KR21) requires 3 pieces of information:

- Number of items on the test

- The mean

- The standard deviation

Considered the most frequent method for determining internal consistency

Alpha Coefficient: a general form of the KR20 used to calculate the reliability of items that are not scored right vs. wrong.



# ***Internal Consistency***

- If Likert scales are used to represent the response choices, analysis for internal consistency can be accomplished using Cronbach's Alpha.
- Job Satisfaction Questionnaire

No.	Component	Cronbach Alpha
1	Intrinsic Factors	.75
2	Extrinsic Factors	.83

# ***Interrater***

- Give a set of student responses requiring judgmental scoring to two or more raters and have them independently score the responses

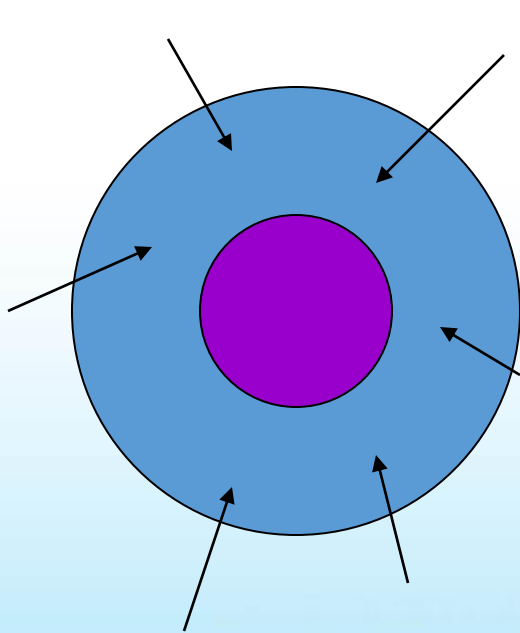
# ***Factors Influencing Reliability Measures***

- Number of assessment tasks
- Spread of scores
- Objectivity

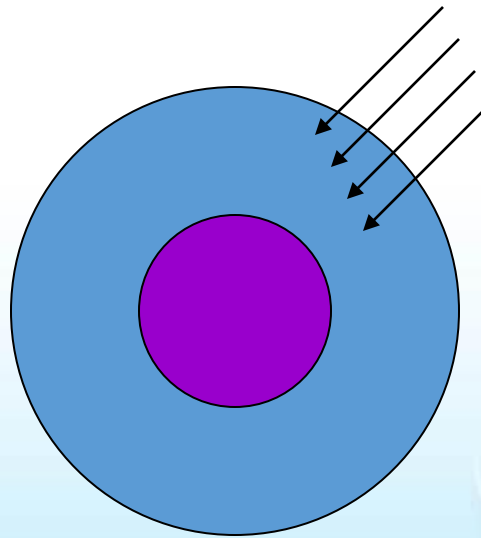
# ***Validity-Reliability***

- Reliability is a necessary but insufficient condition for validity.
- Reliability (consistency) of measurement is needed to obtain valid results, but we can have reliability without validity.

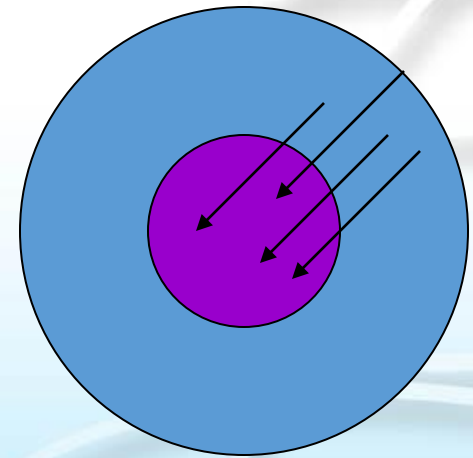
# ***Interactions of Validity and Reliability***



**Lack of validity  
and reliability**



**Good reliability  
But poor validity**



**Good validity  
and reliability**

# ***Other Factors Influencing Validity and Reliability***

# *Objectivity*

- Refers to the degree to which equally competent scorers obtain the same results.
- The test items (objective type)
- The results (not influenced by scorer's judgment or opinion)

# ***Practicality***

- Economical from the viewpoint of both time and money.
- Easily administered and scored.
- Produce results that can be accurately interpreted and applied by available school personnel



# *Usability*

- Ease of administration
- Time required for administration
- Ease of interpretation and application
- Availability of equivalent or comparable forms
- Cost of testing

# ***The Standard Error of Measurement***

- The index used in educational assessment to describe the consistency of a particular person's performance(s)
- Reflection of the consistency of an individual's scores
- The higher the reliability of a test, the smaller the SE of measurable will be

# ***Teachers' Ethical Responsibilities Regarding Assessment***

- Make fair and impartial decisions
- Construct and administer fair and clear assessments
- Motivate pupils to do their best
- Teach pupils the varied types of assessments
- Provide opportunities for pupils to practice test approaches
- Make reasonable accommodations for students with disabilities

# *Summary*

- Concept of Validity
- Factors influencing validity
- Concept of Reliability
- Factors influencing reliability



# *Questions*

How important is the concept of ‘validity’ in classroom assessment?

To what extent do reliability concerns affect the construction of a good test?

# *References*

- Linn, R. L., & Miller, M. D. (2005). Measurement and assessment in testing. International Edition. 9<sup>th</sup> Ed. New Jersey: Pearson Merrill Prentice Hall
- Airasian, P.W. (2005). Classroom assessment: Concepts and applications. 5<sup>th</sup> Ed. Boston: McGrawHill.