



自資高等教育聯盟

Federation for Self-financing  
Tertiary Education

## Module 3A

# **What is educational assessment and the most common assessment tool: Written Test**

Mr. Howard Sou  
hhpsou@gmail.com

Afternoon session August 2015  
Federation for Self-financing Tertiary Education

# Brief Recap of Morning Session

- Assessment is the collection of data for [educational] decision making.
- The meaning of assessment, test and measurement
- Two major categories of item types: Objective Test Items and Performance Tasks

# Focus of the PM Session

- Alignment of Assessment with learning outcomes
- Planning of assessment: Programme level
- Planning of achievement test
- Quality assurance of assessment: item level and test level

# The process of preparing, administering and using assessment

- ▶ 8. Interpreting and using the results
- ▶ 7. Appraising the assessment
- ▶ 6. Administering the assessment
- ▶ 5. Assembling the assessment
- ▶ 4. Preparing relevant assessment tasks
- ▶ 3. Selecting appropriate assessment tools
- ▶ 2. Developing specifications
- ▶ 1. Determining the purpose of assessment

# Determining the purpose of assessment

What is the intended purpose of assessment?

- Accountability?
- Programme improvement?
- Instruction improvement?
- Certification?

# Developing specifications

Developing a *Table of Specifications* (or blue print of assessment) to identify the achievement domains being measured and to ensure that a fair and representative sampling of attributes and/or questions.

It also helps to ensure the *Content Validity* of the assessment

# Table of specification (Programme)

Mode of assessment  Module/Content	End of Module Test	End of Programme Examination	Project	Programme paper/thesis	Tutor evaluation: exercises and participations etc		Total
Module/Topic A							
Module/Topic B							
Module/Topic C							
Module/Topic D							
Total							

# Table of specification (Achievement test)

Cognitive levels	Facts and information	Concepts	Analysis	Synthesis	Applications		Total
Content							
Topic A							
Topic B							
Topic C							
Topic D							
Total							



# Matching Assessment objectives with tools

Objectives	Tools
Recall, Recognize, Identify	Objective test items such as fill-in-the-blank, matching, labeling, or multiple-choice questions
Interpret ,Classify, Summarize , Explain	Papers, exams, problem sets, class discussions, or concept maps
Analyze, Differentiate, Organize	Case studies, critiques, labs, papers, projects, debates, or concept maps

# Norm, Criterion and Standards referenced tests

- ▶ ***Norm-referenced tests (NRTs)*** are made to compare test takers to each other.
- ▶ ***Criterion-referenced tests (CRTs)*** are intended to measure how well a person has learned a specific body of knowledge and skills
- ▶ A recent variation of criterion-referenced testing is "***Standards-referenced testing (SRTs)***" which describes what students should know and be able to do in different subjects at various grade levels.

- ▶ What are the information conveyed by:
  - Norm referencing test scores?
  - Criterion referencing test scores?
  - Standards referencing test scores?

# Z Score

$$\text{Z Score} = \frac{x - \mu}{\sigma}$$

$$\text{Z Score} = \frac{\text{Raw score} - \text{Mean}}{\text{Standard deviation}}$$

# Standard Deviations

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

$\sigma$  = standard deviation

$\sum$  = sum of

$x$  = each value in the data set

$\bar{x}$  = mean of all values in the data set

$n$  = number of value in the data set

# Questions for Reflection

- ▶ What are the information and messages carried by a certificate? a result slip?
- ▶ If I am going to admit students into a programme of studies, how can I make use of public examination results as a tool for selection?
- ▶ If I am going to recruit a staff, how can I make use of public examination results as a tool for selection?

# Quality of a test

- ▶ Reliability - refers to the consistency of measurement;
- ▶ Validity - refers to whether the assessment measures what intended to measure.

Both reliability and validity refer to the results obtained with an assessment instrument and **NOT** to the instrument itself.

# Reliability

- ▶ Reliability is primarily statistical

## Methods of estimating reliability

- ▶ Test-retest method (measure of stability)
- ▶ Split-half method (measure of internal consistency)
- ▶ Kuder-Richardson method and coefficient Alpha (measure of internal consistency)
- ▶ Inter-rater method (measure of consistency of rating)



# Validity

- ▶ Content validity – How well the sample of assessment tasks represents the domain of tasks to be measured
- ▶ Construct validity – How well performance on the assessment can be interpreted as a meaningful measure of some characteristic or quality
- ▶ Test-Criterion Relationship – How well performance on the assessment predicts future performances or estimates current performance on some valued measures other than the test itself.

# Quality of an Item

- ▶ Difficulty index: the percentage of participants who get the item right
- ▶ Discriminating index: the degree to which it discriminate between students with high and low achievement.

# Difficulty index

- ▶  $P = (R/T) \times 100\%$



- ▶ Where  $P$  = Difficulty index

- ▶  $R$  = number of participants who  
get the item right

- ▶  $T$  = total number of participants.



- ▶  $P = 100\%$  implies the item is too easy (everyone get it right)

- ▶  $P = 0\%$  means the item is too difficult (no one get it right)

- ▶ Usually,  $P = 0.4 - 0.8$  is acceptable

# Discriminating index

▶  $D = (RU/T(U)) - (RL/T(L))$

- ▶  $D$  = Discriminating index
- ▶  $RU$  = number of participants in the upper group who get the item right
- ▶  $RL$  = number of participants in the lower group who get the item right
- ▶  $T(U)$ ,  $T(L)$  = total number of participants in the upper and lower groups

# Discriminating index

- ▶ Upper group refers to participants whose scores of the whole paper are at the top 25%
- ▶ Lower group refers to participants whose scores of the whole paper are at the bottom 25%



# Discriminating index

- ▶  $D = 1$  implies the item has high positive discrimination power (i.e. that all upper group participants get it right while all lower group participants get it wrong)
- ▶  $D = 0$  implies the item has no discrimination power (i.e. both the upper and lower group participants either all get it right or all get it wrong)
- ▶  $D = -1$  implies the item has high negative discrimination power (i.e. the all upper group participants get it wrong while all lower group participants get it right)
- ▶ Usually,  $D = 0.2 - 0.3$  is acceptable

- ▶ Questions
- ▶ Comments
- ▶ Sharing of experiences
- ▶ are welcomed !!!

# Thank You!

