TAKLIMAT PRINSIP PEMBINAAN ITEM PENTAKSIRAN

FAKULTI SAINS
5 DIS 2021

ROHAYA TALIB
PhD (Measurement & Evaluation)
M.Ed (Measurement & Evaluation)
BBA (Accounting)

THE OBJECTIVES

At the end of this course, participants are expected to be able to;

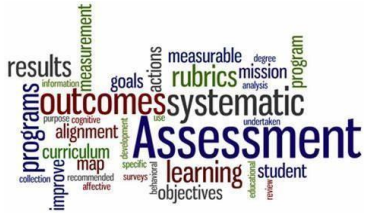
- ✓ develop test blueprint
- ✓ apply the good principles of item development
- ✓ develop items according to the level of cognitive specified on the CLOs based on Bloom Taxonomy

  - ❖ MCQs
  - ❖ Essay Qs

PENTAKSIRAN

A well-constructed comprehensive assessment system provides continuous, coherent, and **high-quality information on student performance** that educators and administrators could use to improve teaching and learning and meet their decision-making needs

(Rhode Island Department of Education & the National Center for the Improvement of Educational Assessment)



CURRICULUM DEVELOPMENT

Tyler's Objective Model

Tyler's model is a linear model and the 'ends-means' model.

OBE
OUTCOME-BASED EDUCATION

Education System that focus /emphasis on the **development** and **achievement** of student's **outcomes**

OBE

Starting with a clear picture of what is important for **students to be able to do** then **organising** the **curriculum, instruction** and **assessment** to make sure that the learning is ultimately happens (Spady, 1994)

LEARNING OUTCOMES

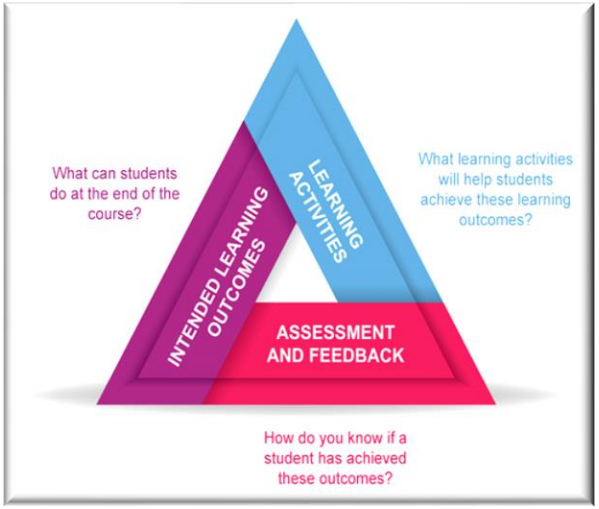
Learning Outcomes are **statements of what a learner is expected to know, understand and/or be able to demonstrate after completion of a process of learning.** (Source : ECTS Users' Guide, 2005)

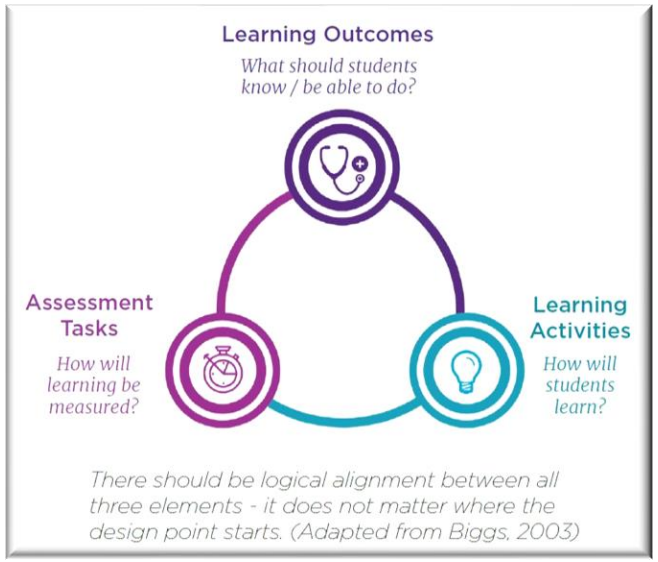
Hence….

- Learning outcomes focus on **what the learner has achieved** rather than the intentions of the teacher;
- Learning outcomes focus on **what the learner can demonstrate** at the end of a learning activity.

CONSTRUCTIVE ALIGNMENT

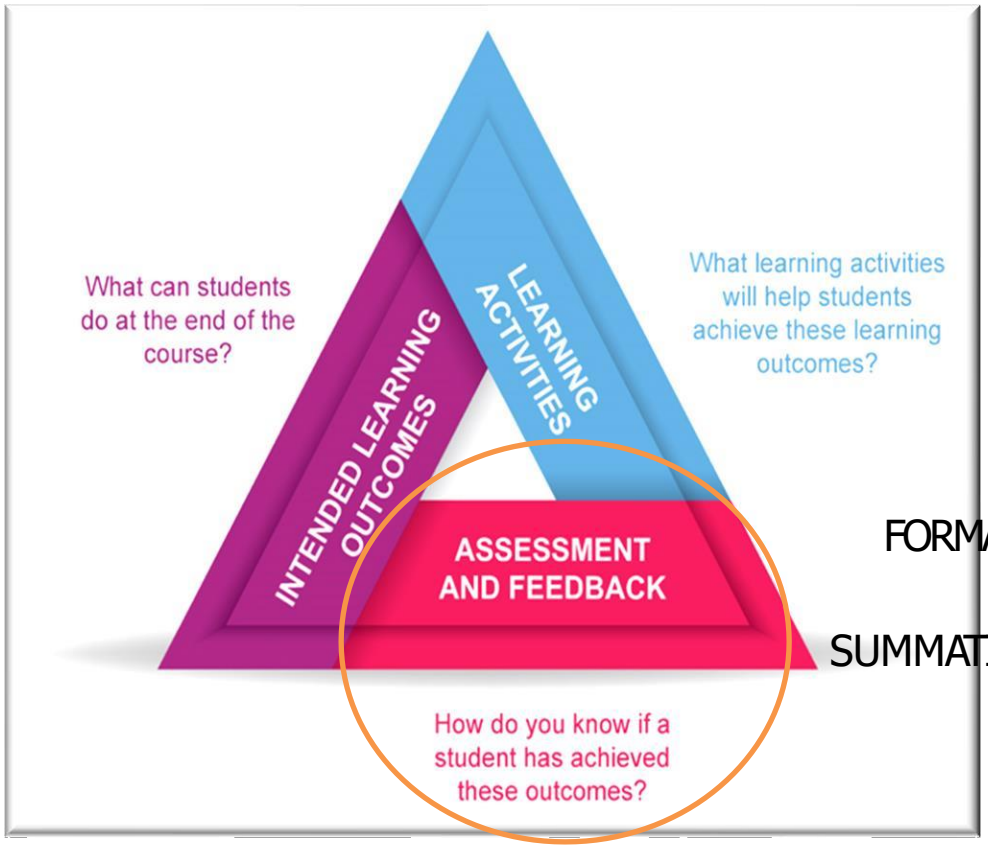
Constructive alignment ensures that the "learner cannot escape without learning what is intended"; they are trapped in a web of consistency between learning intentions, activities and assessment (Biggs, 2003).





We should design learning activities and assessments so they support students in achieving the learning outcomes (Biggs 1996).

Biggs, J. (1996). Enhancing teaching through constructive alignment. Higher Education, 32(3), 347–364.
Biggs, J. (2003) Teaching for Quality Learning at University (2nd ed.). Buckingham: SRHE and OUP.



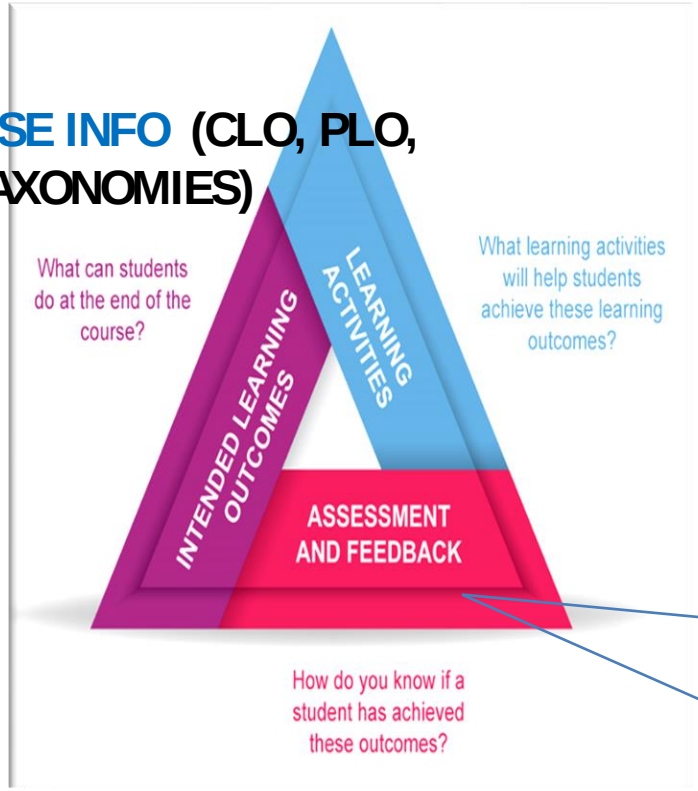
FORMATIVE
- CONTINUOUS , FEEDBACK, IMPROVEMENT

SUMMATIVE
- TESTING, EVALUATION

Biggs, J. (2003) l'eaching foí Quality Leaíning at Univeísity (2nd ed.). Buckingham: SRHE and OUP.

CONSTRUCTIVE ALIGNMENT

COURSE INFO (CLO, PLO, TAXONOMIES)



TRADITIONAL ASSESSMENT

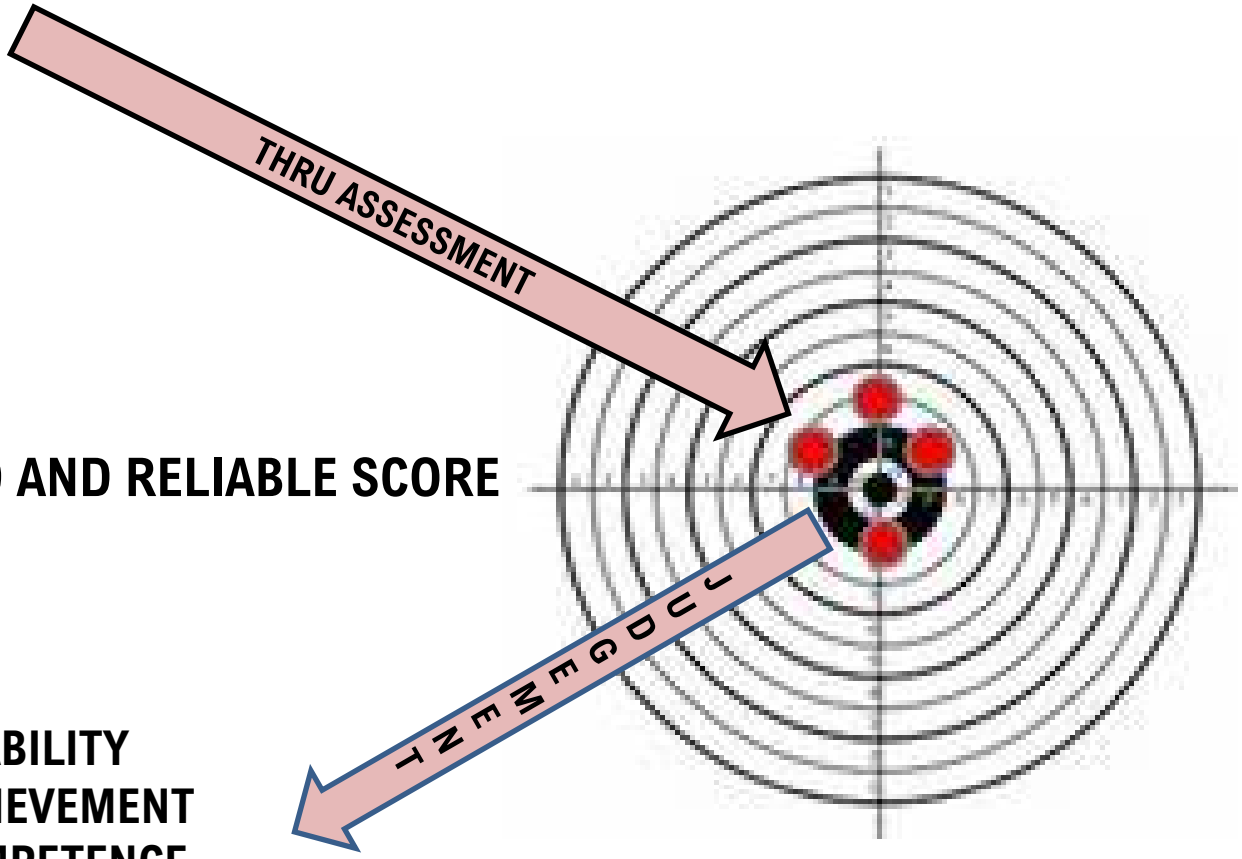
FINAL EXAM /TAKE HOME EXAM

ALTERNATIVE ASSESSMENT

TASK-BASED ASSIGNMENT

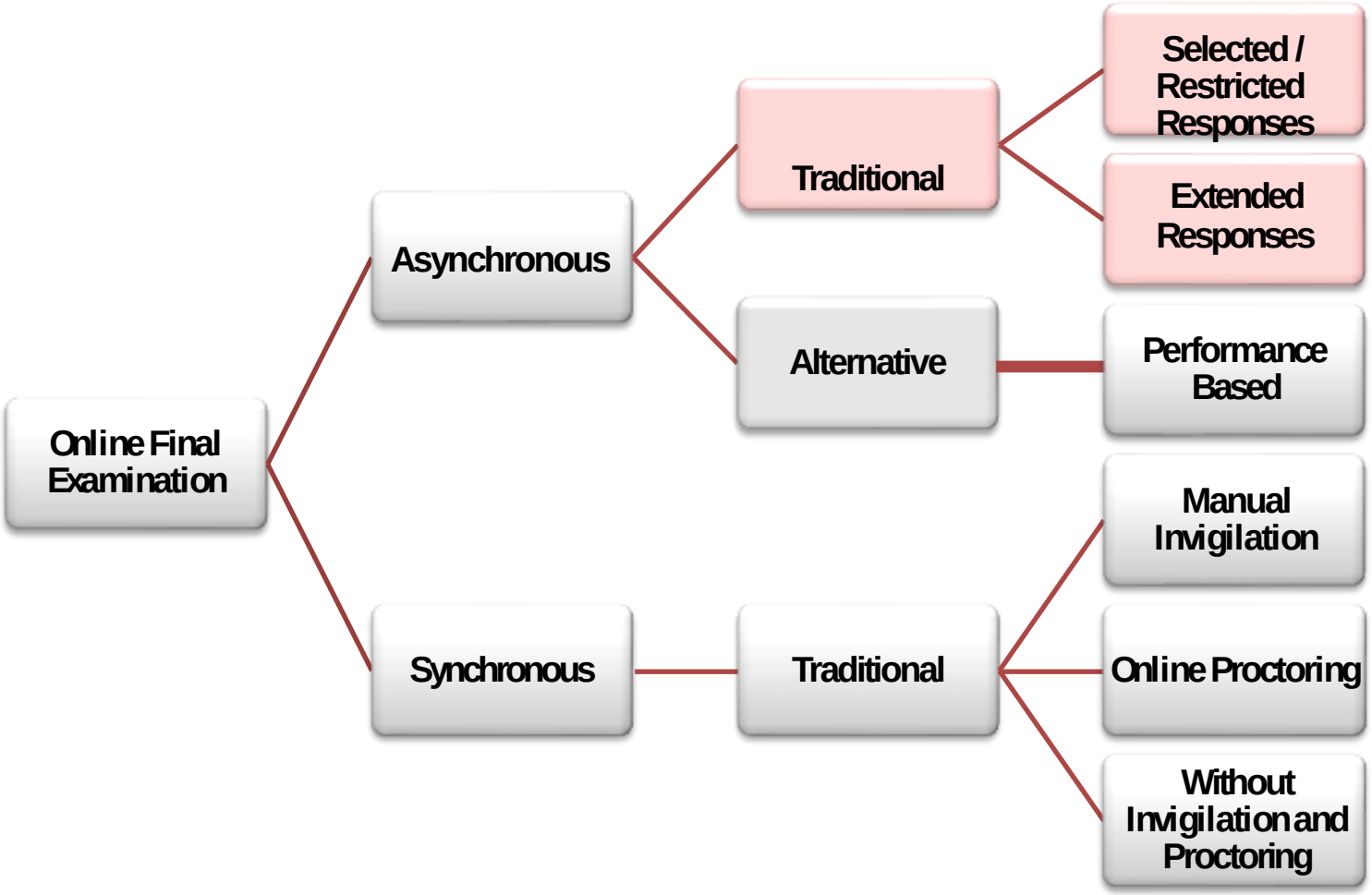
Biggs. J. (2003) Teaching for Quality Learning at University – What the Student Does?. 2nd Edition SRHE / Open University Press, Buckingham.

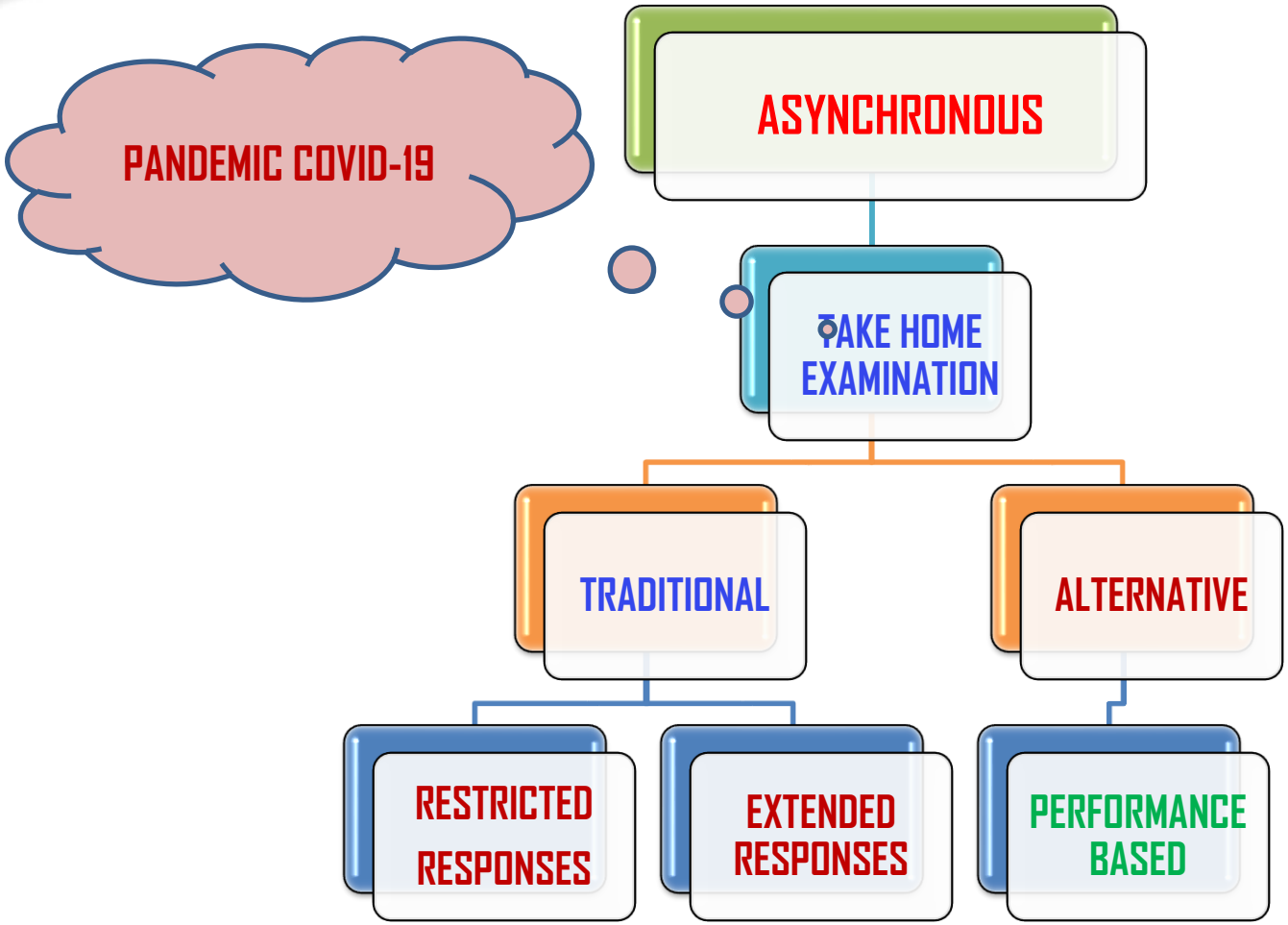
WHY CONSTRUCTIVE ALIGNMENT IS SO IMPORTANT?



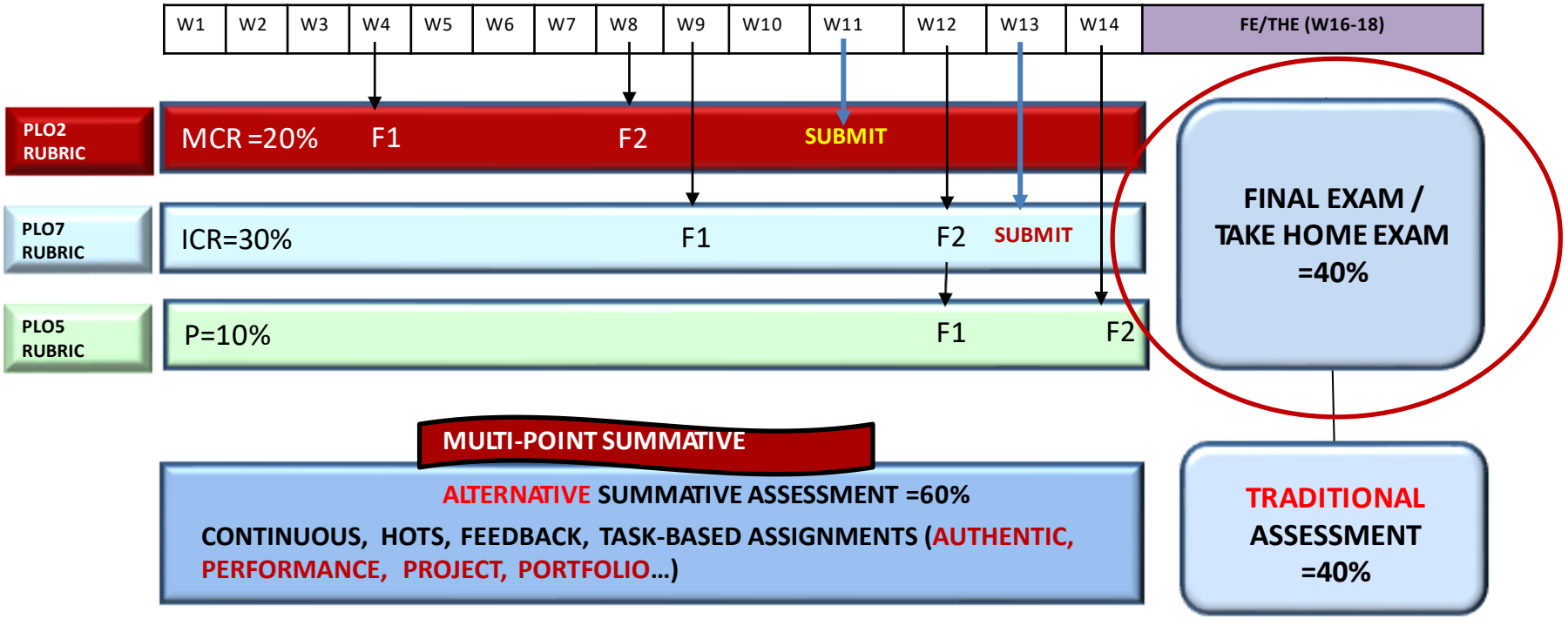
VALID AND RELIABLE SCORE

- ABILITY
- ACHIEVEMENT
- COMPETENCE
- MASTERY


Online Final Examination
Asynchronous
Synchronous
Traditional
Alternative
Traditional
Selected / Restricted Responses
Extended Responses
Performance Based
Manual Invigilation
Online Proctoring
Without Invigilation and Proctoring


PANDEMIC COVID-19
ASYNCHRONOUS
TAKE HOME EXAMINATION
TRADITIONAL
ALTERNATIVE
RESTRICTED RESPONSES
EXTENDED RESPONSES
PERFORMANCE BASED

COURSE PLAN



Sue Bloxham and Pete Boyd (2007). *Effective Assessment in Higher Education: A Practical Guide.*
Milton Keynes, Open University Press, ISBN 9780-335-221073

F1 = FEEDBACK 1, F2 = FEEDBACK 2
(Reserve the last 1 hour in giving Feedback to AA)
P = Presentation
MCR= Meta Content Report
ICR= Instrumentation Cycle Report

ASSESSMENT STRUCTURE

FROM CI

1	FINAL EXAM / TAKE HOME EXAM (MCO)	CLO1- **PLO1**	• SUMMATIVE	40%
2	• META CONTENT REPORT(MCR)	CLO2- **PLO2**	• ALTERNATIVE SUMMATIVE ASSESSMENT	20%
3	• INSTRUMENTATION CYCLE REPORT(ICR)	CLO3- **PLO7**	• ALTERNATIVE SUMMATIVE ASSESSMENT	30%
4	• PRESENTATION (P)	CLO4- **PLO5**	• ALTERNATIVE SUMMATIVE ASSESSMENT	10%

SCORING TOOLS

- **PLO1 (TECHNICAL – CONTENT) – ANSWER SCHEME**
- **PLO2(TECHNICAL - CONTENT) - RUBRIC**
- **PLO7 (NUMERICAL SKILL) - RUBRIC**
- **PLO5 (COMMUNICATION SKILL) - RUBRIC**

TEST VS EXAM

What does Test mean?

According to the American Heritage Dictionary test means **'a series of questions, problems, or physical responses designed to determine knowledge, intelligence, or ability.'**

- A test is a **short exam** that a educator gives to his or her students at the end of a lesson in order to understand how much of what he or she has taught has gone into the students' minds.
- A test is **not** very formal.

Tanner, D.E.(2001). Assessing Academic Achievement. Allyn and Bacon,Needham Heights, MA

TEST VS EXAM

What does Exam mean?

The word **exam** referring to a very **formal test / MASTERY test.**

In the educational level, it is a test that tests knowledge **on a number of lessons**.

Exam is held at the end of a semester or a term; can be **written exam or practical exam**.

Tanner, D.E.(2001). Assessing Academic Achievement. Allyn and Bacon,Needham Heights, MA

UNDERLYING CONCEPT

Provide information about an individual's achievement of a course objective or **M A S T E R Y of an area of the content**

Tanner, D.E.(2001). Assessing Academic Achievement. Allyn and Bacon,Needham Heights, MA

WHAT MAKES AN EXAM A GOOD EXAM?

- **1. Variance in scores**: The goal of discrimination is achieved only if there is sufficient variance in the scores of the test takers. A test which is **too tough** would result in all test takers scoring low marks while one that is **too easy** will lead to overall high scores thereby **not highlighting any discrimination** on any of the criterions and thus neither test is considered good.

- **2. Reliability**: Is a measure of a test's consistency – both over a period of time as well as internal consistency. It measures **precision of test scores** or extent of measurement error in the test (SEM low, Reliability high)

Linn, R.L & Miller, M.D. (2005). Measurement and Assessment in Teaching. Pearson Education Inc., Upper Saddle River, New Jersey

WHAT MAKES AN EXAM A GOOD EXAM?

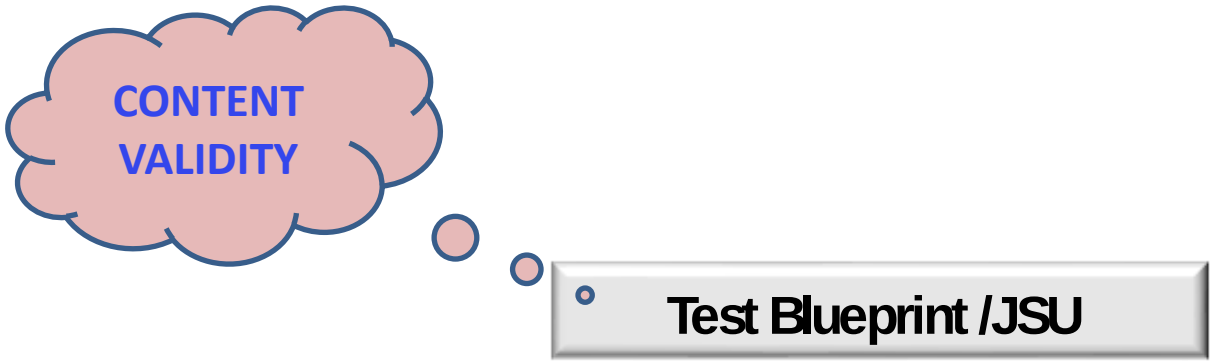
- **3. Validity**: Validity is an indicator of **how well an assessment is measuring what it is supposed to measure.** In other words it measures a test's usefulness.

- **4. Truth in Testing/ Integrity**: A good test has integrity and transparency built into it at multiple stages.

  While the test is being developed, it should be reviewed by a number of experts to make it free of developer bias,

  Once the test is developed it is reviewed on the basis of its content and scoring.

Linn, R.L & Miller, M.D. (2005). Measurement and Assessment in Teaching. Pearson Education Inc., Upper Saddle River, New Jersey


CONTENT
VALIDITY
Test Blueprint /JSU

What is Table of Specifications (TOS)/Test Blueprint?

TOS is **a two-way chart** which describes the **topics** to be covered by a test and **the number of items** which will be associated with each topic.

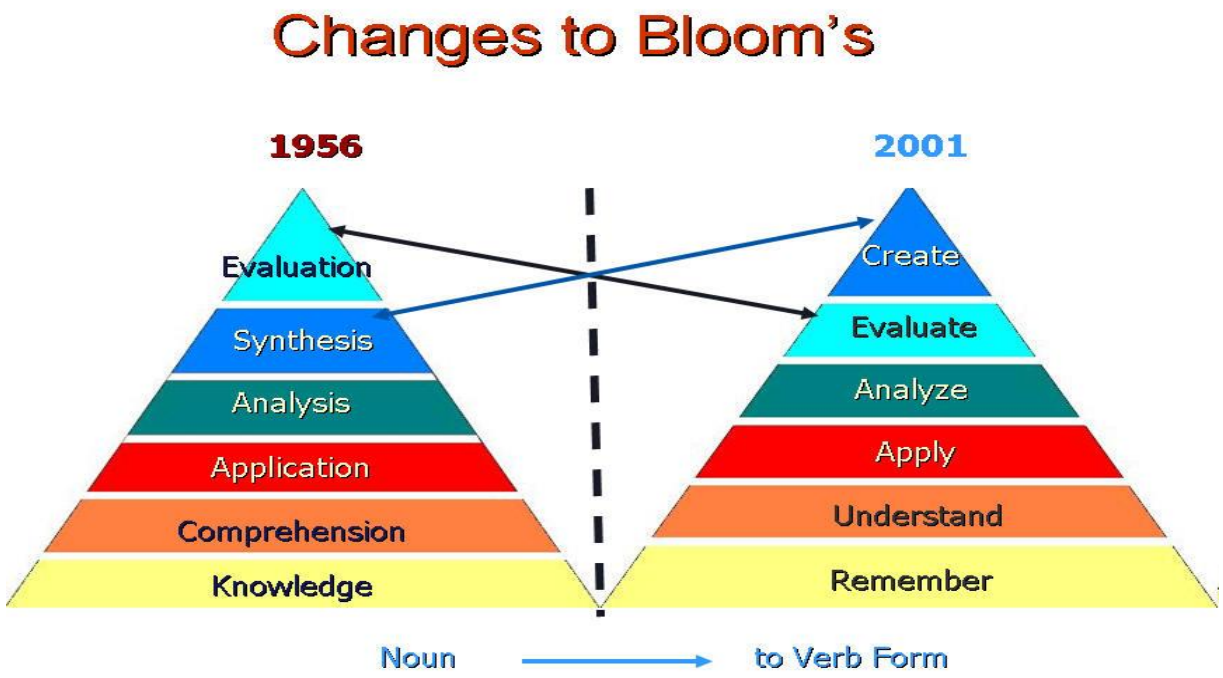
The purpose of TOS is to **identify the achievement domains being measured** and to ensure that **a fair and representative sample of questions appear on the test.**

TOS provides the teacher with evidence that **a test has content validity,** that it covers what should be covered.

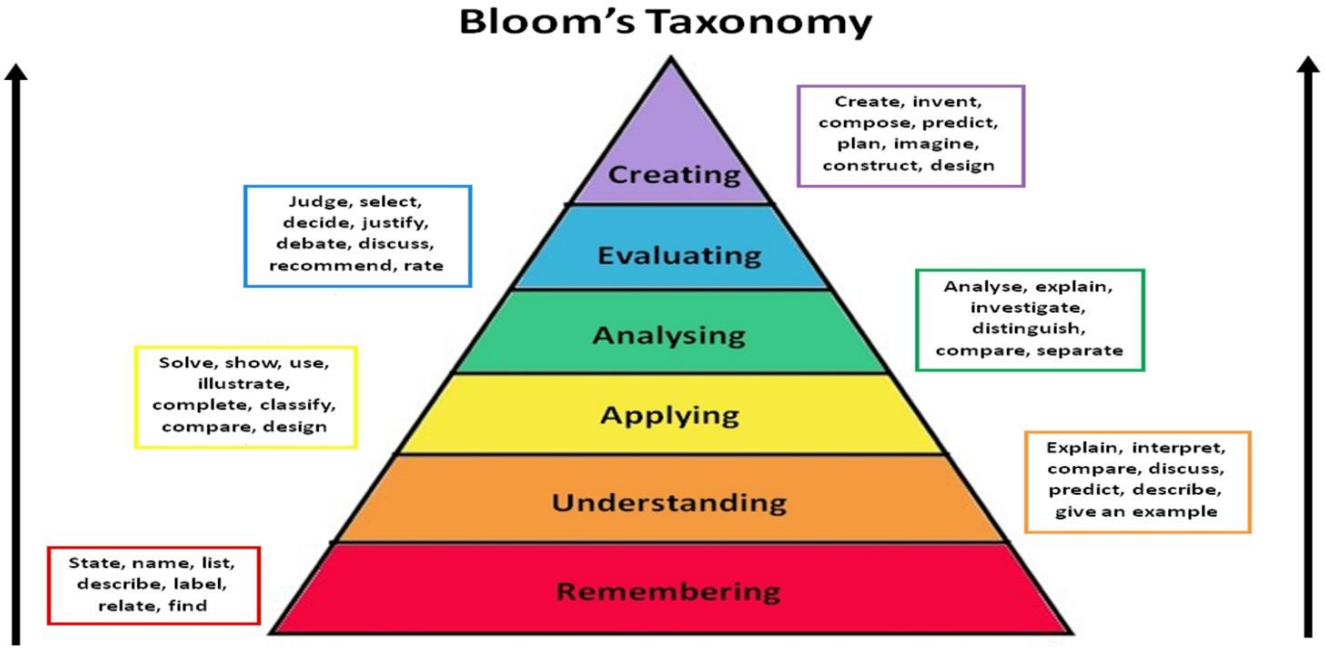
A Table of Specifications helps to ensure that there is **a match between what is taught and what is tested**.

The TOS ensures that there is balance between items that test **lower level thinking skills and those which test higher order thinking skills**

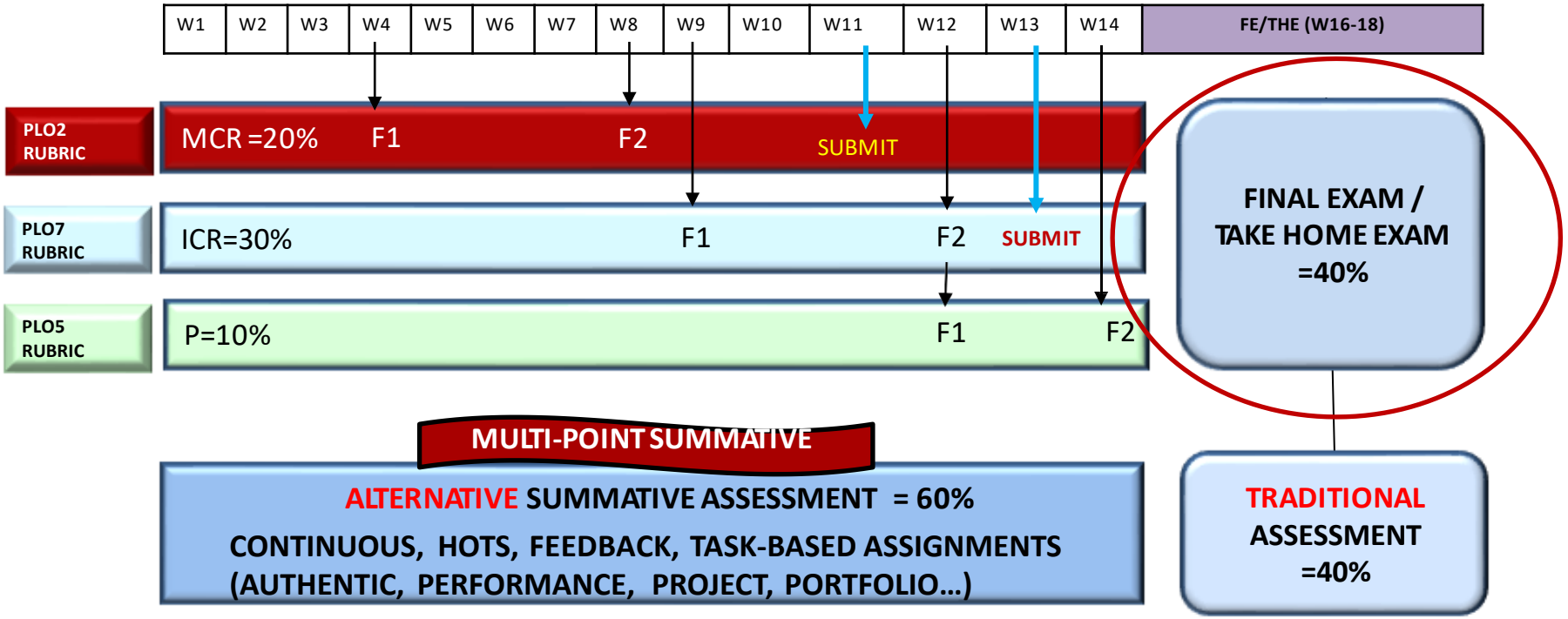
COMPARE



Bloom. B.S. (Ed.) (1956) Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain. Longmans, New York.



Anderson, L.W., & Krathwohl, D.R.(eds.). **(2001).** A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.



Sue Bloxham and Pete Boyd (2007). *Effective Assessment in Higher Education: A Practical Guide.*
Milton Keynes, Open University Press, ISBN 9780-335-221073

F1 = FEEDBACK 1, F2 = FEEDBACK 2
(Reserve the last 1 hour in giving Feedback to AA)
P = Presentation
MCR= Meta Content re[ort
ICR= Instrumentation Cycle Report

PREPARING TABLE OF ITEM SPECIFICATION

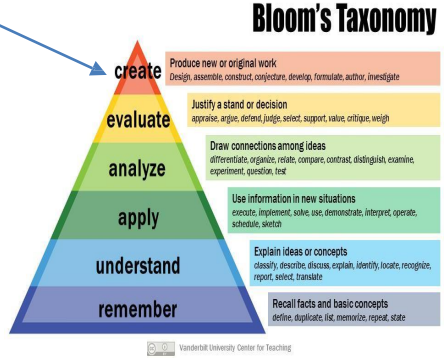
- Select **the learning outcomes** to be tested
- **Outline** the subject matter
- **Making the two-way chart**
- **Distributing the number of test item** (relative weights) according to cells of the table

COURSE INFORMATION

Mapping of the Course Learning Outcomes (CLO) to the Programme Learning Outcomes (PLO), Teaching & Learning (T&L) methods and Assessment methods:

No.	CLO*	PLO (Code)	**Taxonomies and ***generic skills	T&L methods	****Assessment methods
1	Discuss critically the psychological test in terms of its psychometric properties, procedures in designing for research purpose, interpretation of the scores and/or the challenges, trends and issues related to psychological testing.	PLO1 (KW)	C6	Lecture Intermittent Discussion (ID): Think-Pair-Share Round Robin Mind Mapping	Final Exam: 40%
CLO2	Design Meta Content Analysis according to the principles of measurement to extract the conceptual and operational definition of a construct being measured.	PLO2 (CG)	C6	Case Study ID: Read & Examine Brainstorming	Meta-Analysis Report: 20% *Report Rubric*

Levels of Cognitive



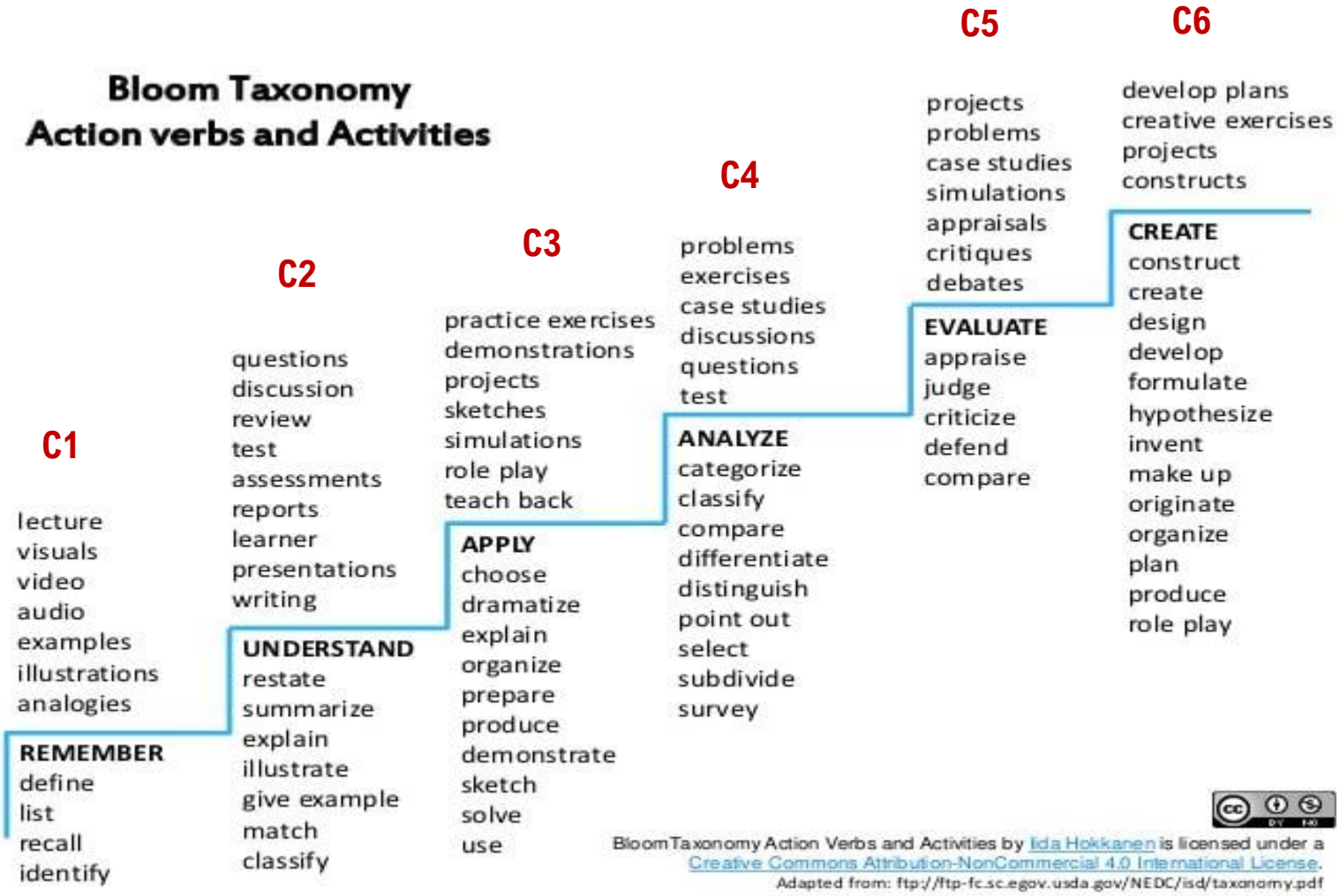
Armstrong, P. (2010). Bloom's Taxonomy. Vanderbilt University Center for Teaching. Retrieved 3 May 2021] from https://cft.vanderbilt.edu/guides-sub-pages/blooms- taxonomy/.

TAXONOMY BLOOM (1956, 2001) - COGNITIVE

Taxonomy / Hierarchy		Learning Outcomes	Action Verbs
C6	**CREATE/ SYNTHESIZE**	Ability to put parts together Ability to integrate different elements / concepts in order to form a sound pattern / structure so a new meaning can be established	Categorizes, Combines, Compiles, Composes, Creates, Devise, Designs, Plans, Organize, Reconstructs, Revises, Write
C5	**EVALUATE**	Ability to judge / value of material for given purpose Ability to come up with judgements about the importance of concepts	Appraises, Concludes, Criticizes, Critiques, Defends, Evaluates, Justifies, Recommends
C4	**ANALYZE**	Ability to break down information into its components Ability to differentiate	Analyses, Break down, Compares, Contrast, Diagram, Differentiates, Distinguishes, Illustrates, Outlines, Selects, Separates
C3	**APPLY**	Ability to utilize an abstraction or to use learned materials in new situation	Applies, Changes, Computes, Constructs, demonstrates, Discovers, Manipulates, Modifies, Operates, Prepares, Relates, Solves, Uses
C2	**UNDERSTAND**	Ability to understand and interpret learned information Ability to understand the meaning of what is known	Comprehends, Converts, Defends, Distinguishes, Estimates, Explains, Extends, generalizes, Gives an example, Interprets, Paraphrases, Rewrites, Summarizes, Translates
C1	**REMEMBER**	Ability to remember facts without necessarily understand Ability to recall data and/or information	Defines, Describes, Identifies, Knows, labels, Lists, Matches, Names, Outlines, Recalls, recognizes, reproduces, Selects, States

- Bloom, B.S., Engelhart, M.D.., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain. New York: David McKay.
- Anderson, L.W., & Krathwohl, D.R.(eds.). (2001). A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.

Taxonomy Bloom (2001) – Action Verb



Adapted from: ftp://ftp-fc.sc.egov.usda.gov/NEDC/isd/taxonomy.pdf

3

Prepare the Test Blueprint

TEST BLUEPRINT			
SCHOOL & FACULTY	School of Education Faculty of Social Sciences and Humanities		
NAME			
COURSE		CODE :	
SECTION		SESSION	

No.	TOPIC/SUBTOPIC	*Objective Items (O) Essay Items (E)*	*CLO*	*PLO*	COGNITIVE LEVELS						*Total*	*Percentage (%)*
					Remember	*Understand*	*Apply*	*Analyze*	*Evaluate*	*Synthesize / Create*		
	Total											
	Percentage (%)											
					LOTs			HOTs				

Prepared By:

_____

Date:

3H/17H*40=7M

7/40*100=18%

TOPICS	HOUR	MARK	PERCENTAGE
Measurement Data	3	7	18%
Scales	2	5	12%
Characters of good Instrument	3	7	18%
Flow of constructing instrument	3	7	18%
Pilot Test	2	5	12%
Establish Reliability using software – Winsteps (Rasch Measurement Model)	4	9	24%
TOTAL	17H	40M	100%

EXAMPLE

In CI, CLO1 AKW Taxonomy C5 Assessment Method FE

TABLE OF SPECIFICATION / TEST BLUEPRINT

FINAL EXAM [redacted]

Bil	Kandungan	Hours	PO	Cognitive Levels						SCORE	PERCENTAGE
				Remember	Understand	Apply	Analyze	Evaluate	Synthesize/ Create		
1	Measurement Data	3	PO1	1(i)=2	1(ii)=2			1(iii)=3		7m	18%
2	Scales	2	PO1			2(ii)=2	2(i)=3			5m	12%
3	Characters of good Instrument	3	PO1		3(i)=2 3(ii)=2			3(iii)=3		7m	18%
4	Flow of constructing instrument	3	PO1			4(i)=2 4(ii)=2		4(ii)=3		7m	18%
5	Pilot Test	2	PO1			5(i)=3	5(ii)=2			5m	12%
6	Establishing Reliability using software –Winsteps : Rasch Measurement Model	4	PO1		6(i)=3		6(ii)=3	6(iii)=3		9m	24%
	SCORE			2	9	9	8	12		40m	100
	TOTAL SCORE			20			20				
	PERCENTAGE			20%			20%				FE =40%

Lower Order Thinking Skills (LOTS) – 20%

Higher Order Thinking Skills (HOTS)-20%

EXAMPLE

In CI, CLO1 AKW Taxonomy C6 Assessment Method FE

TABLE OF SPECIFICATION / TEST BLUEPRINT

FINAL EXAM [redacted]

Bil	Kandungan	Hours	PO	Cognitive Levels						SCORE	PERCENTAGE
				Remember	Understand	Apply	Analyze	Evaluate	Synthesize/ Create		
1	Measurement Data	3	PO1	1(i)=2	1(ii)=2			1(iii)=3		7m	18%
2	Scales	2	PO1			2(ii)=2	2(i)=3			5m	12%
3	Characters of good Instrument	3	PO1		3(i)=2 3(ii)=2				3(iii)=3	7m	18%
4	Flow of constructing instrument	3	PO1			4(i)=2 4(ii)=2			4(iii)3	7m	18%
5	Pilot Test	2	PO1			5(i)=3			5(ii)=2	5m	12%
6	Establishing Reliability using software –Winsteps : Rasch Measurement Model	4	PO1		6(i)=3		6(ii)=3	6(iii)=3		9m	24%
	SCORE			2	9	9	6	6	8	40	100
	TOTAL SCORE			20			20				
	PERCENTAGE			20%			20%				FE =40%

Lower Order Thinking Skills (LOTS) – 20%

Higher Order Thinking Skills (HOTS)-20%

Hierarchy		Learning Outcomes	Action Verbs
C6	CREATING /SYNTHESIZE	Ability to put parts together Ability to integrate different elements / concepts in order to form a sound pattern / structure so a new meaning can be established	Categorizes, Combines, Compiles, Composes, Creates, Devise, Designs, Plans, Organize, Reconstructs, Revises, Write
C5	EVALUATE	Ability to judge / value of material for given purpose Ability to come up with judgements about the importance of concepts	Appraises, Concludes, Criticizes, Critiques, Defends, Evaluates, Justifies, Recommends
C4	ANALYZE	Ability to break down info Ability to differentiate	Analyses, Break down, Compares, Contrast, Diagram, Differentiates, Distinguishes, Illustrates, Separates
C3	APPLY	Ability to utilize an abstra new situation	Applies, Changes, Computes, Constructs, demonstrates, Discovers, Manipulates, Modifies, Operates, Prepares, Relates, Solves, Uses
C2	UNDERSTAND	Ability to understand and interpret learned information Ability to understand the meaning of what is known	Comprehends, Converts, Defends, Distinguishes, Estimates, Explains, Extends, generalizes, Gives an example, Interprets, Paraphrases, Rewrites, Summarizes, Translates
C1	REMEMBER	Ability to remember facts without necessarily understand Ability to recall data and/or information	Defines, Describes, Identifies, labels, Lists, Matches, Names, Outlines, Recalls, Recognizes, Reproduces, Selects, States

Question 1=7m

1(i) List -2m
1(ii) Explain - 2m
1(iii) Recommend- 3m

COURSE INFORMATION

2

CONTENT TO BE TESTED

Week 1	Introduction to Scaling and Instrumentation	
Week 2	Measurement Data (Nominal, Ordinal, Interval, Ratio)	✓
Week 3	Types of Scales	✓
Week 4	Instrumentation Plan	✓
Week 5	Instrument Development Process	✓
Week 6	Instrument Conceptualization [Questionnaire]	✓
Week 7	Item Construction	✓
Week 8	Pilot Test	✓
Week 9	Semester Break	
Week 10	Establish Validity	✓
Week 11	Establish Reliability	✓
Week 12	SPSS/Winsteps (Data Input)/	
Week 13	SPSS/Winsteps (Data Analysis)	✓
Week 14	SPSS/Winsteps (Data Interpretation)	
Week 15	Revision	
Week 16-18	Exam	

4

Balance, Fair, Representative

TEST BLUEPRINT			
SCHOOL & FACULTY	School of Education Faculty of Social Sciences and Humanities		
NAME			
COURSE		CODE :	
SECTION		SESSION	

No.	TOPIC/SUBTOPIC	Objective Items (O) Essay Items (E)	CLO	PLO	COGNITIVE LEVELS						Total
					Remember	Understand	Apply	Analyze	Evaluate	Synthesize / Create	
1	Measurement Data (Nominal, Ordinal, Interval, Ratio)	O	1	1	1,2						2
2	Types of Scales	O	1	1		3		4			2
3	Instrumentation Plan	O	1	1	5			6	7	8	4
4	Instrument Development Process	O	1	1		9	10	11			3
5	Instrument Conceptualization [Questionnaire]	O	1	1		12	13,14		15	16	5
6	Item Construction	O	1	1	17	18		19	20	21	5
7	Pilot Test	O	1	1		22		23	24		3
8	Validity	O	1	1	25		26			27	3
9	Reliability	O	1	1					28		1
10	Data Analysis and Interpretation	O	1	1			29			30	2
	Total				5	5	5	5	5	5	**30**
	Percentage (%)				50%			50%			**100%**
					LOTs			HOTs			

Prepared By:

/

Date:

It is recommended that the open-ended questions posed in the final examination should consist of 10 to 20% of the total marks(40%)

Hafizah Husain, Badariah Bais, Aini Hussain & Salina Abdul Samad (2012). How to Construct Open Ended Questions. https://www.sciencedirect.com/science/journal/18770428

Malaysian Education Blueprint 2015-2025_HOTs

Take Home Exam

Open Book Exam

Extended Time

It is recommended that the open-ended questions posed in the final examination should consist of 30%of the total marks (40%)

Stretch

Assessments should demonstrate stretch

Include both high and low-achieving students

Holly Lavender and Cathryn Shaw (2013). Understanding Test Blueprints and Their Impact on Student Growth. Ohio Department of Education

Markah	Gred	Mata Nilai
90 - 100	A+	4.00
80 - 89	A	4.00
75 – 79	A-	3.67
70 – 74	B+	3.33
65 – 69	B	3.00
60 – 64	B-	2.67
55 – 59	C+	2.33
50 – 54	C	2.00
45 – 49	C-	1.67
40 – 44	D+	1.33
35 – 39	D	1.00
30 – 34	D-	0.67
00 – 29	E	0.00

← Mastery Score

FINAL EXAM (INDIVIDUAL)	40%	HOTS 30%
		LOTS 10%
ALTERNATIVE ASSESSMENT (GROUP)		60%

Mastery Score

Markah	Gred	Mata Nilai
90 - 100	A+	4.00
80 - 89	A	4.00
75 – 79	A-	3.67
70 – 74	B+	3.33
65 – 69	B	3.00
60 – 64	B-	2.67
55 – 59	C+	2.33
50 – 54	C	2.00
45 – 49	C-	1.67
40 – 44	D+	1.33
35 – 39	D	1.00
30 – 34	D-	0.67
00 – 29	E	0.00

5

Start Constructing Items Based on Test Blueprint

MPPR1333
Introduction to
Scaling and
Instrumentation

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

UNIVERSITI TEKNOLOGI MALAYSIA
FAKULTI PENDIDIKAN

FINAL EXAM
(TAKE HOME EXAM)

SEMESTER II SESSION 2020/2021

INTRODUCTION TO SCALING AND INSTRUMENTATION

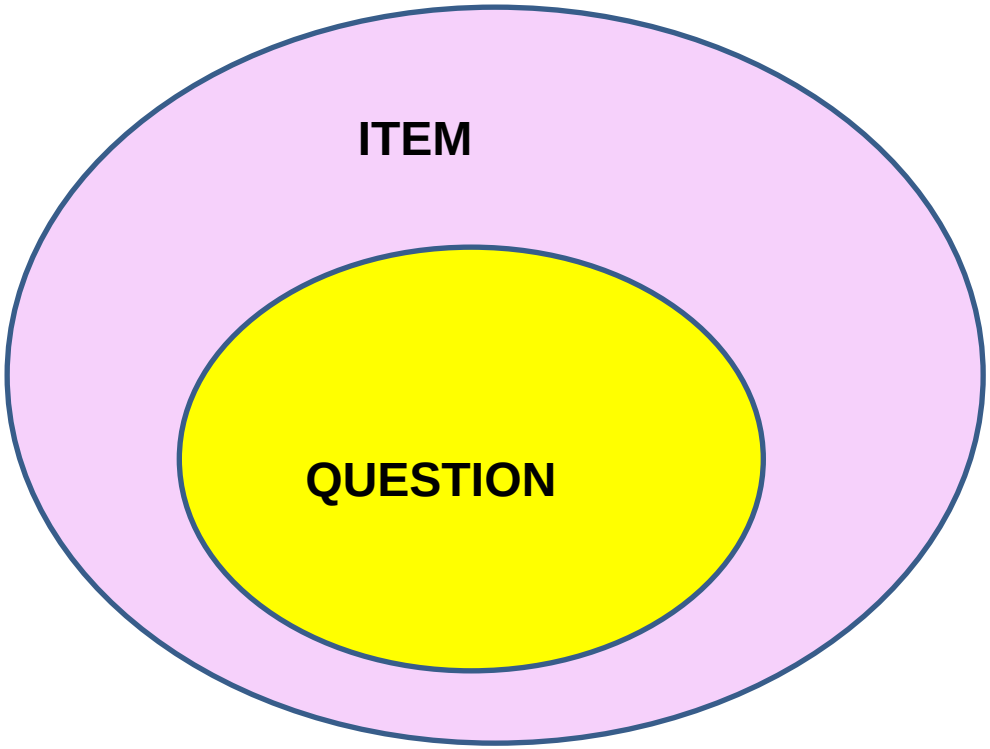
Five Hours

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO

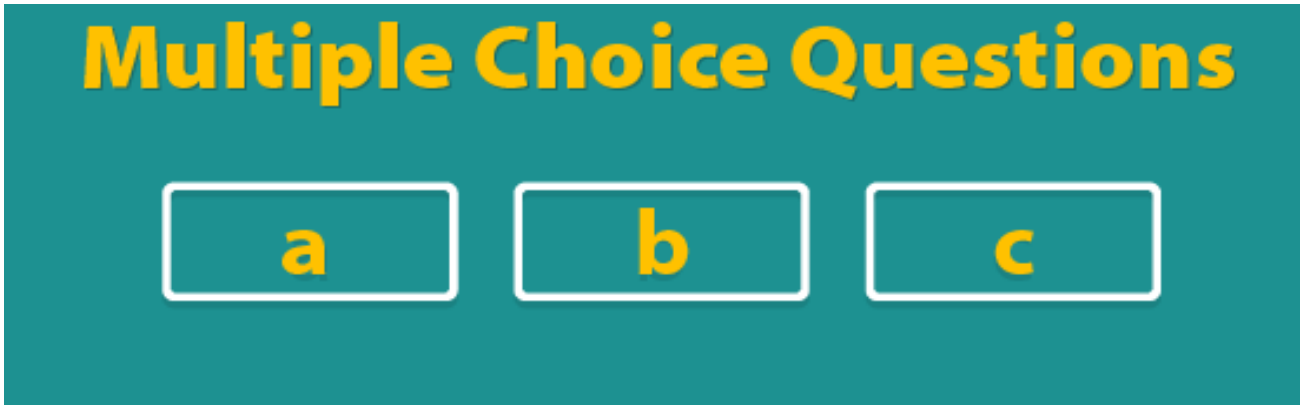
Instructions:

1. This paper consists of **thirty (30) multiple choice** questions.
2. Answer **all** questions in teh answer sheet given.

MCQS/ ITEM DEVELOPMENT

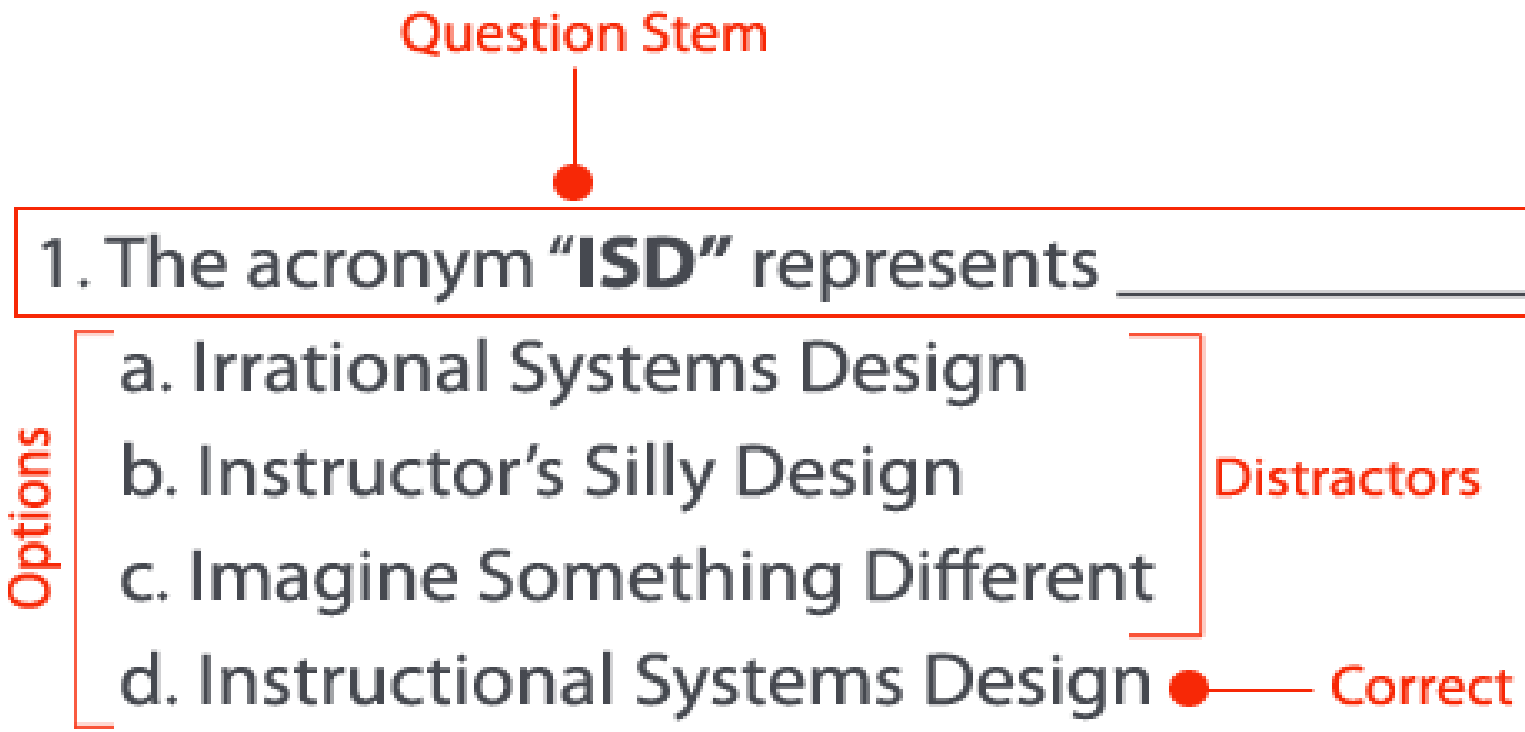


RULES IN WRITING



Theelearningcoach.com/elearning_design/rules-for-multiple-choice-questions/

ANATOMY OF A MULTIPLE CHOICE QUESTION



A standard multiple-choice test item consists of two basic parts:

(I) a problem (stem)
(II) a list of suggested solutions (alternatives).

The stem may be in the form of either **a question** or an **incomplete statement**, and the list of alternatives contains one correct or best alternative (**answer**)
and a number of incorrect or inferior alternatives (**distractors**).

The purpose of the distractors is to appear as plausible solutions to the problem for those students who have not achieved the objective being measured by the test item. Conversely, the distractors must appear as implausible solutions for those students who have achieved the objective.

#1: Test comprehension and critical thinking, not just recall

Multiple choice questions are criticized for testing the superficial recall of knowledge. You can go beyond this by asking learners to interpret facts, evaluate situations, explain cause and effect, make inferences, and predict results.

#2: Use simple sentence structure and precise wording

Write test questions in **a simple structure that is easy to understand**. And try to be as **accurate as possible in your word choices.** Words can have many meanings depending on colloquial usage and context.

#3: Place most of the words in the question stem

If you're using a question stem, rather than an entire question, **ensure that most of the words are in the stem.** This way, the answer options can be short, making them less confusing and more legible.

#4: Make all distractors plausible

All of the wrong answer choices should be completely reasonable. This can be very hard to accomplish, but avoid throwing in those give-away distractors as it detracts from the test's validity. If you're really stuck, get help from your friendly **SME.**

(*BTW, this word can also be spelled as "distracter."*)

#5: Keep all answer choices the same length

This can be difficult to achieve, but expert test-takers can use answer length as a hint to the correct answer. Often the longest answer is the correct one. When you can't get all four answers to the same length, use two short and two long.

#6: Avoid double negatives

Don't use combinations of these words in the same question**: not, no, nor, the -un prefix,** etc. For example, this type of question could confuse test-takers: '**Which of the following comments would NOT be unwelcome in a work situation?'**

Flip it around and write it in the positive form: '**Which of the following comments are acceptable in a work situation?'**

#7: Mix up the order of the correct answers

Make sure that most of your correct answers aren't in the "b" and "c" positions, which can often happen. **Keep correct answers in random positions and don't let them fall into a pattern that can be detected.** When your test is written, go through and reorder where the correct answers are placed, if necessary.

#8: Keep the number of options consistent

Making the number of options consistent from question to question helps learners know what to expect. Research doesn't seem to agree on whether 3 or 4 or 5 options is best. **Recommended to use 4 options. It feels fair.**

#9: Avoid tricking test-takers

Tests exist to measure knowledge. **Never** use questions or answer options that could **trick a learner**. If a question or its options can be interpreted in two ways or if the difference between options is too subtle, then **find a way to rewrite it.**

#10: Use 'All of the Above' and 'None of the Above' with caution

When you run out of **distractors**, ***All of the Above*** **and** ***None of the Above*** can come in handy. But they may not promote good instruction.

Here's why. *All of the Above* can be an obvious give-away answer when it's not used consistently. Also, the *All of the Above* option can **encourage guessing** if the learner thinks one or two answers are correct. In addition, the downside to *None of the Above* is that **you can't tell if the learner really knew the correct answer.**

HOW TO IMPROVE A TEST'S RELIABILITY

1. Add more questions to the test.
A **longer test is going to be more reliable**.

2. **Word test questions very clearly** so that no other interpretations are possible.

3. Write **test instructions** so that they are **easily understood.**

4. Make sure the **answer choices are clearly different from each other and that distractors** (wrong answers) are 100% wrong.

5. Create test items of **similar difficulty**, when possible.

NEGATIVE ITEMS

For most educational objectives, a student's achievement is more effectively measured by **identifying a correct answer rather than an incorrect answer**.

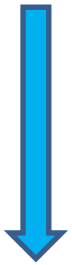
Just because the student knows an incorrect answer does not necessarily imply that he or she knows the correct answer.

Negative items must be carefully worded to avoid confusing the student. The negative word should be placed in the stem, NOT in the alternatives, and should be emphasized by using underlining, italics, bold face, or CAPITALS. In addition, each of the **alternatives should be phrased positively** to avoid forming a confusing double negative with the stem.

NEGATIVE ITEMS

All of the following are **correct** procedures for putting out a fire in a pan on the stove **except**:

a. Do not move the pan.
b. Pour water into the pan.
c. Slide a fitted lid onto the pan.
d. Turn off the burner controls



All of the following are **correct** procedures for putting out a fire in a pan on the stove **except**:

a. **Leave** the pan where it is.
*b. Pour water into the pan.
c. Slide a fitted lid onto the pan.
d. Turn off the burner controls.

Research. In a survey of 46 authoritative references in the field of educational measurement, 31 of the 35 authors that discussed the negative variety recommend that they be avoided (Haladyna & Downing, 1989a).

COMPLEX MCQ/ ITEMS

The fluid imbalance known as edema is commonly associated with:
1. Allergic reactions.
2. Congestive heart failure.
3. Extensive burns.
4. Protein deficiency.

The correct answer is:
A 1, 2, and 3.
B 1 and 3.
C 2 and 4.
D 4 only.
E 1, 2, 3, and 4.

RESEARCH. Numerous studies indicate that items of the **combined-response variety** are **lower in reliability, lower in discrimination, higher in difficulty, and equal in validity** when compared with similar items of the single-correct-answer and best-answer varieties (Albanese, 1990; Haladyna & Downing, 1989b). They have also been found to be **lower in reliability, higher in difficulty, and equal in validity** when compared with similar multiple true-false items (Frisbie, 1990).

1. BASE EACH ITEM ON A SPECIFIC PROBLEM STATED CLEARLY IN THE STEM.

Direct Question Example

Which of the following was the principal keyboard instrument in 16th century Europe?

a. Clavichord.
*b. Harpsichord.
c. Organ.
d. Pianoforte.

Incomplete Sentence Example

The principal keyboard instrument in 16th century Europe was the:

a. Clavichord.
*b. Harpsichord.
c. Organ.
d. Pianoforte.

2. INCLUDE AS MUCH OF THE ITEM AS POSSIBLE IN THE STEM, BUT DO NOT INCLUDE IRRELEVANT MATERIAL

Rather than repeating redundant words or phrases in each of the alternatives, place such material in the stem to decrease the reading burden and more clearly define the problem in the stem.

Poor Example

If the pressure of a certain amount of gas is held constant, what will happen if its volume is increased?

a. The temperature of the gas will decrease.
*b. The temperature of the gas will increase.
c. The temperature of the gas will remain the same.

Better Example

If you increase the volume of a certain amount of gas while holding its pressure constant, its temperature will:

a. Decrease.
*b. Increase.
c. Remain the same.

3. INCLUDE AS MUCH OF THE ITEM AS POSSIBLE IN THE STEM, BUT DO NOT INCLUDE IRRELEVANT MATERIAL

Poor Example

Suppose you are a mathematics professor who wants to determine whether or not your teaching of the unit on probability has had a significant effect on your students. You decide to analyze their scores from a test they took before the instruction and their scores from another exam taken after the instruction. Which of the following t-tests is appropriate to use in this situation?

*a. Dependent samples.
b. Heterogeneous samples.
c. Homogeneous samples.
d. Independent samples.

Better Example

When analyzing your students' pretest and posttest scores to determine if your teaching has had a significant effect, an appropriate statistic to use is the t-test for:

*a. Dependent samples.
b. Heterogeneous samples.
c. Homogeneous samples.
d. Independent samples.

The stem of the poor example above is excessively long for the problem it is presenting. The stem of the better example has been reworded to exclude most of the irrelevant material, and is less than half as long.

4. WORD THE ALTERNATIVES CLEARLY AND CONCISELY.

Poor Example

The term *hypothesis*, as used in research, as defined as:

a. A conception or proposition formed by speculation or deduction or by abstraction and generalization from facts, explaining or relating an observed set of facts, given probability by experimental evidence or by factual or conceptual analysis but not conclusively established or accepted.
b. A statement of an order or relation of phenomena that so far as is known is invariable under the given conditions, formulated on the basis of conclusive evidence or tests and universally accepted, that has been tested and proven to conform to facts.
*c. A proposition tentatively assumed in order to draw out its logical or empirical consequences and so test its accord with facts that are known or may be determined, of such a nature as to be either proved or disproved by comparison with observed facts.

Better Example

The term hypothesis, as used in research, is defined as:

a. An assertion explaining an observed set of facts that has not been conclusively established.
b. A universally accepted assertion explaining an observed set of facts.
*c. A tentative assertion that is either proved or disproved by comparison with an observed set of facts.

5. KEEP THE ALTERNATIVES MUTUALLY EXCLUSIVE

Poor Example

How long does an *annual* plant generally live?
*a. It dies after the first year.
b. It lives for many years.
c. It lives for more than one year.
*d. It needs to be replanted each year.

Better Example

How long does an *annual* plant generally live?
*a. Only one year.
b. Only two years.
c. Several years.

In the poor example above, alternatives *a* and *d* overlap, as do alternatives *b* and *c*. In the better example, the alternatives have been rewritten to be mutually exclusive.

6. KEEP THE ALTERNATIVES HOMOGENEOUS IN CONTENT

Poor Example

Idaho is widely known as:
- *a. The largest producer of potatoes in the United States.
- b. The location of the tallest mountain in the United States.
- c. The state with a beaver on its flag.
- d. The "Treasure State."

Better Example

Idaho is widely known for its:
- a. Apples.
- b. Corn.
- *c. Potatoes.
- d. Wheat.

The poor example contains alternatives testing knowledge of state agriculture, physical features, flags, and nicknames. If the student misses the item, it does not tell the teacher in which of the four areas the student is weak. In the better example, all of the alternatives refer to state agriculture, so if the student misses the item, it tells the teacher that the student has a weakness in that area.

7. KEEP THE GRAMMAR OF EACH ALTERNATIVE CONSISTENT WITH THE STEM.

Poor Example

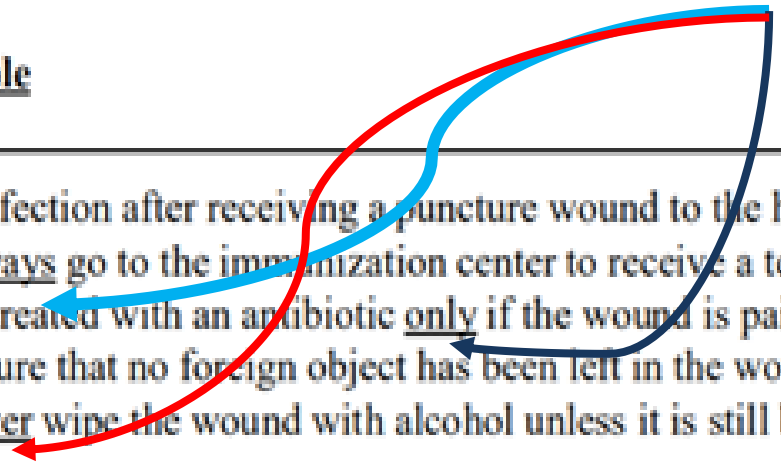
A word used to describe a noun is called an:
*a. Adjective.
b. Conjunction.
c. Pronoun.
d. Verb.

Better Example

A word used to describe a noun is called:
*a. An adjective.
b. A conjunction.
c. A pronoun.
d. A verb.

The word "an" in the stem of the poor example above serves as a clue to the correct answer, "adjective," because the other alternatives begin with consonants. The problem has been corrected in the better example by placing the appropriate article, "an" or "a," in each alternative.

8. AVOID THE USE OF SPECIFIC DETERMINERS



Poor Example

To avoid infection after receiving a puncture wound to the hand, you should:

a. Always go to the immunization center to receive a tetanus shot.
b. Be treated with an antibiotic only if the wound is painful.
*c. Ensure that no foreign object has been left in the wound.
d. Never wipe the wound with alcohol unless it is still bleeding.

Better Example

To avoid infection after receiving a puncture wound to the hand, you should always:

a. Go to the immunization center to receive a tetanus shot.
b. Be treated with an antibiotic if the wound is painful.
*c. Ensure that no foreign object has been left in the wound.
d. Wipe the wound with alcohol unless it is still bleeding.

9. AVOID INCLUDING KEYWORDS IN THE ALTERNATIVES.

Poor Example

When conducting library research in education, which of the following is the best source to use for identifying pertinent journal articles?

a. *A Guide to Sources of Educational Information.*
*b. *Current Index to Journals in Education.*
c. *Resources in Education.*
d. *The International Encyclopedia of Education.*

Better Example

When conducting library research in education, which of the following is the best source to use for identifying pertinent journal articles?

a. *A Guide to Sources of Educational Information.*
*b. *Education Index.*
c. *Resources in Education.*
d. *The International Encyclopedia of Education.*

10. AVOID THE ALTERNATIVES "ALL OF THE ABOVE" AND "NONE OF THE ABOVE" IN GENERAL)

Alternative	Use	Weakness
"All of the above"	Answer	Can be identified by noting that two of the other alternatives are correct
	Distractor	Can be eliminated by noting that one of the other alternatives is incorrect
"None of the above"	Answer	Measures the ability to recognize incorrect answers rather than correct answers
	Distractor	Does not appear plausible to some students

Research. While research on the use of "all of the above" is not conclusive, the use of "none of the above" has been found in several studies to decrease item discrimination and test score reliability (Haladyna & Downing, 1989b).

11. USE AS MANY FUNCTIONAL DISTRACTORS AS ARE FEASIBLE.

Functional distractors are those chosen by students that *have not* achieved the objective and are ignored by students that *have* achieved the objective. In other words, they have positive discrimination. The following table categorizes distractors according to functionality:

Description	Discrimination	Meaning
Functional	Positive	More non-achievers choose them than achievers
Nonfunctional	Low or none	Achievers and non-achievers choose them equally, or they are rarely chosen at all
Dysfunctional	Negative	More achievers choose them than non-achievers

Whether or not a distractor is functional can be determined through item analysis, a statistical procedure which is discussed in books such as the one by Oosterhof (1990) listed in the bibliography of this booklet.

USE AS MANY FUNCTIONAL DISTRACTORS AS ARE FEASIBLE.

Poor Example

Obsidian is an example of which of the following types of rocks?

*a. Igneous.
b. Metamorphic.
c. Sedimentary.
d. Transparent.
e. None of the above.

Better Example

Obsidian is an example of which of the following types of rocks?

*a. Igneous.
b. Metamorphic.
c. Sedimentary.

Assuming that alternatives *d* and *e* in the poor example above are rarely selected by students, the item is improved by removing these nonfunctional distractors.

ESSAY Qs

Educators choose essay questions over other **forms of assessment because essay items** challenge students to **create a response** rather than to **simply select a response.**

Some educators use them because essays have the potential **to reveal students' abilities to reason, create, analyze, synthesize, and evaluate.** In short, essay items are used for the advantages they offer.

There are **two major purposes for using essay questions**.

1. To assess students' understanding of and ability to think with subject matter content.
2. To assess students' writing abilities.

ESSAY Qs

John M. Stalnaker (1951, p.495)

"A test item which **requires a response composed by the examinee, usually in the form of one or more sentences**, of a nature that **no single response or pattern of responses can be listed as correct**, and the **accuracy and quality of which can be judged subjectively only by one skilled or informed in the subject**."

Based on Stalnaker's definition, an essay question should meet the following criteria:

1. Requires examinees to **compose** rather than **select** their response.
2. **Elicits student responses** that must consist of more than one sentence.
3. **Allows different or original responses or pattern of responses.**
4. Requires **subjective judgment** by a competent specialist to judge the **accuracy and quality of responses.**

ADVANTAGES, LIMITATIONS, AND COMMON MISCONCEPTIONS OF ESSAY QUESTIONS

ADVANTAGES

1. ASSESS HIGHER-ORDER OR CRITICAL THINKING SKILLS.

2. EVALUATE STUDENT THINKING AND REASONING.

3. PROVIDE AUTHENTIC EXPERIENCE

*(Constructed Responses Are Closer To Real Life Than Selected Responses,
*Problem Solving And Decision-making Are Vital Life Competencies)

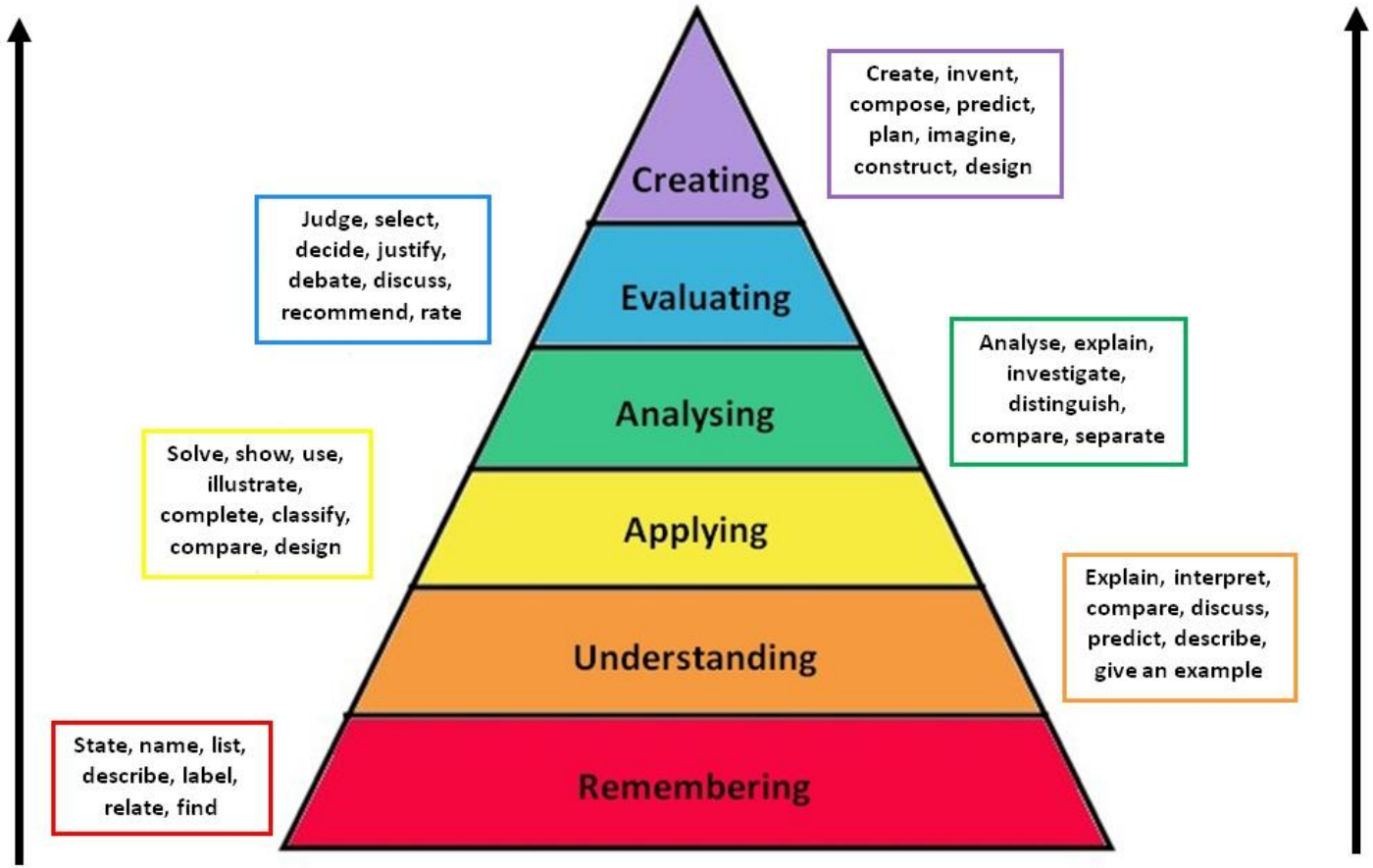
ADVANTAGES, LIMITATIONS, AND COMMON MISCONCEPTIONS OF ESSAY QUESTIONS

LIMITATIONS

1. **ASSESS A LIMITED SAMPLE OF THE RANGE OF CONTENT (the number of essay questions that can be included in a test is limited)**

2. **MARKING DIFFICULT AND TIME CONSUMING**
   (Different readers may vary in their grading of the same or similar responses (**interscorer reliability)** and one reader can vary significantly in his grading consistency depending on many factors (**intrascorer reliability**)

Bloom's Taxonomy



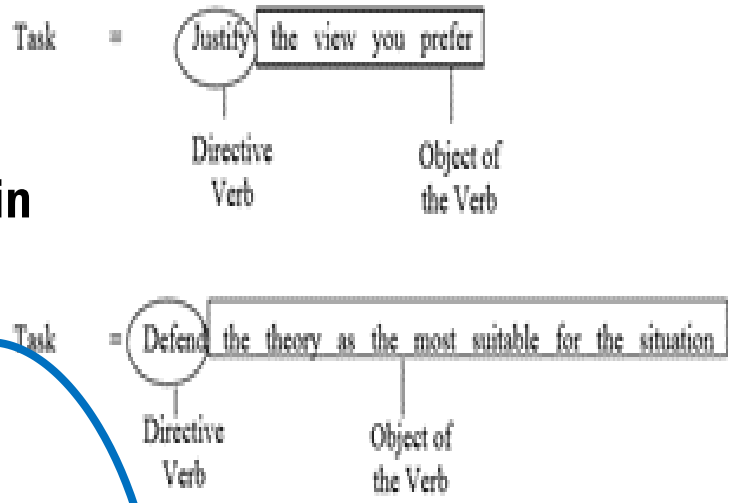
HOW SHOULD ESSAY QUESTIONS BE CONSTRUCTED?

1. **Clearly define the intended learning outcome to be assessed by the item. CLEAR AND SPECIFIC**

2. **Clearly define and SITUATE THE TASK within a problem situation.**

*** Avoid using essay questions for intended learning outcomes that are better assessed with other kinds of assessment.**

******Some types of learning outcomes can be more **efficiently** and **more reliably** assessed with selected-response questions than with essay questions.



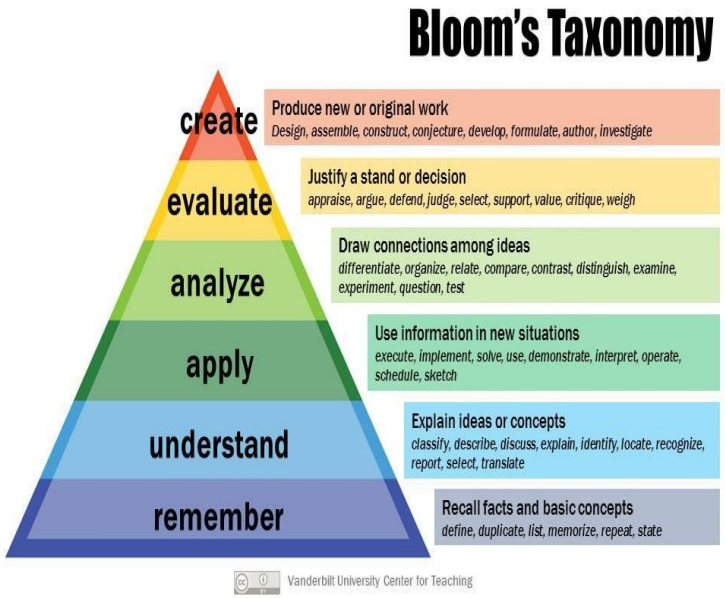
TAXONOMY BLOOM
ACTION VERB/ DIRECTIVE VERB

REFERENCES

Haladyna, T. M., & Downing, S. M. (1989a). A taxonomy of multiple-choice item-writing rules. Applied Measurement in Education, 2(1), 37-50.

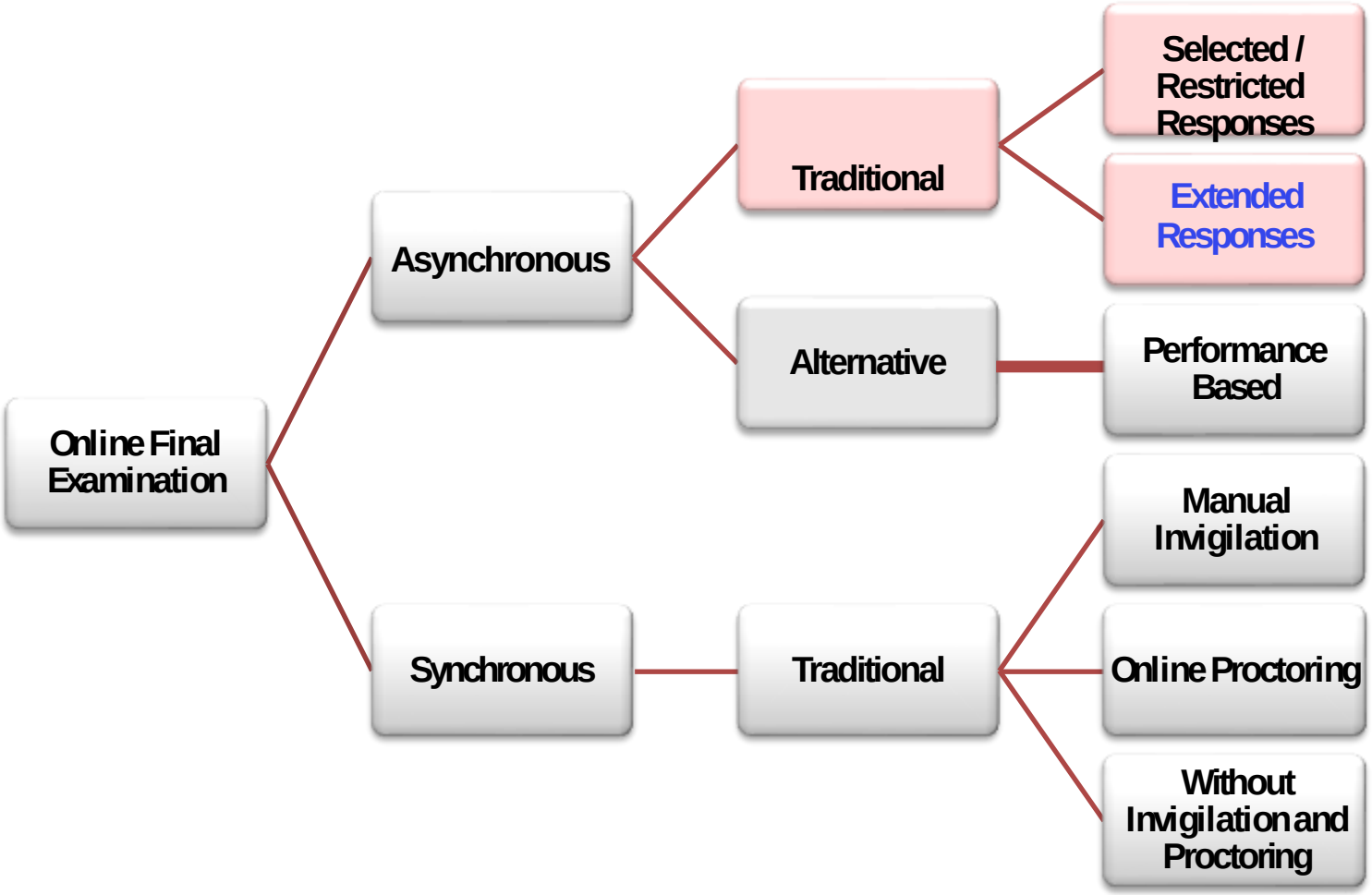
Haladyna, T. M., & Downing, S. M. (1989b). Validity of a taxonomy of multiple-choice item writing rules. Applied Measurement in Education, 2(1), 51-78.

Traditional Final Exam



- Can all PLOs be measured in the final exam? (Answer : **PLO1 and PLO2**)

- Trad-FE only appropriate to measure cognitive domain (6 levels of Taxonomy Bloom)


Online Final Examination
Asynchronous
Synchronous
Traditional
Alternative
Traditional
Selected / Restricted Responses
Extended Responses
Performance Based
Manual Invigilation
Online Proctoring
Without Invigilation and Proctoring

Open-ended questions require complex thinking and produce a variety of solutions (Badger 1992)

Badger, E., B. Thomas, (1992). Open-ended questions in reading. Practical Assessment, Res. and Evaluation, 3(4).
Finkel, D.L., (2000). Teaching with Your Mouth Shut. Boynton/Cook Publishers.

Finkel (2000) states specifically that open-ended question should be like the parable of the puzzle that intrigue the mind, but also challenging enough so that students would be interested in the answer.

Finkel, Donald L. (2000). Teaching with Your Mouth Shut. Portsmouth, NH: Heinemann, Boynton/Cook Publishers.

The Features of Open-ended Questions

Open-ended questions should be able to challenge the students

- to demonstrate a deep understanding of the materials being studied,
- the ability to use techniques from other disciplines, and
- the ability to assess the value of design, product, or system

Felder, R. M. (1987). On Creating Creative Engineers. Engineering Education, 77(4), 222-227.

The Features of Open-ended Questions

Cooney et al (2004) stipulates that the open-ended question should include the following features:

1. Authentic (real-world) Problem Solving
2. Require Multiple answer
3. Require Reasoning
4. Clearly Stated
5. Scoring Rubric (multiple point)

1. IT INVOLVES A SIGNIFICANT CONCEPT IN A RELATED FIELD.

The assessment items conveyed to the students what needs to be emphasized and what is important.

It involves giving students the chance to display their understanding by linking the entire topic and how it can lead to real world problem solving (authentic).

Cooney, T.J., W.B. Sanchez, K. Leatham, D.S. Mewborn (2004) Open-Ended Assessment in Math: Online version: http://books.heinemann.com/math).

The Features of Open-ended Questions

2. THERE COULD BE MULTIPLE ANSWERS TO OPEN-ENDED QUESTIONS.

1. Authentic (real-world) Problem Solving
2. Require Multiple answer
3. Require Reasoning
4. Clearly Stated
5. Scoring Rubric (multiple point)

Questions that require students to explain their thinking will encourage a variety of responses or reactions because not all students think the same.

For example, can an equilateral triangle have right angle? If yes, explain your reasons.

Typically, students focused on the angle and concluded that it is not possible, because all the sides of an equilateral triangle must have the same length and an equilateral triangle cannot have 270 degrees.

But one student replied that **NO,** because if it has a right angle it will have a hypotenuse. And it will become the longest side. But since all sides are of the same length so it cannot happen

3. NEED TO COMMUNICATE THE REASONING PROCESS.

1. Authentic (real-world) Problem Solving
2. Require Multiple answer
3. Require Reasoning
4. Clearly Stated
5. Scoring Rubric (multiple point)

One strong point of using open-ended questions is that students are given the opportunity to communicate what is in their minds

Example

Mary claimed that the area of any 30-60-90 triangle can be calculated even if only one side of the given length. Is Mary right or wrong? Explain your answer.

Here is an answer given by students: Mary is right. If you know the length of one side, you can divide or multiply with 3 or 2. After that, the height is multiplied with the base, divide by 2, and you will get the answer.

Another student gave the following response: Mary is wrong. Since all the angles are different, then all the sides are of different length. By knowing only one side initially, you would need to know the length of the other two sides (height and base) to calculate the area of the triangle.

The first student sees the relevance relations among the sides of the triangle 30-60-90, while the second student who may not realize this, does not see the relevance context of this problem.

When students are required to communicate their reasoning process, it was easy to understand what they know and they can use this to solve the given problem

4. OPEN-ENDED QUESTIONS SHOULD BE CLEARLY STATED.

1. Authentic (real-world) Problem Solving
2. Require Multiple answer
3. Require Reasoning
4. Clearly Stated
5. Scoring Rubric (multiple point)

Simulation activities in class setting

The fact is that open-ended questions should not be incomprehensible.
These types of questions should have a clear purpose even if there are many different answers.
In addition, students need to know what is expected of them and what lecturers consider as a good and complete response.

Many lecturers find that sharing a variety of answers with their students and asked them to evaluate these responses helped the students to determine what constitutes a good response.

This is because students are often not used to explain their thoughts in writing and it is important to help them to improve their communication, analytical and critical thinking skills.

The Features of Open-ended Questions

5. HAVE A SCORING RUBRIC.

Each item evaluated must have at least a 3-4-points rubric

But the purpose of the open-ended questions is to provide students with the opportunity to communicate their understanding in something other than the scenario of exactly right or wrong answer.

To give students partial credit is a common perception, and using the rubric to formulate the process helps to ensure fairness.

One of the features for a good open-ended question is to get the answer that agrees with the partial credit as defined by the rubric. As an example in the question of whether an equilateral triangle have a right angle?

Answer yes or no does not involve a partial credit, but **the answer to why** will allow a partial credit to be given.

1. Authentic (real-world) Problem Solving
2. Require Multiple answer
3. Require Reasoning
4. Clearly Stated
5. Scoring Rubric (multiple point)

Table 1. Open-ended question as modified from the closed-ended question

Example of closed-ended questions	Modified to open-ended questions
Which of the following numbers are prime? 7, 57, 67, 117	*Fred thinks that 57 and 67 are prime because they both end in 7, which is a prime number. Dick says he is wrong. Who is correct and why?*
What are the next three numbers in the following sequence? 1, 4, 7, 10, 13, ___, ____, ____	*Consider the following sequence: 1, 4, 7, 10, 13, ... Is 100 a member of this sequence? Explain your reasoning.*
Round 37.67 to the nearest 10th.	*Generate three different numbers that when rounded to the nearest 10th give 37.7.*
Find the LCM of 18 and 24	*Why can't 48 be the LCM of 18 and 24?*

Boaler, J. (1998). Open and Closed Mathematics: Student Experiences and Understandings. J. for Res. in Math. Ed., 29(1), 41 -62

a) *Determine the optimum solutions at the end user level for improving overall voltage sag performance and reliability.*

b) *Design a simple test distribution system for the customer facility as shown in Figure 2. PCC-1 is the utility bus at 13.8 kV and PCC-2 is the customer bus at 480 V. The data of PCC-2 bus are as follows:*

$Tr_1 = 1000\ kVA$, *5.0% impedance on 1000 kVA base*

$I_{load} = 1000\ A$

The measured distortion is equal to 90 A at 5^{th} harmonic and 44 A at 7^{th} harmonic. The short circuit impedance on 13.8 kV bus is equal to 1.7% on 1000 kVA base.

(b) Question 2

Noncompliance

1. Authentic (real-world) Problem Solving
2. ~~Require Multiple Answer~~
3. ~~Require Reasoning~~
4. Clearly Stated
5. Scoring Rubric (multiple point)

Table 2. Qualitative analysis of open-ended questions that met the feature

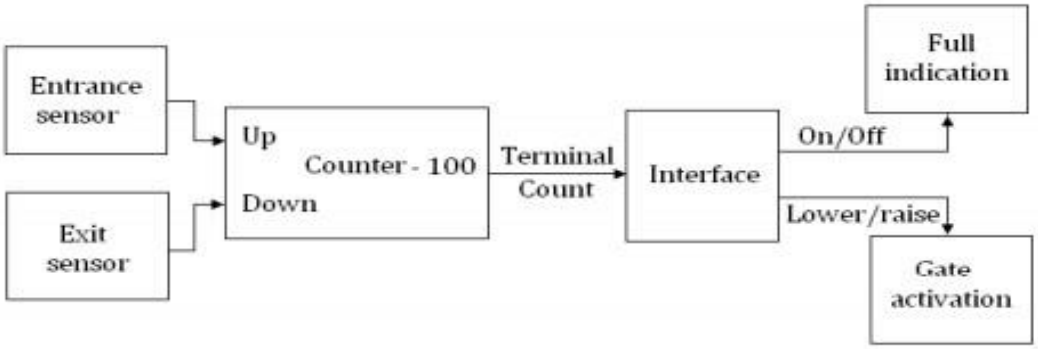
Question	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5
1	√	x	x	√	y
2	√	Part a) √	Part a) √	√	y
		Part b) x	Part b) x	√	y
3	√	Part a) x	Part a) x	√	y
		Part b) √	Part b) √	√	y
		Part c) √	Part c) √	√	y
		Part d) x	Part d) x	√	y
4	√	Part a) √	Part a) √	√	√
		Part b) √	Part b) √	√	√

Hafizah Husain, Badariah Bais, Aini Hussain & Salina Abdul Samad (2012). How to Construct Open Ended Questions. How to Construct Open Ended Questions. https://www.sciencedirect.com/science/journal/18770428

spaces in a one-hundred space parking garage and provide an indication for a full condition by illuminating a 'FULL' display sign and lowering a gate bar at the entrance. Optoelectronic sensors are used at the entrance and exit of the garage, an up/down counter and associated circuitry is used to count the number of parked cars and parking space available and an interface circuit that uses the counter to turn the 'FULL' sign on or off as required and lower or raise the gate bar at the entrance. The general block diagram is as shown in Figure 7. Unfortunately, one day, patrons complained that when they enter the garage (the gate was raised and the 'FULL' sign was off), and yet they can find no empty space.

a) As the technician in charge of this facility, Discuss with examples and illustrations, if any, what and where do you think the problem is.

b) Explain on the troubleshooting procedure and how will you troubleshoot and repair this system as quickly as possible



1. Authentic (real-world) Problem Solving
2. Require Multiple Answer
3. Require Reasoning
4. Clearly Stated
5. Scoring Rubric (multiple point)

(d) Question 4

Table 2. Qualitative analysis of open-ended questions that met the feature

Question	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5
1	√	x	x	√	y
2	√	Part a) √	Part a) √	√	y
		Part b) x	Part b) x	√	y
3	√	Part a) x	Part a) x	√	y
		Part b) √	Part b) √	√	y
		Part c) √	Part c) √	√	y
		Part d) x	Part d) x	√	y
4	√	Part a) √	Part a) √	√	√
		Part b) √	Part b) √	√	√

HafizahHusain, BadariahBais, AiniHussain & Salina AbdulSamad (2012). How to Construct Open Ended Questions. https://www.sciencedirect.com/science/journal/18770428

6 → Send to Panel Review for Improvement

7 → Revise

8 → Send to Director to endorse

Validity & Reliability Compliance

MPPR1333
Introduction to Scaling and Instrumentation

UNIVERSITI TEKNOLOGI MALAYSIA

UNIVERSITI TEKNOLOGI MALAYSIA
FAKULTI PENDIDIKAN

FINAL EXAM
(TAKE HOME EXAM)

SEMESTER II SESSION 2020/2021

INTRODUCTION TO SCALING AND INSTRUMENTATION

Five Hours

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO

Instructions:

1. This paper consists of **thirty (30) multiple choice** questions.
2. Answer **all** questions in teh answer sheet given.

WEIGHTAGE LOTS + HOTS FE=40%

			Case 1	Case 2
MCQ				
*LOTS	15	@2	30	**40**
*HOTS	25	@2	50	**40**
ESSAY				
*LOTS			5	**NA**
*HOTS			15	

LOTS	35 35/100*0.4
HOTS	65 65/100*0.4

LOTS	**16%**
HOTS	**24%**

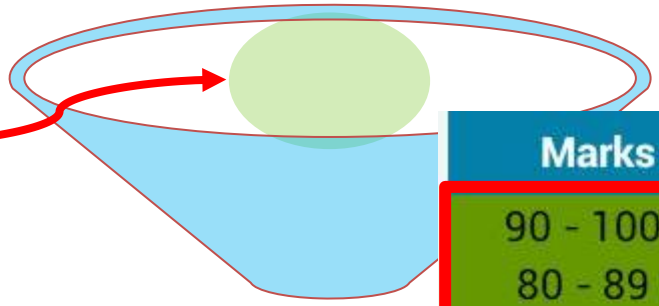
LOTS	40 40/80*0.4
HOTS	40 40/80*0.4

LOTS	**20%**
HOTS	**20%**

WEIGHTAGE & MASTERY

		CASE 1	CASE 2
FE	HOTS	24%	20%
	LOTS	16%	20%

ASSIGNMENT	60%



MASTERY

Marks	Grade	Point
90 - 100	A+	4.00
80 - 89	A	4.00
75 - 79	A-	3.67
70 - 74	B+	3.33
65 - 69	B	3.00
60 - 64	B-	2.67
55 - 59	C+	2.33
50 - 54	C	2.00
45 - 49	C-	1.67
40 - 44	D+	1.33
35 - 39	D	1.00
30 - 34	D-	0.67
00 - 29	E	0

Q&A

Remember….

We are measuring the effects of teaching

Assessment must reflect what is taught

Teaching must reflect what is to be assessed

References

Anderson, L.W., & Krathwohl, D.R.(eds.). (2001). A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.

Bloom. B.S. (Ed.) (1956) Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain. Longmans, New York.

Biggs. J. (2003) Teaching for Quality Learning at University – What the Student Does 2nd Edition SRHE / Open University Press, Buckingham.

Chase, C.I. (1999). Contemporary assessment for educators. New York: Longman.

Downing, S. M., & Haladyna, T. M. (2006). Handbook of Test Development. New York, NY: Routledge.

Holly Lavender and Cathryn Shaw (2013). Understanding Test Blueprints and Their Impact on Student Growth. Ohio department of Education