

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



1999م

كلية طب الكندي

1420هـ

# Assessment and Evaluation



**Prof. Dr. Yousif AbdulRaheem**

**FICMS/CM**

**Consultant in Preventive Medicine & Public Health**

**Department of Community & Family Medicine**

**Al Kindy College of Medicine**

**University of Baghdad**

# Objectives

At the end of this lecture we should be able to:

- 1. Recognize the importance of assessment in medical education**
- 2. Differentiate between assessment and evaluation**
- 3. Appreciate the scientific step in different type of assessment methodologies.**
- 4. Distinguish the different types of educational measurement (validity, reliability, practicability, & positive impact )**
- 5. Describe basic ideas behind blueprinting**

LICENSE to DRIVE





# Knowledge: How we will test it?

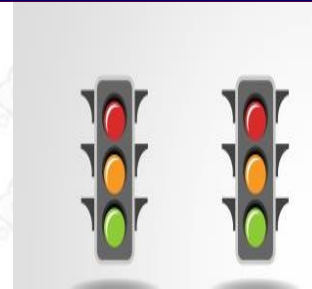


# Performance: How we will test it?





# Attitude: How we will test it?



# How we will be sure that we are testing the right domains?

## Indicators for proper license giving

- Number of accidents
- Number of drivers wearing seat belt
- Other Driver behaviors
- Number of hospital admission because of traffic accidents
- National morbidity & mortality for traffic accidents.



Being involved in student assessment is among the most critical of all tasks facing the teacher.



Generally, teachers take such involvement quite seriously but, sadly, the **quality of many assessment and Evaluation** procedures remains **undesirable**

◆ **Assessment:** A system of measurement for professional accomplishments in light of educational objectives, using defined criteria

✓ **By grading on a rough scale**

✓ **Or by assigning numerical values.**

➤ **Assessment refers to assessment of medical **student performance****

◆ **Evaluation:** A process that attempts to systematically determine the **relevance**, **effectiveness**, and **impact** of **educational activities** in light of their objectives.

➤ **Structure**

➤ **Process or**

➤ **Outcome**

➤ Evaluation refers for faculty, staff, learning environment, curriculum, assessment process, and program evaluation



# A Continuum, or Levels, of Educational Outcomes/Results



- ◆ **Student performance** (Attendance, Satisfaction, Learning (Knows), Competence (Skills), Performance/Behavior in Practice - Individual and Organizational (Does))
- ◆ **Teacher Achievement** (Degree, research, community participation...etc)
- ◆ **College Products**: post graduates, research center, Counseling Clinic...etc)
- ◆ **Patient Outcomes**
- ◆ **Community Health Status**

➤ **Evaluation tools** (Questionnaires, Focus groups, Thesis project, Patient feedback, Allied healthcare professionals' assessment, Peer evaluation & Self assessment)



**Take decision to  
improve**

**REPORTING  
(Documentation)**

**Activity EVALUATION**

**Student ASSESSMENT  
(Various strategies)**

**College & Curriculum  
Expectations Aims-  
Objectives**

*Feedback*





➤ How will we know if our students have achieved the desired learning outcomes?

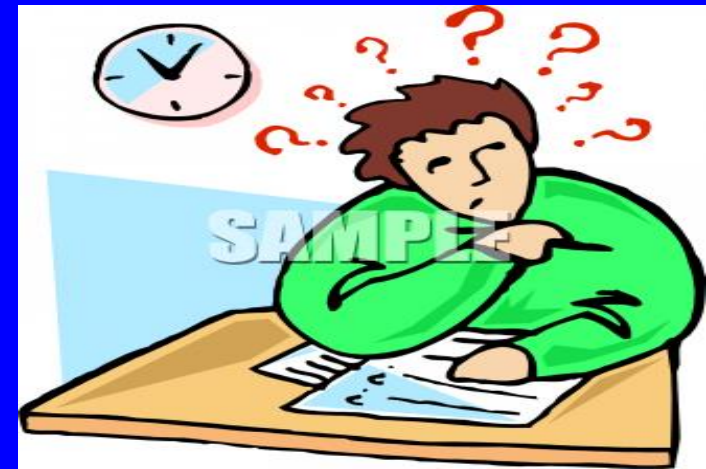


➤ How will we measure the extent to which they have achieved these learning outcomes?



# Assessment

- The process of documenting, usually in measurable terms, knowledge, skills, attitudes and beliefs of the student and decide whether the student can pass to up step grade.
- Most powerful drive for student learning behavior
- **Silent killer of learning**

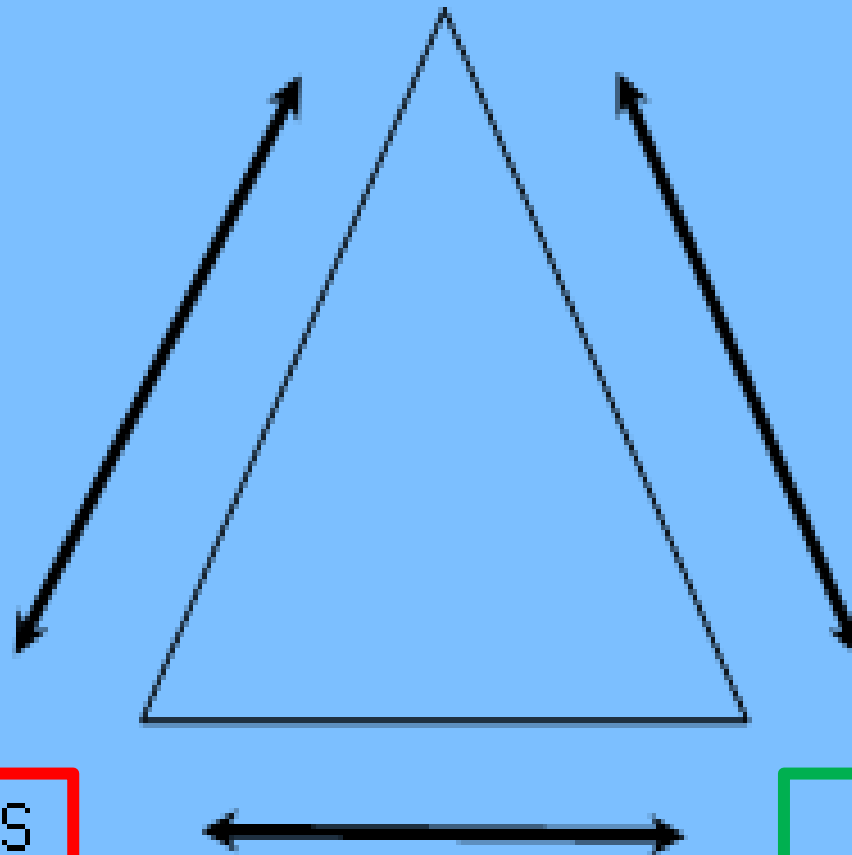


**We should assess what  
we taught and teach  
what we will assess.**





# Learning Outcomes



Learning activities

Assessment

**Triangle of Effective Learning**

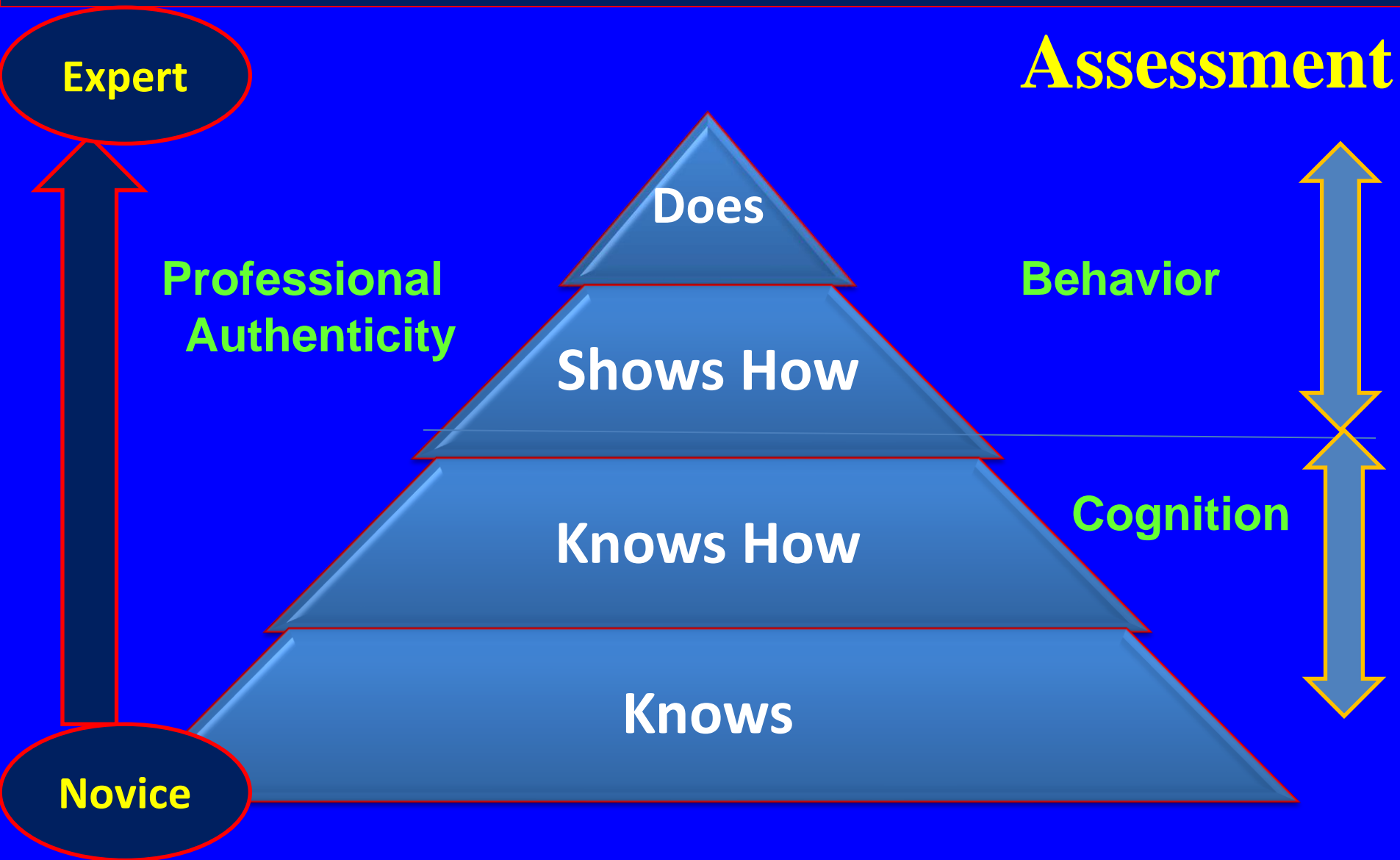


# Linking Learning Outcomes, Teaching and Learning Activities and Assessment

If the learning outcomes are clearly written, the assessment is quite easy to plan!



# MILLER'S PYRAMID





# Keeping in mind.....



- The Primary Purpose of Assessment is to Improve **Student Learning**
- Assessment should be used to improve educational activity
- Medical schools should take methods of assessment that fits with their curriculum.
- Always assess **deep learning**

# Purposes of Assessment:

- ✓ Ranking of student.
- ✓ Measuring improvements over time
- ✓ Evaluating the effectiveness of the course
- ✓ To diagnose student difficulties.
- ✓ Evaluation of teaching method.
- ✓ Motivating student to study.
- ✓ Assuring the achievement of learning objectives
- ✓ Safeguarding the public



# Values and Attitudes about Assessment

- ✓ Sharing learning goals with the students.
- ✓ Providing feedback that helps students recognize their next steps and how to take them.
- ✓ Providing students with examples of what we expect from them.
- ✓ Involving students in self-assessment.
- ✓ Being confident that every student can improve.

# What should we assess?

## LEARNING COMPONENTS or LEARNING OBJECTIVES

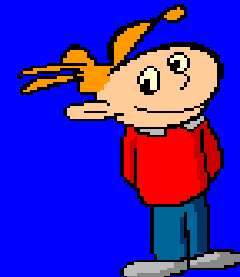
**A.** Knowledge (Cognitive)



**B.** Skills (Psychomotor)



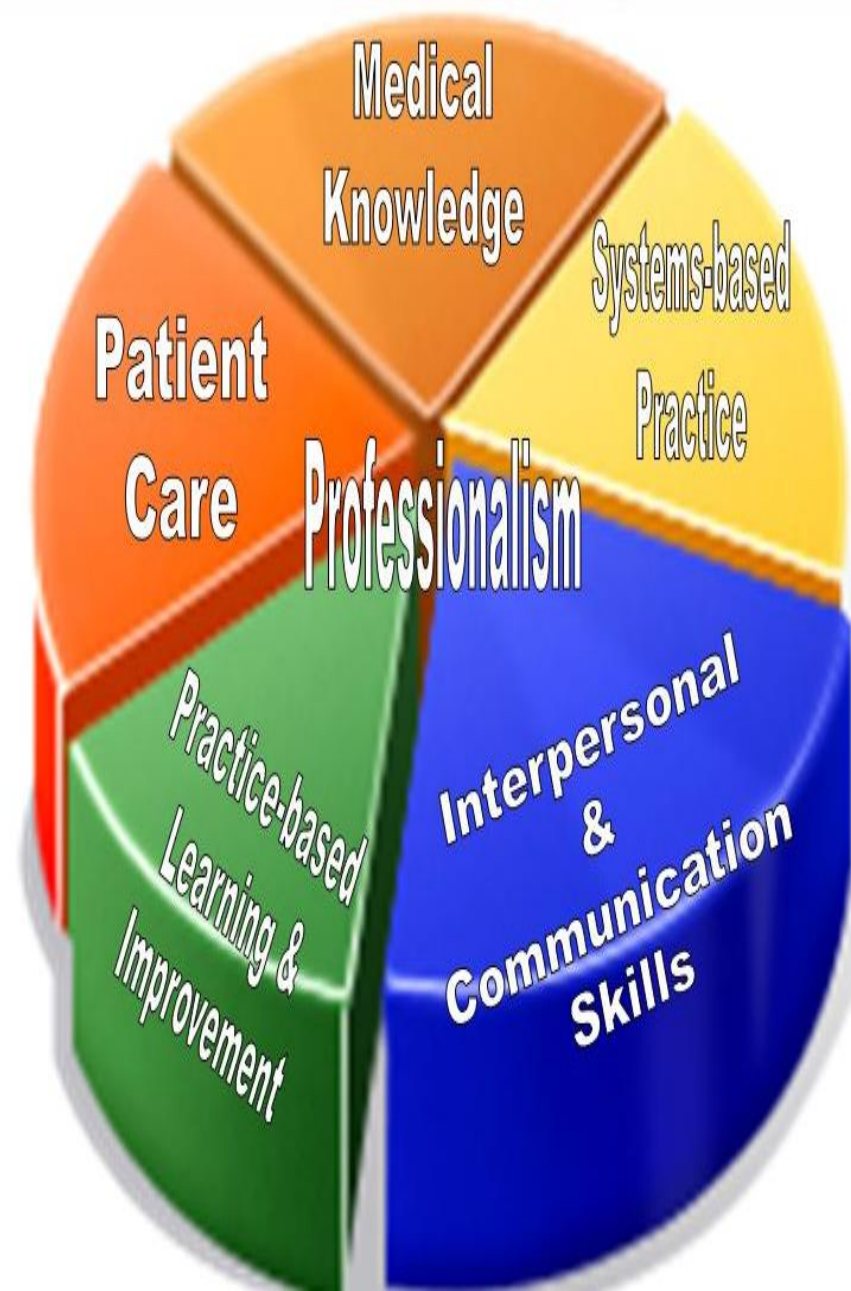
**C.** Attitudes (Affective )





# Standard Graduate Competencies

- ✓ Medical knowledge
- ✓ Patient care (skills)
- ✓ System-based practice
- ✓ Evidence-based & continuous self learning & improvement
- ✓ Inter-personal and communication skills
- ✓ Professionalism



# Assessment methods (AM) Educational Measurements



## Written:

- ◆ **MCQ** - Multiple Choice Questions
- ◆ Traditional Essay
- ◆ **MEQ** – Modified Essay Questions
- ◆ **EMI** - extended matching items
- ◆ **FB** - Filling the blanks
- ◆ **TF**- True or false questions
- ◆ **PMP**– Patient Management Problem

# Assessment methods (AM)

## Practical

- ◆ Oral, viva
- ◆ Short & long case Traditional practical
- ◆ OSCE
- ◆ ACC – Assessment of Clinical Competence
- ◆ SP - Standardised Patient

## Others

- ◆ Portfolio
- ◆ Log Book
- ◆ Research,
- ◆ 360 global rating

# WHAT WE SHOULD KNOW ABOUT EDUCATIONAL MEASUREMENT

- The methods of assessment should satisfy the following requirements:
  - 1) **Validity**: Does it measure what it is supposed to measure?
  - 2) **Reliability**: Does it produce consistent results?
  - 3) **Practicality**: Is it practical in terms of time and resources?
  - 4) **Positive impact** on learning.



# Validity: What does validity mean?

- Assessment is valid if it actually measures what it is supposed to measure.
- Validity depends on Learning Outcomes
- Ex1: Are MCQs valid to assess the skill of the student to insert IV cannula?
- Ex2: Is OSCE exam valid to assess the ability of the student to interpret ECG findings

# Types of validity

- ❖ Face validity-
- ❖ Content validity
- ❖ Construct validity
- ❖ Criterion-related validity



1) Face validity: Test overall appearance

Ex1: Medical exam not surgery exam

Ex2: Neuro exam not dermatology exam

Ex3: undergraduate not postgraduate exam

**2) Content validity:** It is a measure of the degree to which the assessment contains a **representative sample** of the material taught in the course (**according to the weight of each topic**)

➤ **Ex1:** Does Final exam in medicine for 4<sup>th</sup> or 5<sup>th</sup> or 6<sup>th</sup> year assess the whole or most course topics according to their weight ?

**3) Construct validity:** Seeks agreement between a theoretical concept and a specific measuring procedure.  
disability questionnaire

➤ **Ex1:** A test of intelligence must include measures of multiple intelligences, rather than just logical-mathematical and linguistic ability.

➤ **Ex1:** A test for “disease management” should include measure the diagnosis and treatment?

## 4) Criterion-related validity

- Measures how well one measure predicts an outcome for another measure. A test has this type of validity if it is useful for predicting performance or behavior in another situation (past, present, or future).
- Ex1: If Final year exam test accurately predicts how well the physician will perform on the job, the test is said to have criterion validity.



# Reliability: What does mean?

- Is the extent to which an experiment, test, or any measuring procedure shows the same result on repeated trials. four key types of reliability are:
  - **“Stability”**: related to time consistency
  - **“Internal”**: related to the instruments
  - **“Inter-rater”**: related to the examiners’ criterion
  - **“Intra-rater”**: related to the examiners’ criterion

# PRACTICALITY

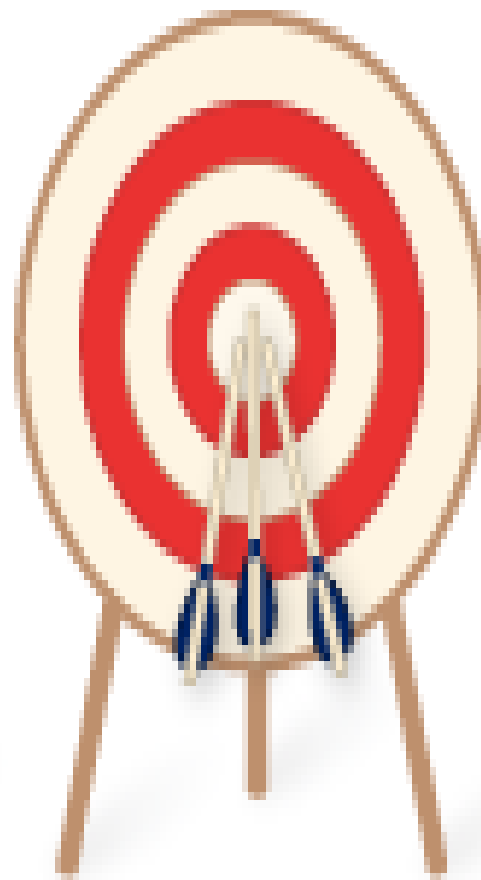
It refers to the economy of time, effort and environment in testing. In other words, a test should be...

- Easy to design
- Easy to administer
- Easy to observe
- Easy to mark
- Easy to interpret  
(the results)

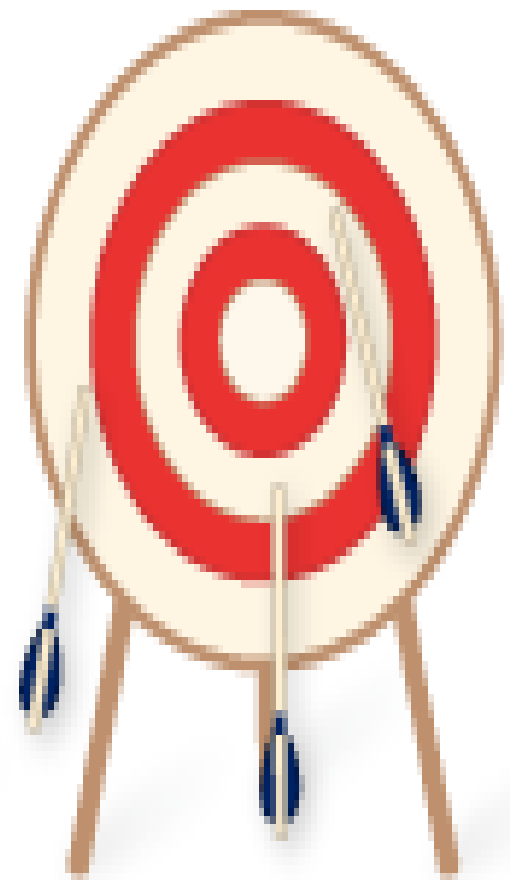




**Reliable but  
not valid**



**Reliable  
and valid**



**Unreliable and  
hence not valid**

# Reliable and Valid



# When should students be assessed?

- During the course of training
- At the end of the course



## Assessment – Two Types

- ◆ Formative (Assessment for learning)
- ◆ Summative (Assessment of learning)



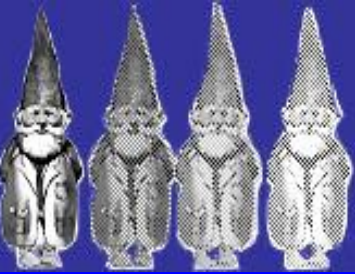


# 1. Formative assessment (Diagnostic)

- No credit
- Done during the course.
- Provide feedback to the student on his/her progress.
- Provide feedback to teacher with data for modification of his/her teaching.

# 2. Summative assessment (Certifying)

- With credit
- Done at the end of period.
- For placing student in order of merit



# Formative vs. Summative Evaluation

## Formative

## Summative

**Purpose**

Improvement

Judgment

**Timing**

Throughout

End

**Uses**

Give feedback

Grades

**Competence**

Identify Strengths and  
weaknesses

Judge

Develop plan

Certify

# The Garden Analogy

- If we think of our students as plants ...
- **Summative assessment:** Simply measures the end product not the growth of the plants.
- **Formative assessment:** Equivalent of feeding and watering the plants appropriate to their needs for growth.



# What can be assessed?

- 1) **Medical knowledge**
- 2) **Patient care (Clinical skills + Ethical)**
- 3) **Communication skills ( patient approach + Emotional reactions)**
- 4) **System Based approach (Problem-solving skills & Interpretations)**
- 5) **Evidence based Practice & continuous self learning (updating & research)**
- 6) **Professionalism**

# What can be assessed?

	Knowl gde	Patient Care	System BA	Com Skill	EBM	Prof
Oral	++	+	+/-	--	+ -	+
Pract.	+	++	++	+ -	+	+ -
MCQ	++	+ -	+ -	--	+ -	--
MEQ	+	++	++	+ -	--	--
Essay	++	+ -	+ -	+ -	+ -	--
OSCE	+	+	+ -	+ -	++	+ -

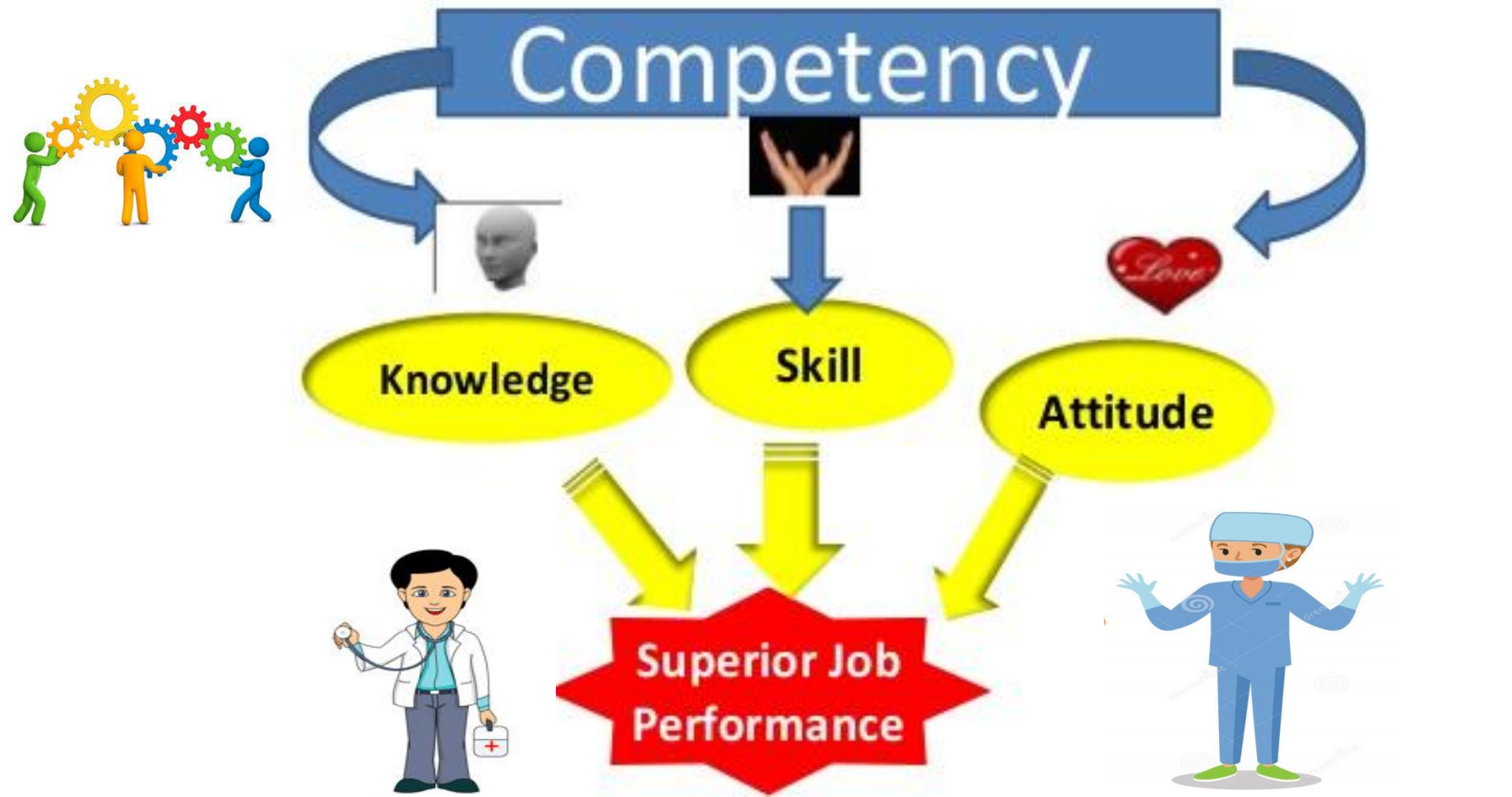
**No single test alone can assess  
all these domains**



# COMPETENCY VARIABLES

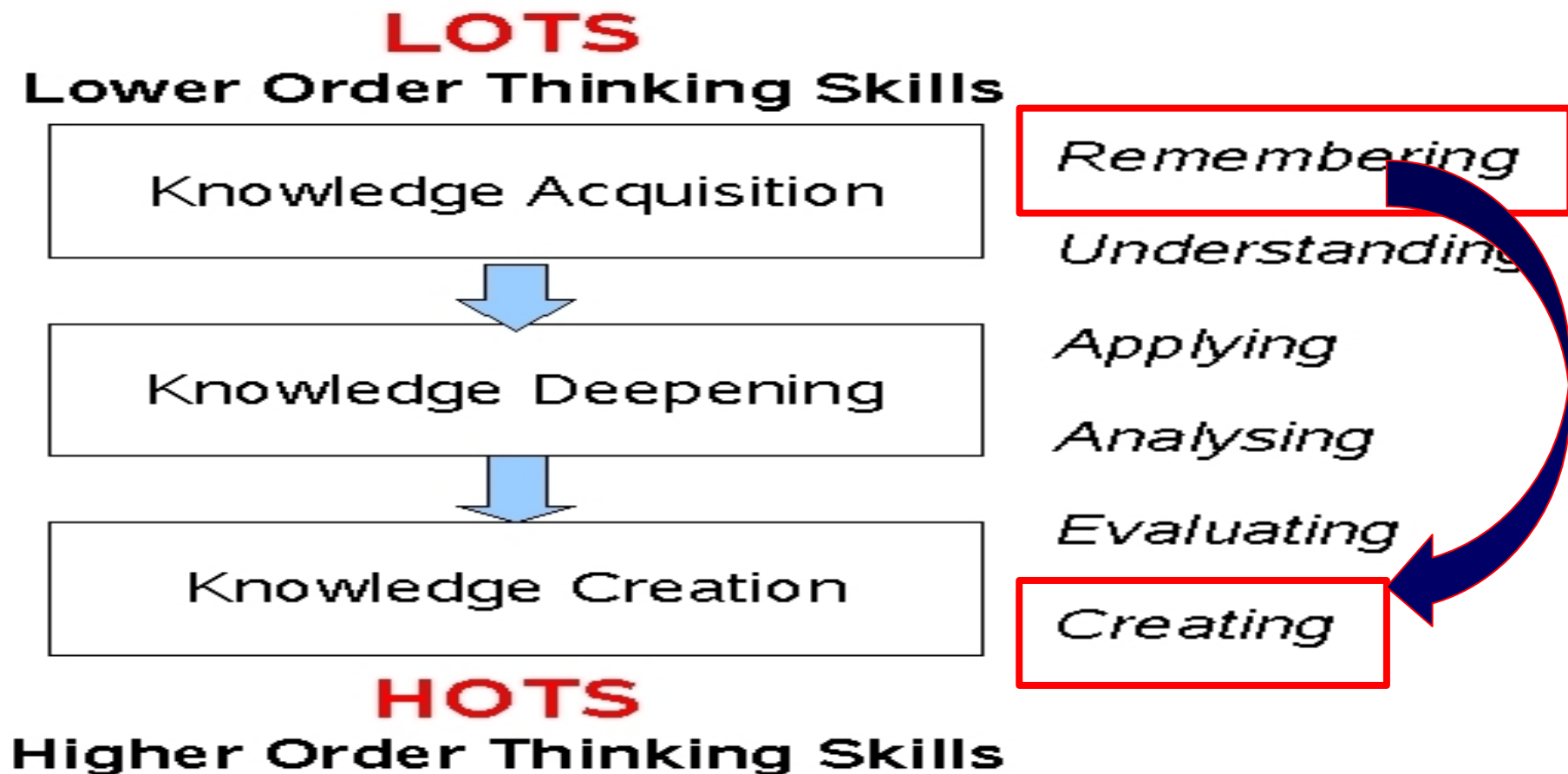
- ✓ skills
- ✓ knowledge
- ✓ assessment methods
- ✓ evidence
- ✓ context

Competency=Knowledge x Skill x Attitude[KSA]



# Bloom's taxonomy: Knowledge

- Ability to retrieve and handle information
- With understanding of basic and clinical sciences



Combining parts to make a new whole

**Create**

Judging the value of information or ideas

**Evaluate**

Breaking down information into component parts

**Analyze**

Applying the facts, rules, concepts, and ideas

**Apply**

Understanding what the facts mean

**Understand**

Recognizing and recalling facts

**Remember**

# **Bloom's Taxonomy**



# What the doctor is able to do?

## Skills

- **Competence in clinical skills**
- **Competence to perform procedures**
- **Competence to investigate a patient**
- **Competence to manage a patient**
- **Competence in health promotion and disease control**
- **Competence in skills of communication**

# How the doctor approaches their practice? **Attitude**

- With appropriate attitudes, ethical understanding and understanding of legal responsibilities.
- With appropriate decision making skills and clinical reasoning and judgment
- Attitude for personal development

**How we will assess all that?**



# Exercise #1

- Use the learning objectives in your module and identify the knowledge, skills and attitude for each learning objective
- Suggest the best method for assessment for each domain
- Large group discussion





# **What is a Blue Print?: means 'detailed plan of action'**

- **A map and a specification for an assessment program that ensures that all aspects of the curriculum and educational domains are covered by valid assessment.**
- **Simply, Blueprint links assessment to learning objectives.**
- **It also indicates the marks (weightage) for each content and test.**

➤ **Effective assessment is possible only when the Course objectives, Competencies to be tested & Method of assessment attain the 'Best fit' and clear to the student.**

**Blue print** makes the assessment valid, clear, explicit and transparent to everyone involved in the process of learning.



<b>Content/system/topic</b>	<b>Long essay (10 marks)</b>	<b>Short essay (5 marks)</b>	<b>Short answer (3 marks)</b>	<b>MCQs (1 mark)</b>	<b>Must/desirable/nice to know category</b>	<b>Total marks</b>
Cardiovascular system	1			3	Must know	13
Respiratory system		1	1	4	Must know	12
Gastrointestinal system		2		2	Must know/desirable to know	12
Renal system	1			2	Must know/desirable to know	12
Hepatobiliary system		2		2	Must know/desirable to know	12
Endocrine system		1	1	2	Must know/nice to know	10
Female reproductive and breast		1		3	Must know/desirable to know	8
Musculoskeletal and soft tissues tumors		1		2	Must know/desirable to know	7
Male reproductive and LUT			1	3	Must know	6
CNS/PNS/eye/skin			1	2	Desirable to know	5
Lymphoreticular System			1		Must know	3
Total number of questions	2	8	5	25		100

- **It is a two dimensional table of test specification (eg. Topics taught and methods used)**
- **Or a three dimensional one (eg. Topics, Methods and Types of questions).**
- **It helps to match various competencies with the course content and the appropriate modality of assessment.**
- **It makes assessment 'fair' to the students as they can have clear idea of what is being examined and can direct**

- It is a three dimensional chart which shows the placement of each question in respect of the objective and the content area that it tests.
- The aim of blueprinting is to reduce the two major threats to **validity**:  
Content & Construct
- And to increase the **reliability** by increase the number of test items. The calculate a realistic time for answering each test item and summed up for the whole test.





**If you don't  
know where  
you're going,  
any road'll  
take you there.**

**-George Harrison**

Learn more at  
[SpiritualCleansing.org](http://SpiritualCleansing.org)



# Preparation of a Blueprint:

- 1) Define the scope of assessment (end semester, final). Whether it is Formative / Summative assessment?**
- 2) Identify the purpose of assessment (learning objectives)**
- 3) Recognize the domains of learning to be included. What domain is being assessed, knowledge/ skills/ attitude?**
- 4) What is the Level of students: Undergraduates /Post- Graduates?**



- 5) What is the time, infrastructure...**
- 6) Decide the methods to be adopted.**
- 7) Decide the weightage to be given to content areas, domains of learning and methods of assessment.**

**Weightage can be decided on the basis of two parameters**

- ✓ i. The perceived impact / importance of a topic in terms of its impact on health.**
- ✓ ii. The frequency of the occurrence of a particular disease or health problem**

# How do we provide relative weighting of the content?

Importance X Frequency

## ➤ Importance/impact

- 1) Non-urgent, little prevention potential
- 2) Serious, but not immediately life threatening
- 3) Life threatening emergency and/or high potential for prevention impact

## ➤ Frequency

- 1) Rarely seen
- 2) Relatively common
- 3) Very common (prevalent)

# Sample opinion on weighting

- ✓ Involve course coordinator, teachers, evaluation coordinators, previous learners, MOH expert
- ✓ Consensus method
- ✓ Readjust weighting if necessary



# Example of weighting

Content presentation	Importance	Frequency	I X F	Weight %
Hypertension	3	3	9	0.22
IHD	3	2	6	0.15
Heart failure	1	2	2	0.049
DM	2	3	6	0.15
Meningitis	3	1	3	0.073
Endocarditis	3	1	3	0.073
Osteo arthritis	2	1	2	0.049
Septicemia	3	2	6	0.15
Cellulitis/Skin infn	2	2	4	0.098
HIV	2	1	2	0.049
Total			41	

- 1) It is a well known fact that, no single method can be used to assess all domains. Hence, combination of various methods have to be used to improve the validity of assessment.**
- 2) One can be prepared one for the whole examination or can be separate for Written and Practical Examinations respectively**
- 3) The marking scheme should be clear so as to bring complete transparency to the assessment.**

<b>Content / System / Topic</b>	<b>Long Essay (10 Marks)</b>	<b>Short essay (5 Marks)</b>	<b>Short Answer (3 marks)</b>	<b>MCQs (1 mark)</b>	<b>Level of Knowledge</b>	<b>Total marks</b>
Cardiovascular System	1	1	1	2	Recall Application Synthesis	20
Respiratory System	-	2	2	4	Recall Synthesis	20
Gastrointestinal System	1	1	-	5	Comprehension Application Evaluation	20
Renal System	-	2	1	7	Comprehension, Analysis	20
Hepatobiliary System	-	2	1	7	Recall Application,	20

<b>Content / System / Topic</b>	<b>History taking</b>	<b>General Physical Examination</b>	<b>Systemic Examination</b>	<b>Investigations</b>	<b>Patient Management</b>
<b>Cardiovascular System</b>	3		3		
<b>Respiratory System</b>		3			3
<b>Renal System</b>		3		3	
<b>Central Nervous System</b>	3		3		
<b>Hematology</b>				3	3



# Learning Objectives

Department of ....., CVS Module

	Knowledge	Practice	Attitude
Year 1			
Year 2			
Year 3			

# Learning Objectives

Department of .....

	Knowledge	Practice	Attitude
Year 4			
Year 5			
Year 6			

# Each year has learning objectives



<b>Medical Content Category</b>	<b>% of Exam</b>
Allergy and Immunology	2%
Cardiovascular Disease	14%
Dermatology	3%
Endocrinology, Diabetes, and Metabolism	9%
Gastroenterology	9%
Geriatric Syndromes	3%
Hematology	6%
Infectious Disease	9%
Nephrology and Urology	6%
Neurology	4%
Obstetrics and Gynecology	3%
Medical Oncology	6%
Ophthalmology	1%
Otolaryngology and Dental Medicine	1%
Psychiatry	4%
Pulmonary Disease	9%
Rheumatology and Orthopedics	9%
Miscellaneous	2%
<b>Total</b>	<b>100%</b>

# Medicine 4<sup>th</sup> year

	Knowledge			Practice		Total Wt	
	Should	Preferable	Worth	Should	Preferable	K	S
	70-75%	20%	<5%	90%	10%		
Res	Community Based Education			Common Clinical Presentation		10 h	8hr
GIT						20 h	12 hr
Total						120	120

# Medicine 4<sup>th</sup> year

	Knowledge		
	Should	Preferable	Worth
	70-75%	20%	<5%
Resp	Community Based Education 10 lectures + 2 seminars 20%		
GIT	15 lectures + 1 seminars 15%		
Renal	10 lectures + .... 10%		
Total	60 lectures + seminars 100%		

# Medicine 4<sup>th</sup> year

## Skills

**Should**

**Preferable**

**90%**

**10%**

System	History	Explanation	Exam	Procedure
CVS	Chest pain	Discharge drugs	Cardiac	BP
RS	Haemoptysis	Smoking	Respiratory	Peak flow
GIS	Abdominal pain	Gastroscopy	Abdominal	Rectal Examination
Repro	Absent menses	Abnormal smear	Pelvic	Cervical smear
NS	Headache	Tension headache	Eyes	Ophthalmoscopy
MS	Backache	GALS	Hip	Show manual handling



# Medicine 4<sup>th</sup> year Exam Blueprint

	Knowledge			Practice		
	MCQs	Essay	Other	Short	Long	OSCE
CVS	7	5%	0	5	5	1
Res	4	10%	0	5	0	2
GIT	5	15%	0	0	5	2
Rnal	4	0	2	0	0	1
Total	50%			50%	100%	

# CVS Module 3rd year Exam Blueprint

Topics	Knowledge			Practice		
	MCQs	Essay	Other	OSBE		
Anat	7	5%	0	5		
Phys	4	10%	0	5		
Bioch	5	15%	0	0		
Micro	4	0	2	0		
Total	50%			50%	100%	

# Complaining of undergraduate Medical Students after the examinations

## I. Theory Examinations

- “Too Lengthy paper..... Time was not enough to write”.
- “All Questions from few Systems only! No Question from.....”.
- “Questions were too vague, What to write ?
- Write short note!! Every think, What to cut?”
- Postgraduate questions!!!!

# Complaining of undergraduate Medical Students after the examinations

## II- Practical Examinations

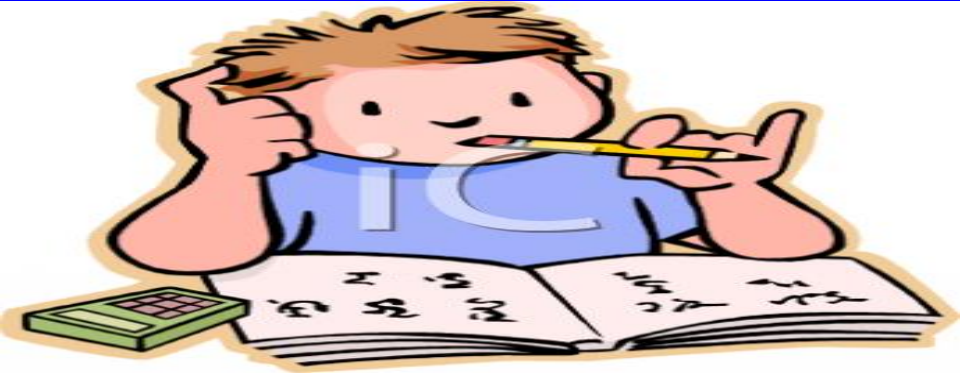
- The exam is not well prepared
- Time for station too short or too long
- We have never told about this exam
- “ My case was the most difficult”
- “I had never seen this case before”
- “ Most of the theory questions, long case , short case and viva questions, all from one/few systems only”

# Why it happens?

- The intended learning outcomes are not stated clearly
- The content of what to assess is left to the decision of one or two Examiners.
- Exam paper is set by one teacher
- Practical Examinations are conducted by some other teacher, without any Co-ordination (most of the times).
- The Examiner imparts instruction according to what 'she/he thinks is appropriate or important'.

# Exercise #2

- ✓ Use a worksheet & start working on a blueprint for your purpose
- ✓ Get together with colleagues and choose one blueprint to work on
- ✓ Large group discussion



# Planning and designing an exam

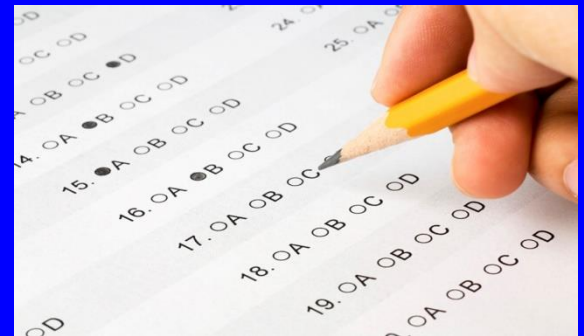
- Start before the course/Module begins
- Construct a blueprint
- Create questions that match the objectives of the course/ module
- Discuss with colleagues to review questions (team work)
- Remember to weight more important & prevalent topics
- Focus on concepts, not trivia
- Focus on the application of knowledge, not recall



# Most common methods of assessment

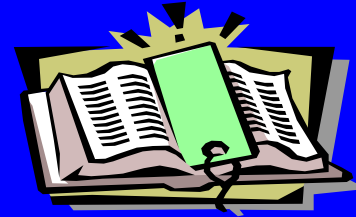
**1- Multiple Choice Questions:** They are designed to test **knowledge**

- ✓ Outcomes assessed- knowledge of basic skills and clinical sciences
- ✓ Patient investigation
- ✓ Patient management
- ✓ Health promotion
- ✓ Clinical reasoning and judgment



# How to construct MCQs

- State the question positively
- Construct options that are grammatically consistent with the item.
- Do not use MCQs when other instruments are more appropriate.
- Anatomy of the MCQ
  - ✓ **Stem**: contains the text
  - ✓ **Lead – In**: wanted information
  - ✓ **Options**: all answer choices
  - ✓ **Key**: the correct answer
  - ✓ **Distractor**: the incorrect answers



# Example

➤ A 65-year-old man comes to the physician for a follow-up examination after the results of a bronchoscopy showed squamous cell carcinoma. When the physician tells the patient the diagnosis, the patient becomes tearful and responds, “No, you’re wrong! This must be a mistake. This can’t happen to me. Let’s do more tests.” **This patient is most likely at which of the following stages of grief?**

- (A) Anger
- (B) Bargaining
- (C) Denial
- (D) Depression

# Example

A 1-year-old infant is known to have heart disease and is noted to be cyanosed. **Which of the following is the most likely diagnosis?**

- a. Atrial septal defect
- b. Patent Ductus Arteriosus
- c. Ventricular septal defect
- d. **Tricuspid atresia**



# Multiple choice questions- principles(rules)



- The question itself or the stem must be appropriate, and pose a clear question
- Have 1 correct response to the question
- Distracters or the incorrect responses among the choices should be plausible
- Each item should focus on an important concept
- Each item should assess application of knowledge, not recall of isolated fact
- All distractors should be homogeneous

# Strengths

- Efficient for large groups
- Can assess a range of thinking skills
- Minimal human intervention
- Grading is usually quick and straightforward
- Highly Acceptable tool for assessment across most of the academic disciplines.

# Weaknesses

- Difficult to write good questions
- It takes time to write & validate questions

# Best Evidence for Writing MCQs

- Invest in the stem; make sure the item can be answered without looking at the options
- Include language in the stem instead of repeating it in each option
- Each item should be clear and plausible
- Avoid using different grammar (e.g., verb tense) in stem and distractor
- Avoid negatively phrased items, and vague language (e.g., always, many)
- Avoid irrelevant items



# More MCQ Tips

- Randomly distribute the correct answer position throughout the test
- Present the options in a logical order (e.g., alpha, chronological)
- Allow 1-2 minutes per question, depending on the reading time
- Instructor should take exam first, then double or triple the time estimate

# Extended Matched Items (EMIs)

- **Purpose: Test knowledge and reasoning skills**
- **Identify the most appropriate course of action for the following patients**
- **Outcomes assessed: All outcomes with an emphasis on clinical reasoning**
  - ✓ **Patient diagnosis**
  - ✓ **Patient investigation**
  - ✓ **Clinical reasoning and judgment**
  - ✓ **Patient management**
  - ✓ **Health promotion**

# Extended Matching examples

- 1) Woman, mid 20s ,14 weeks pregnant to have glycosuria on routine resting in GP ante natal clinic
- 2) Woman, mid 50s , HT, BMI 30, just found to have fasting blood sugar 15 mmol/l, HBA1C 8.5%
- 3) Man, 45 years, smoker, BMI 35 complaining of fatigue, thirst & dry mouth



# Modified Essay Questions MEQ

- Can be used to assess the students' ability to integrate facts and problem solving skills, which are fundamental to good practice.
- Question requiring consolidation of facts and critical thinking will be asked.
- Patient management problem (PMP) consists of short questions that require also short answers that revolve around managing a patient.

- **They are intended to equip the student with knowledge, skills and attitude that are necessary in professional practice.**
- **The questions are asked about a case and options are provided that leads to proper management of the patient.**
- **Constructing a PMP begins with a description of the symptoms of the patient and the resources that are available, followed by a series of questions relating to management, diagnosis and investigations.**

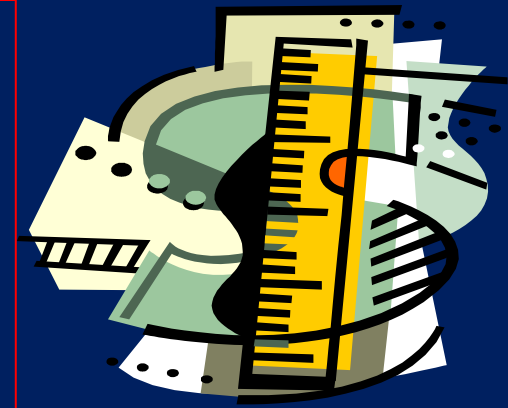
# Clinical Assessment



- ◆ Oral, viva
- ◆ Traditional practical (short & long cases)
- ◆ OSCE / OSBE
- ◆ ACC – Assessment of Clinical Competence
- ◆ SP - Standardised Patient

# What is an OSCE ?

## Objective Structured Clinical Examination



- Objective – means fair and without bias (e.g case difficulty).
- Structured refer to the organization of the examination.
  - ✓ Most examination in the world are not fair.
  - ✓ Use of checklist ensures objectivity.
  - ✓ What is your comment?



- **Several stations (8-20), reflect aspects or areas clinicians have to master.**
- **Station: This is the region where the skill is demonstrated by the candidate**
- **The assessors check against the competence shown in check list.**
- **All examination entails the clinical aspects. In other words the station are clinical in nature.**
- **Validity issues !!!**
- **Reliability!!!**

# OSCE blueprints and included in

- **History taking**
- **Physical examination**
- **Clinical reasoning**
- **Professionalism**
- **Procedural skills**
- **Communication skills**
- **Resting station**

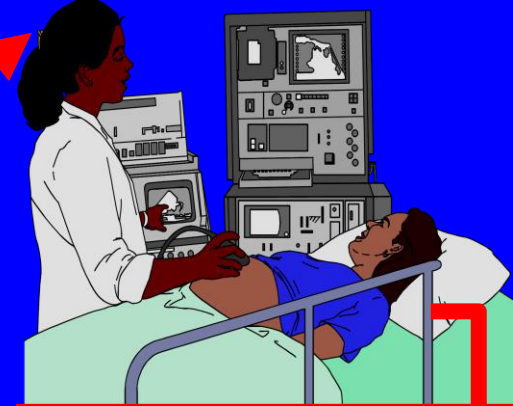
**IN**



**Station 1**



**Station 2**



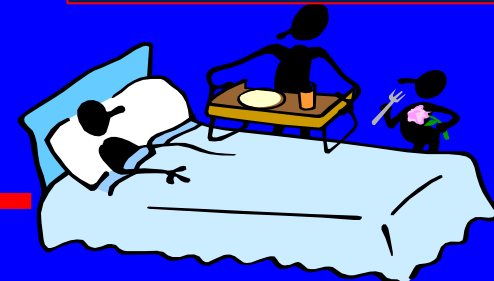
**Station 3**

**Station Rotation**

**Station 4**



**Station 5**



**Station 6**



**Station 7**



**OUT**

CanStock



# Typical observer sheet: Im Injection

Competence	Yes	No
<b>1. Greets patient</b>	<input checked="" type="checkbox"/>	
<b>2. Explains procedure</b>		<input checked="" type="checkbox"/>
<b>3. Assembles equipment</b>	<input checked="" type="checkbox"/>	
<b>4. Withdraws Medicine and expels air</b>	<input checked="" type="checkbox"/>	
<b>5. Swaps the correct area</b>		<input checked="" type="checkbox"/>
<b>6. Inserts needle horizontally and gently</b>	<input checked="" type="checkbox"/>	
<b>7. Propels medicine</b>		<input checked="" type="checkbox"/>

## *Standardized Patient History Taking Checklist (Sample Items)*

Y	N	1. What is the character of the pain ?
Y	N	2. Does the pain radiate ?
Y	N	3. Are there any precipitating factors ?
Y	N	4. Nausea? <b>OR</b> Sweating? <b>OR</b> Dyspnea ?
Y	N	5. Is there a history of smoking ?
Y	N	6. Family history ?
Y	N	7. Past medical history ?

# OSCE Limitations



- **Requires a lot of tedious planning**
- **Requires skilled personnel**
- **Requires substantial resources**
- **Demands a lot of dedication on the part of the participants**
- **Very difficult in large number of student**
- **It is a real Challenge! BUT... It is absolutely Worthwhile**

# Work-based (Clinical) Assessment

- The first rationale for WBA is related to observe the ability of acquisition the competencies required to enter independent practice (competency-based ME)
- Assessing learning acquired in the workplace (lab, Skill lab, hospital, PHCC).
- It influences on the process and outcomes of teaching and learning.





# Factors affect MEASUREMENTS

## Test Takers

Students fatigue, hunger, headache, emotional upset, anxiety, tends to reduce the consistency of the SCORE EXAM

## Test Itself (INTRA TEST)

Poorly constructed items, items with clues, very easy, very difficult, very high vocabulary reading level)—tends to guessing particularly when it is long

# **ERRORS OF MEASUREMENTS**

## **Test Administrations**

**Light, temp, noise, seating arrangement, instruction, time allotment, attitude of test examinee(MAKE THE TEST UNRELIABLE)**

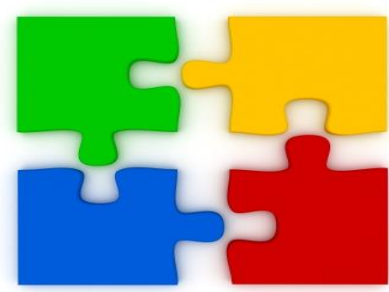
## **Test Scoring**

**Subjective scoring  
MISKEY/ providing wrong answer, Electronic mistakes**

# **Now that the exam is over: We need to**

- **Review exam items**
- **Analyze the students answers**
- **Flag items that did not perform well**
- **Talk to the module /course teachers to learn why items were flawed**

# The next step: Evaluating the test “ITEM ANALYSIS”



- Important step in the development of any assessment strategy
- Examine response patterns
- Examine item quality
- It is a statistical technique which is used for selecting and rejecting the items of the test on the basis of their difficulty value and discriminated power

# Three Elements in an Item Analysis

- 1) Examination of the facility (difficulty) level of the items, (MCQ, TF, matching)
- 2) Determination of the discriminating power of each item, and
- 3) Examination of the effectiveness of distractors in a multiple choice or matching items.

# Facility Value FV (Item Difficulty)

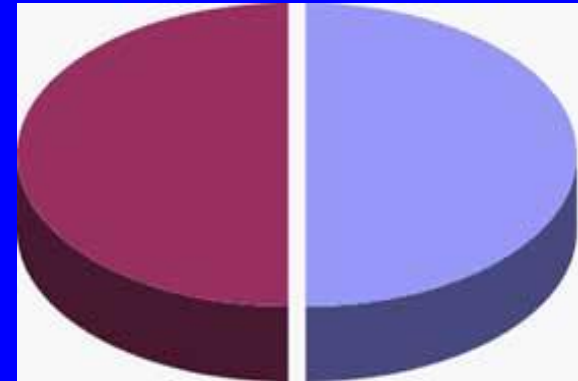
- Proportion of examinees selecting the correct response
- Determine item difficulty (0% - 100%)
- Easy = 85-90% answer correctly
- Difficult = 25-35% answer correctly
- Item analysis will show you this statistic for each item
- If an item was too easy!!!! Or was too difficult, revisit your lecture materials and update for the next iteration of your lecture

# Item Discrimination

- ID indicates correlation (+1 to -1) between the student's total score on the test (rank) and student selection
- Large, positive ID indicates students with higher scores answered correctly
- Low, positive suggest ability not related to success
- Negative indicates that low-achieving students performed better than high-achieving students

# STEPS

- Evaluate answer card - MARKING
- Arrange marks list in descending order
- Divide whole list into 2 half
  - ✓ High achiever
  - ✓ Low achiever
- If list is bigger –  
divide upper 30% or  
25% and lower 30% or  
25%





# Example

A 1-year-old infant is known to have heart disease & is noted to be cyanosed.

➤ Which of the following is the most likely diagnosis?

- a) Atrial septal defect
- b) Patent Ductus Arteriosus
- c) Ventricular septal defect
- d) Tricuspid atresia



	a	b	c	d
HAG(n=50)	0	02	08	40
LAG(n=50)	12	26	02	10

Easy:  $\geq 80\%$

Medium: 35%-80%

Hard:  $< 30$

➤  $FV = 50 / 100 = 50 \%$

Start test with easy (high FV) items towards more difficult items (low FV)

Good:  $\geq 0.4$

Medium:  $> 0.2 - < 0.4$

Low:  $\leq 0.2$

➤  $DI = \frac{2 \times (H - L)}{N}$

$2 \times (40 - 10) / 100 = 0.6$

Item with high DI are preferred. DI = 1 indicates item perfectly discriminates between those who knows from those who don't

# Distractor Efficiency


- Distractor should be PLAUSIBLE
- Any distractor that is not picked by at least 5% of student – is not a good distractor


	a	b	c	d
HAG(n=50)	0	02	08	40
LAG(n=50)	12	26	02	10

**GOOD**

**Change the distractor**

Item	Group		Answers				Correct Answers	Difficulty Index
			A	B	C	D		
1	H	20	3	14	2	1	21	?
	L	20	10	7	3	0		
2	H	20	0	0	18	2	27	?
	L	20	0	3	9	8		
3	H	20	3	8	4	4	10	?
	L	20	10	2	4	4		
4	H	20	3	3	4	10	14	?
	L	20	2	4	10	4		
5	H	20	15	2	2	1	16	?
	L	20	1	10	4	5		

Item	Group		Answers				Correct Answers	Difficulty Index
			A	B	C	D		
1	H	20	3	14	2	1	21	0.52
	L	20	10	7	3	0		
2	H	20	0	0	18	2	27	0.67
	L	20	0	3	9	8		
3	H	20	3	8	4	4	10	0.25 
	L	20	10	2	4	4		
4	H	20	3	3	4	10	14	0.35
	L	20	2	4	10	4		
5	H	20	15	2	2	1	16	0.40
	L	20	1	10	4	5		

Item	Group		Answers				Corr. Ans	Diff. Index	Discrim Index
			A	B	C	D			
1	H	20	3	14	2	1	21	0.52	0.35
	L	20	10	7	3	0			
2	H	20	0	0	18	2	27	0.67	0.45
	L	20	0	3	9	8			
3	H	20	3	8	4	4	10	0.25	0.30
	L	20	10	2	4	4			
4	H	20	3	3	4	10	14	0.35	0.30
	L	20	2	4	10	4			
5	H	20	15	2	2	1	16	0.40	0.70 
	L	20	1	10	4	5			

# Interpretation of the Difficulty Index

Range of DFI	Interpretation	Action
<b>0 – 0.34</b>	<b>Difficult</b>	<b>Revise or discard</b>
<b>0.35 – 0.79</b>	<b>Right difficulty</b>	<b>Retain</b>
<b>0.80 - above</b>	<b>Easy</b>	<b>Revise or discard</b>

# Interpretation of the DSI

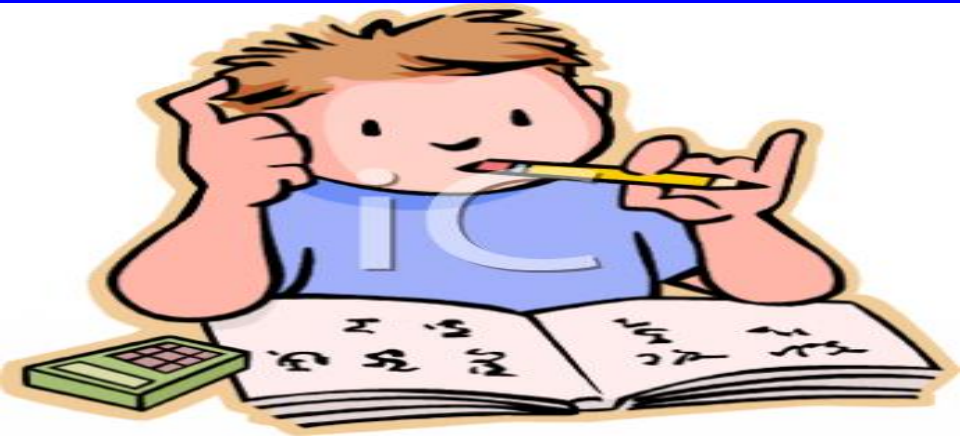
<b>Range</b>	<b>Verbal Description</b>	<b>Action</b>
<b>.40 &amp; above</b>	<b>Very Good Item</b>	<b>Include</b>
<b>.30 – .39</b>	<b>Good Item</b>	<b>Include!!</b>
<b>.20 – .29</b>	<b>Fair Item</b>	<b>Revise</b>
<b>.09 – .19</b>	<b>Poor Item</b>	<b>Discard</b>
<b>0</b>	<b>Very poor</b>	<b>Discard</b>
<b>-ve</b>	<b>Very poor</b>	<b>Discard</b>





# Examining Distractor Effectiveness

- **Distractors: To attract students who do not know the correct answer.**
- **An ideal item is one that all students in the upper group answer correctly and all students in the lower group answer wrongly.**
- **And the responses of the lower group have to be evenly distributed among the incorrect alternatives.**

- People who do not know the item will have responses equally distributed across the wrong answer.
- It is not desirable to have one of the distracters chosen more often than the correct answer



# Distracter Analysis for 50 student

Answer Choices	Selected Number	%	% Expected
Choice A	2	4	20 
Choice B	26	52	20 
Choice C	7	14	20
Choice D (Right)	15	30	40

- ✓ **A is chosen by only 4%; may be too obviously wrong.**
- ✓ **C is working well.**
- ✓ **B has drawn far too many students; should be examined to determine what caused this choice.**
- ✓ **Teacher should consider re-teaching the topic and explaining why B is wrong.**
- ✓ **A and B need to be rewritten for the next edition of the test.**

# QUESTION BANKING



- ✓ Large pool of question with set of information regarding question
- ✓ Goal – ability to deposit, discover and retrieve question.
- ✓ Random selection of questions according to the weight



# Exercise 3



**Construct item  
analysis for the  
following exam**



<b>Answer Choices</b>	<b>Upper 50%</b>	<b>Lower 50%</b>	<b>Total</b>
<b>Choice A</b>	<b>2</b>	<b>4</b>	<b>6</b>
<b>Choice B</b>	<b>2</b>	<b>4</b>	<b>6</b>
<b>Choice C</b>	<b>13</b>	<b>9</b>	<b>22</b>
<b>Choice D</b>	<b>3</b>	<b>8</b>	<b>11</b>
<b>Choice E (Correct)</b>	<b>30</b>	<b>25</b>	<b>55</b>

$$\begin{aligned} \text{DFI} &= (30+25)/100 \\ &= 0.55 \quad \text{Right Difficulty} \end{aligned}$$

$$\begin{aligned} \text{DSI} &= 2(30-25)/100 \\ &= 0.1 \quad \text{Poor Discard} \end{aligned}$$

**Distractor analysis** = choice C  
need to be revised



Medical  
student



# Final Message

- ✓ **Assessment is integral part of educational process**
- ✓ **No single test can assess all the student domains. Good assessment requires a variety of methods.**
- ✓ **No single method can test the whole medical competence and performance.**
- ✓ **Validity & Reliability**
- ✓ **Team work assessment**
- ✓ **Encourage external evaluator!!**

**Thank You**



**Professor  
Yousif AbdulRaheem**