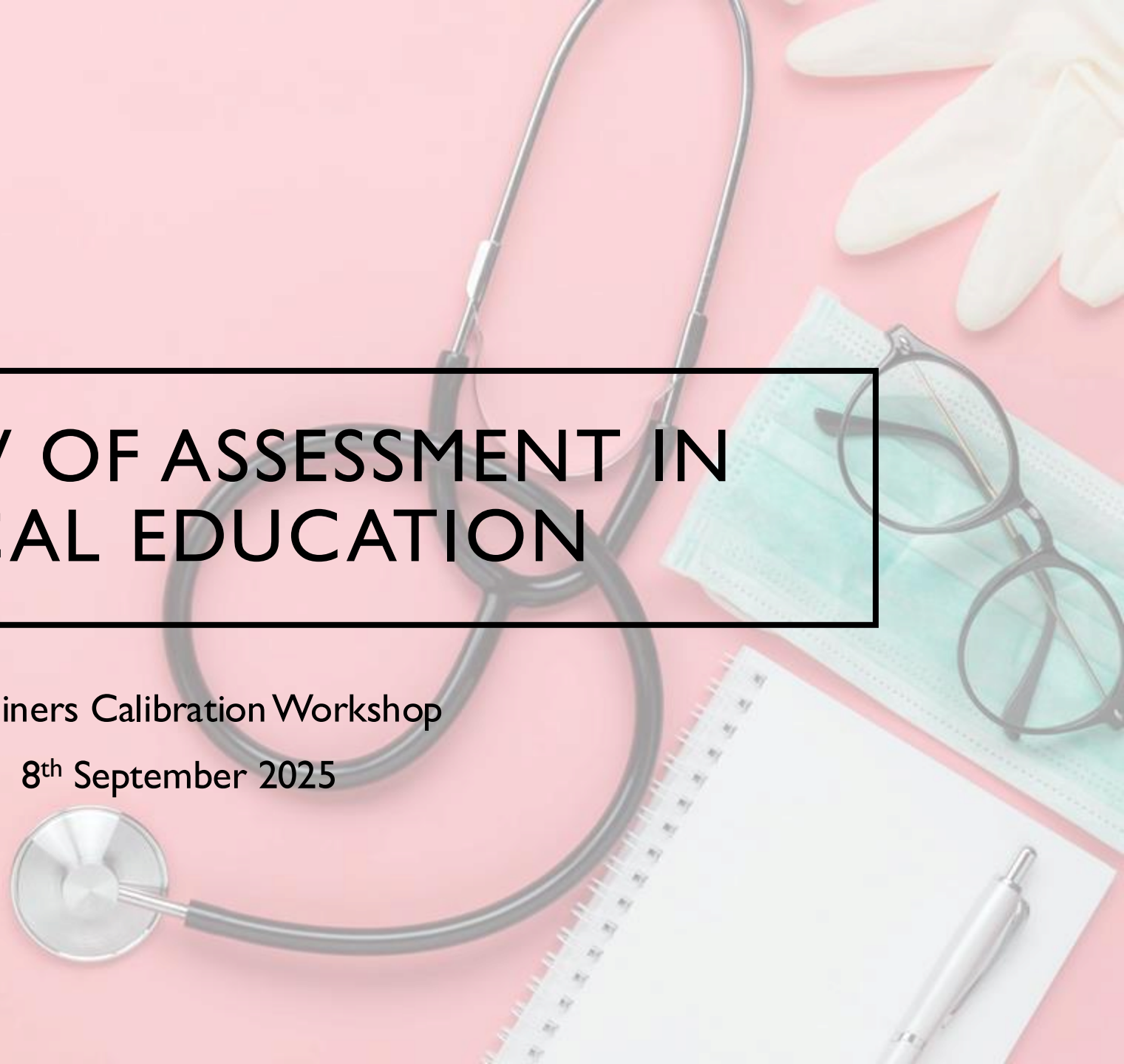
OVERVIEW OF ASSESSMENT IN MEDICAL EDUCATION

Examiners Calibration Workshop

8th September 2025



MEASUREMENT, ASSESSMENT & EVALUATION

Measurement

Quantification

Using instruments / scale

Nominal/ ordinal/ interval/ ratio

To apply a standard scale or measuring device to an object, series of objects, events, or conditions, according to practices accepted by those who are skilled in the use of the device or scale.

Assessment

Systematic process to measure or evaluate the performance of individuals to draw inferences (purposeful)

All test are assessment, not all assessment are test

Knowledge, skills, attitude

Evaluation

Using data from assessment to make inference about students

Grades, certificate, degree

Procedures used to determine whether the subject meets a predetermined criteria

WHY ASSESSMENT MATTERS

Ensures competence & patient safety

Drives learning behaviours

Evidence of programme quality

Accreditation compliance (MMC, MQF)

References: van der Vleuten, 1996

THE ASSESSMENT GOALS

Protect Public

To protect the public by identifying incompetent graduates

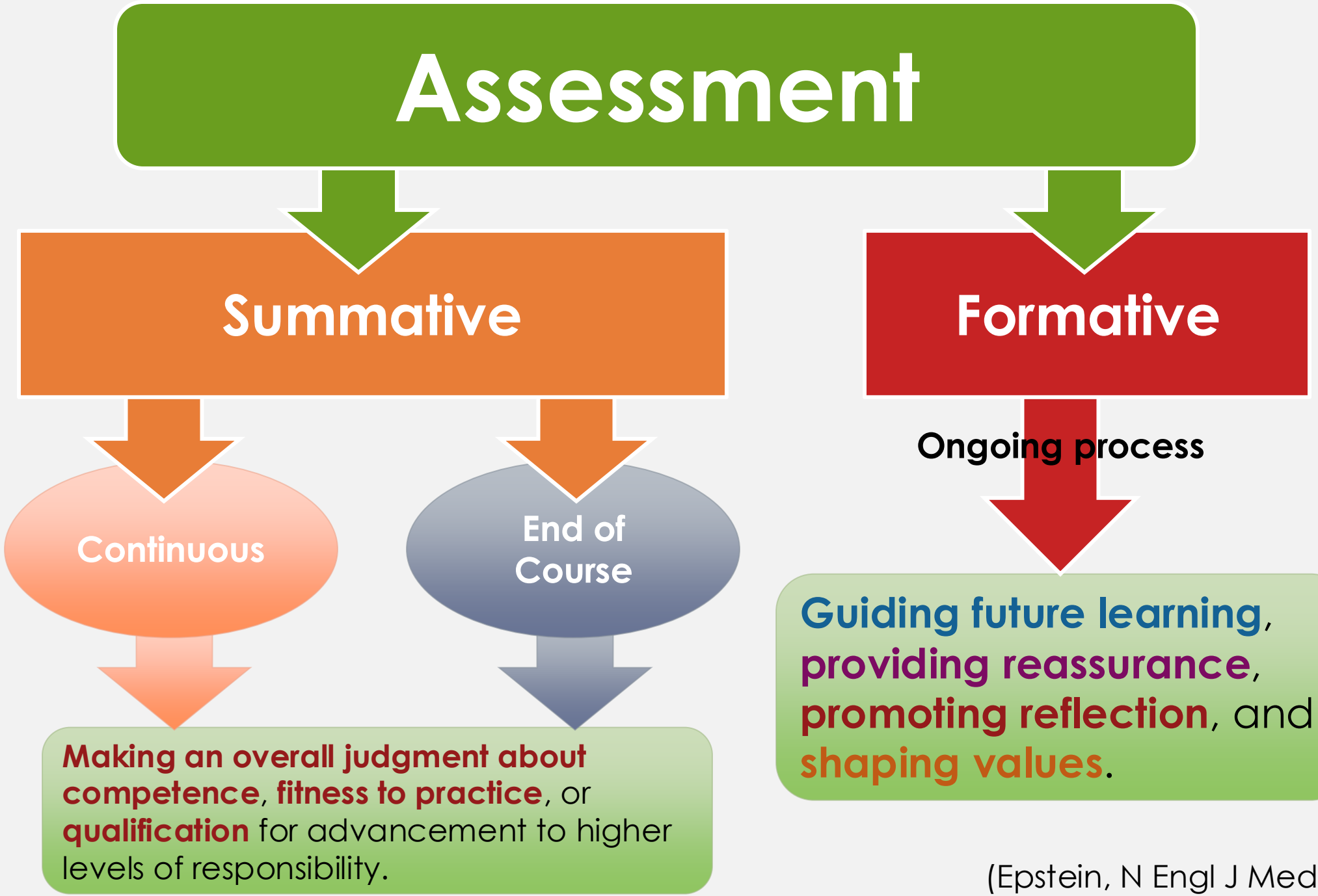
Further Training

To provide a basis for advanced training

Optimize capabilities

To optimize the capabilities of all learners by providing motivation and direction for future learning

(Epstein, N Engl J Med, 2007)



(Epstein, N Engl J Med 2007)

ASSESSMENT FOR, OF, AND AS LEARNING

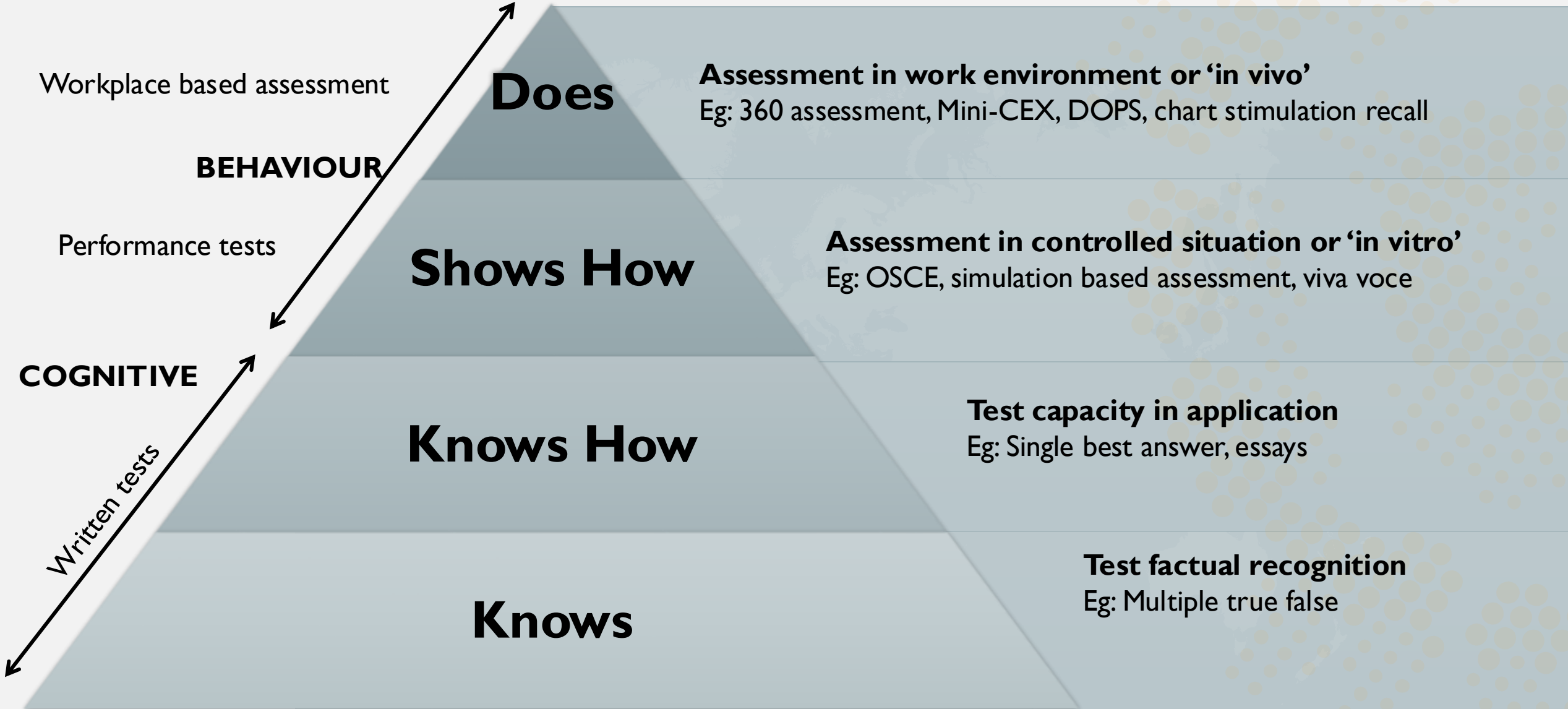
- Assessment FOR learning (formative)
- Assessment OF learning (summative)
- Assessment AS learning (reflection, self-regulation)

Table 1 Characteristics of Formative and Summative Assessments

Characteristic	Formative Assessment	Summative Assessment
Purpose	To improve teaching and learning	Evaluation of learning outcomes
	To diagnose student difficulties	Placement, promotion decisions
Formality	Usually informal	Usually formal
Timing of administration	Ongoing, before and during instruction	Cumulative, after instruction
Developers	Classroom teachers to test publishers	Classroom teachers to test publishers
Level of stakes	Low-stakes	High-stakes
Psychometric rigor	Low to high	Moderate to High
Types of questions asked	What is working	Does student understand the material
	What needs to be improved	Is the student prepared for next level of activity
	How can it be improved	
Examples	Observations	Projects
	Homework	Performance assessments
	Question and answer sessions	Porfolios
	Self-evaluations	Papers
	Reflections on performance	In-class examinations
	Curriculum-based measures	State and national tests

Miller's pyramid of clinical competence

(Miller, 1990)



ASSESSMENT OF PROFESSIONALISM

- Essential in medicine – ethics, values, teamwork
- Tools: STEPS, MSF, portfolios
- Feedback & coaching critical

References: Roslan et al., 2017

PRINCIPLES OF GOOD ASSESSMENT

- Validity – measures what it intends
- Reliability – consistent results
- Feasibility – practical in real settings
- Acceptability – acceptable to stakeholders
- Educational Impact – drives learning behaviors

References: van der Vleuten, 1996

Messick's framework (1995)

Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application

David A. Cook, MD, MHPE, Thomas J. Beckman, MD, FACP
Division of General Internal Medicine, Mayo Clinic College of Medicine, Rochester, Mi

Construct Validity

Content	Response Process	Internal Structure	Relation to Other Variables	Consequences
Relationship between test content & measured construct	**Cognitive and behavioural responses of students and raters to task**	**Data internal to assessment**	**Correlation between other relevant assessment**	**Intended / unintended impact of assessment -Methods of determining exam pass/fail score**
• Blueprint • Training of item writers • Vetting	• Familiarity with format • Calibration • SP training	• Item analysis • Factor analysis	• Correlation with similar assessment	• Standard setting • Performance in subsequent training • Patient outcomes

CHALLENGES IN ASSESSMENT

- Balancing validity vs feasibility
- Faculty training & calibration
- Student stress & fairness
- Resource-intensive logistics

References: Downing, 2004

TRENDS & INNOVATIONS

- Simulation-based assessment
- Entrustable Professional Activities (EPAs)
- Competency-Based Medical Education (CBME)
- AI and analytics for feedback

References: ten Cate, 2005; Frank et al., 2010

KEY TAKEAWAYS

- Assessment ensures safe, competent doctors
- Use multiple methods across knowledge, skills, attitudes
- Adopt programmatic, feedback-rich approaches
- Anchor decisions in standards, evidence, and fairness

THANK YOU

Observed Long Case Calibration

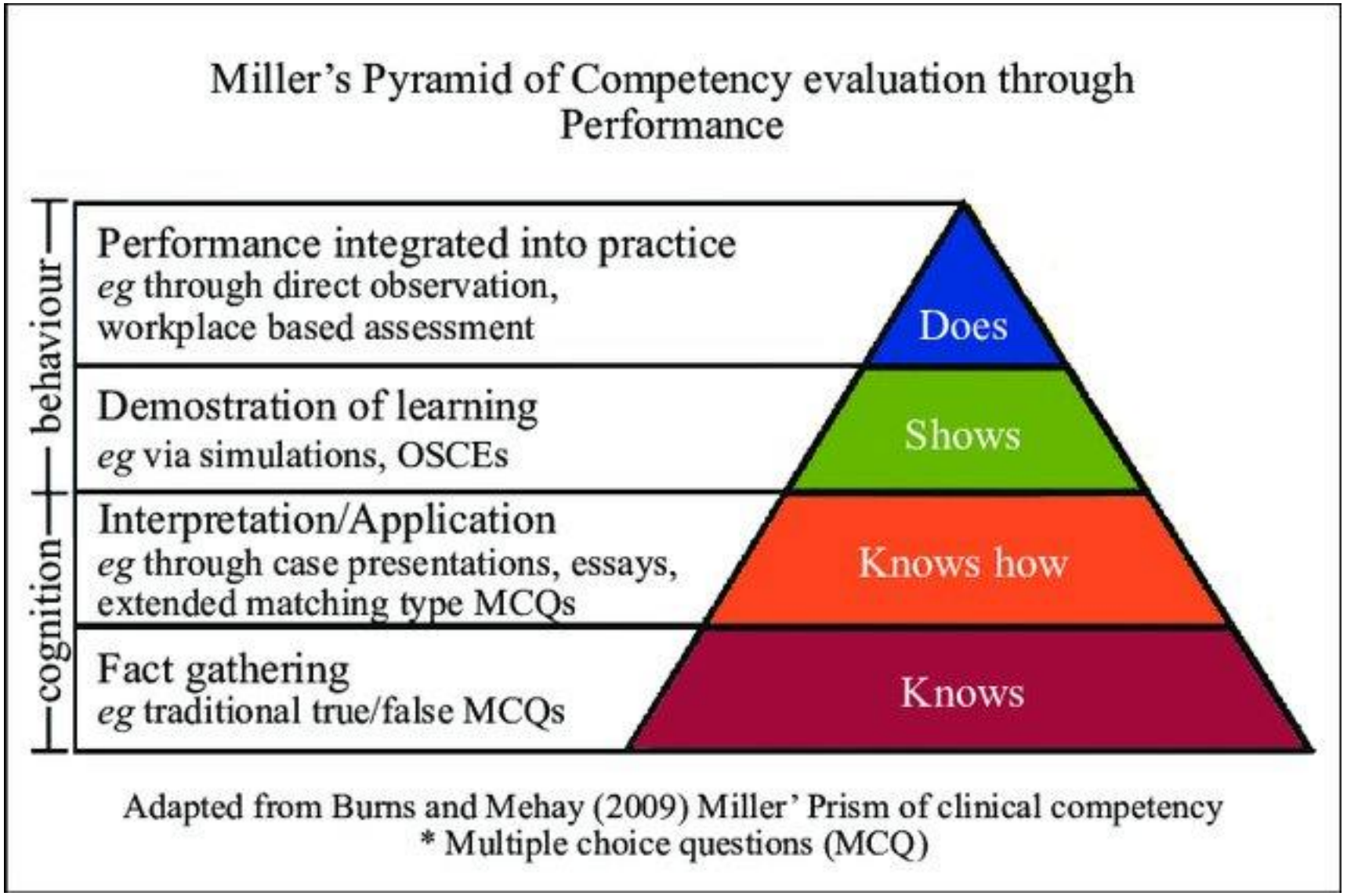
Assoc. Prof. Dr. Ahmad Marzuki Omar
Department of Internal Medicine
Kulliyyah of Medicine

Workshop on Examiners' Calibration for OSCE and OLC
Medical Education and Quality Unit
Kulliyyah of Medicine

8 September 2025

Introduction

Introduction



Introduction

Assessment focus	Proposed methods of assessment
Readiness to cope with challenges in practice	Same as does + Entrustment-based assessments (e.g., discussion)
Performance integrated into workplace	360^0 assessment, Case-based discussions, Clinical competency assessments, DOPS, Multi-source feedback, Portfolios, Work place-based assessment
Performance in controlled settings	OSCEs, Practicals, Simulations, Standardized clients / patients
Application and Manipulation of knowledge, Relationships between concepts and principles	Case presentations, Essays, Gaming, Extended matching MCQs, Problem-solving approaches
Fact gathering, Processes, Scientific principles	Essays, Oral exams, Reports, Traditional MCQs, Various tests

DOES: Evaluating cognitive, psychomotor, and affective domains in an integrated setting.

Expectations

What Are Expected From the Candidates

- Take focused history and present relevant positive and negative findings from the history
- Recognise, demonstrate and present physical findings
- Formulate a diagnosis, identify problems and offer possible differential diagnoses based on logical reasoning
- Request and interpret relevant laboratory and other investigations
- Outline appropriate management of the patient
- Demonstrate good rapport and interaction with patient

What Are Expected From the Examiners

Attend the briefing of the examination

Review and calibrate the cases

Familiarise with the marking format and scoring rubric

To observe the candidate clerking and examination for 20-25 minutes, the next 20-25 minutes for the discussion

Candidates are not required to present the case in full, may present the summary of the case

Total time of examination per case: 45 minutes

Examiners’ Calibration

Examiners' Calibration

Purpose

- Reduces variability in scoring across different examiners
- Ensures fairness in student assessment
- Enhance the validity and reliability of clinical performance evaluations
- Reduce rater bias, such as leniency, severity, and halo effects

Examiners’ Calibration

History

- If patients have multiple problems, determine which problem is appropriate for the candidate’s level to be assessed and prioritise that problem
- Due to time constraints, the examiner should focus on evaluating specific clinical problems rather than completing a checklist of history-taking
- Before examination, examiners should take the patient's history to verify the information and assess whether the patient is a reliable historian and can consistently provide an accurate history

Examiners' Calibration

Physical Examination

- Examiners need to examine the patients before the start of the examination
- Students should not be penalised if clinical findings cannot be readily elicited by the assessors themselves

Examiners' Calibration

Diagnosis

- Assess the candidate's ability to make a provisional diagnosis and consider possible differential diagnoses based on the case scenario

Examiners' Calibration

Investigation

- Determine the essential investigations required for diagnosing and managing the case
- Candidates should be able to select appropriate investigations and interpret the results correctly
- Provide normal reference values for interpretation if necessary

Examiners' Calibration

Management

- Assess the candidate's understanding of the principles of case management

Examiners' Responsibilities & Professionalism

Examiners' Responsibilities & Professionalism

Engage in rater training

- To understand performance standard and recognise biases

Adopt a criterion-referenced approach

- Evaluate student performance against clearly defined standards rather than comparing students to each other

Stay objective

- Assess students based on observed behaviours and competencies rather than personal impressions and emotions

Examiners' Responsibilities & Professionalism

Ensure consistency and integrity

- Follow standardised guidelines and reference criteria to ensure fairness and reliability in scoring

Adhere to institutional criteria rather than personal expectations

- Keep personal biases separate from professional evaluations

Use the entire rating scale properly

- Apply all levels of the scale appropriately rather than overusing middle or extreme ratings

Examiners' Responsibilities & Professionalism

Consider all aspects of performance

- Evaluate each component of the performance rather than focusing on just one trait

Use "Not applicable" when necessary

- If a performance aspect is not observed, select "not observed" or "unable to judge" instead of making assumptions

Provide justifiable scores

- Be prepared to explain and defend ratings based on clear performance criteria

Examiners' Responsibilities & Professionalism

Ensure responsible grade reporting

- Consider the validity of assessments and weight grades accordingly for accurate performance representation

Recognise and mitigate rater bias

- Be aware of potential biases such as leniency, severity, or halo effects

Remain cognizant of fatigue effects

- Do not let fatigue affect your scoring

Rating & Scoring

Clinical Rating Form

INTERNATIONAL MULTI-AWARD WINNING INSTITUTION FOR SUSTAINABILITY

Student's sticker

BACHELOR OF MEDICINE & BACHELOR OF SURGERY(MBBS)
KULLIYYAH OF MEDICINE
SESSION 2023/2024

CLINICAL RATING FORM

OBSERVED LONG CASE

FINAL PROFESSIONAL EXAMINATION

Matric No.:	Discipline: MED | PAED | PSY | FAMMED | SUR |O&G | ORTHO
Case Diagnosis:	
Examiners:	Room:

Instruction to the examiner: Circle the appropriate score

Assessment Component	Weightage (A)		Score and Status (B)											Mark (B/10 X A)
	Standard	Alternative	Bad Fail		Fail		Borderline	Satisfactory		Good		Excellent		
History	20		0	1	2	3	4	5	6	7	8	9	10	
Physical / Mental State Examination	20		0	1	2	3	4	5	6	7	8	9	10	
Diagnostic Ability	15		0	1	2	3	4	5	6	7	8	9	10	
Investigation	15		0	1	2	3	4	5	6	7	8	9	10	
Patient Management	20		0	1	2	3	4	5	6	7	8	9	10	
Professionalism	10		0	1	2	3	4	5	6	7	8	9	10	
													TOTAL	

REMARKS

.....

.....

.....

SIGNATURE

.....

Examiner's name:

Date:

GLOBAL RATING	
	FAIL
	BORDERLINE
	PASS

Scoring Rubric

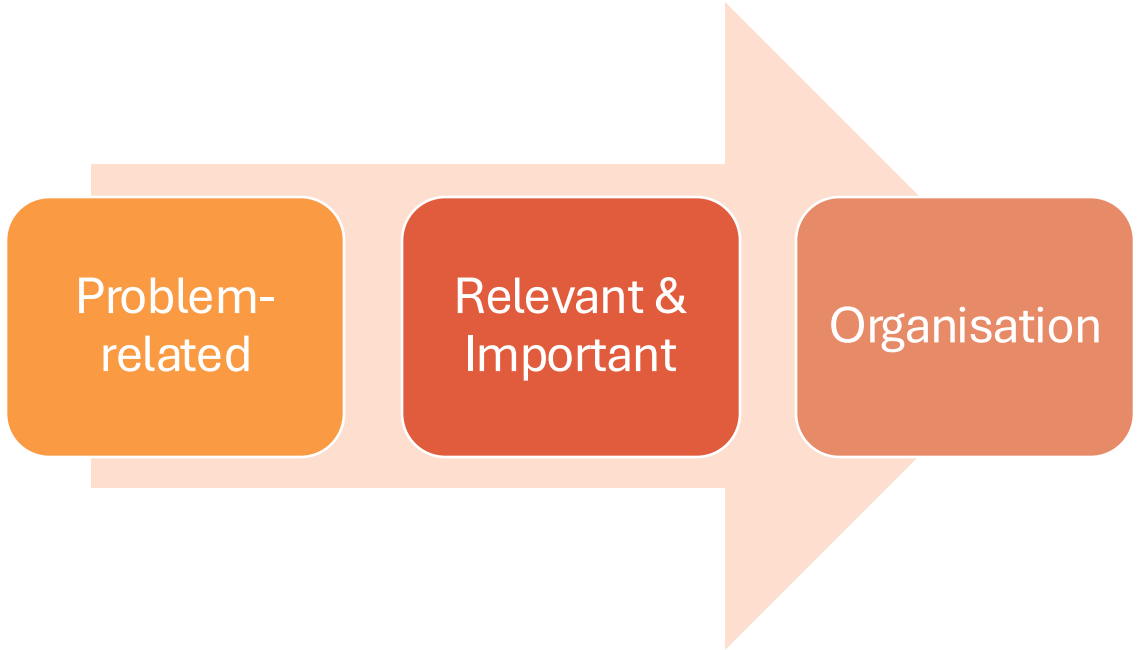
NO.	CLINICAL COMPONENT	STATUS	SCORE
1	HISTORY		
	Elicits problem related data, stresses important points, has a well-organised approach.	Excellent	9 -10
	As above but not always well-organised.	Good	7 - 8
	As above but misses a few relevant information.	Satisfactory	5 - 6
	As above but concentrates on data not relevant to problem, misses important information.	Borderline	4
	Approach not well-organised, not problem-related, misses many important information.	Fail	2 - 3
	Irrelevant history or no attempt to answer at all despite prompting.	Bad fail	0 - 1
2	PHYSICAL / MENTAL STATE EXAMINATION		
	Elicits and interprets correctly all signs, systematic with good technical and organisational approach.	Excellent	9 -10
	As above with satisfactory organisational approach but less systematic.	Good	7 - 8
	As above but misses a few relevant physical / mental state signs with lack organisational approach.	Satisfactory	5 - 6
	As above but some technical and organisational imperfection, some important data missed, invents signs.	Borderline	4
	Approach technically and organisationally imperfect or unacceptable, important data missed, invents signs.	Fail	2 - 3
	Irrelevant clinical examination or no attempt to answer at all despite prompting.	Bad fail	0 - 1
3	DIAGNOSTIC ABILITY		
	Makes reasoned deduction from available data, able to give correct provisional and differential diagnoses.	Excellent	9 -10
	As above, able to give provisional diagnosis but lack comprehensive differential diagnoses.	Good	7 - 8
	As above but shows minor faulty deductions, able to give correct provisional but not all relevant differential diagnoses.	Satisfactory	5 - 6
	Makes major faulty deductions from available data, give wrong provisional diagnosis, but able to give some differential diagnoses.	Borderline	4
	Does not follow logical approach to deduction from data (haphazard), faulty deduction, give wrong provisional diagnosis and differential diagnoses.	Fail	2 - 3
	No attempt to answer at all despite prompting.	Bad fail	0 - 1

Scoring Rubric

4	INVESTIGATION		
	Plans/requests and interprets investigations appropriate to the problem with attention to specificity, reliability, patient safety and comfort, and cost. Able to explain reasons for and nature of investigations.	Excellent	9 -10
	As above, able to explain reasons and nature of investigations but less comprehensive.	Good	7 - 8
	As above but misses a few relevant investigations.	Satisfactory	5 - 6
	Misses a few relevant investigations and suggests investigations not appropriate to the problem.	Borderline	4
	Makes inappropriate decision in ordering investigations, misinterprets data.	Fail	2 - 3
	No attempt to answer at all despite prompting.	Bad fail	0 - 1
5	PATIENT MANAGEMENT		
	Suggests appropriate comprehensive management, exhibits awareness of the role and possible complications of the proposed intervention.	Excellent	9 -10
	As above but less comprehensive management.	Good	7 - 8
	As above but misses a few relevant points in the comprehensive management suggested.	Satisfactory	5 - 6
	Misses many important points about comprehensive management.	Borderline	4
	Suggests inappropriate management, show lack of awareness of role of proposed interventions and their possible complications.	Fail	2 - 3
	No attempt to answer at all despite prompting.	Bad fail	0 - 1
6	PROFESSIONALISM		
	Excellent rapport with the patient, able to gain the patient's trust and confidence.	Excellent	9 -10
	Good rapport with the patient.	Good	7 - 8
	Acceptable rapport with the patient.	Satisfactory	5 - 6
	Not able to establish rapport at all with the patient.	Borderline	4
	Rough and inconsiderate to the patient.	Fail	2 - 3
	Obviously upsets the patient by being inconsiderate.	Bad fail	0 - 1

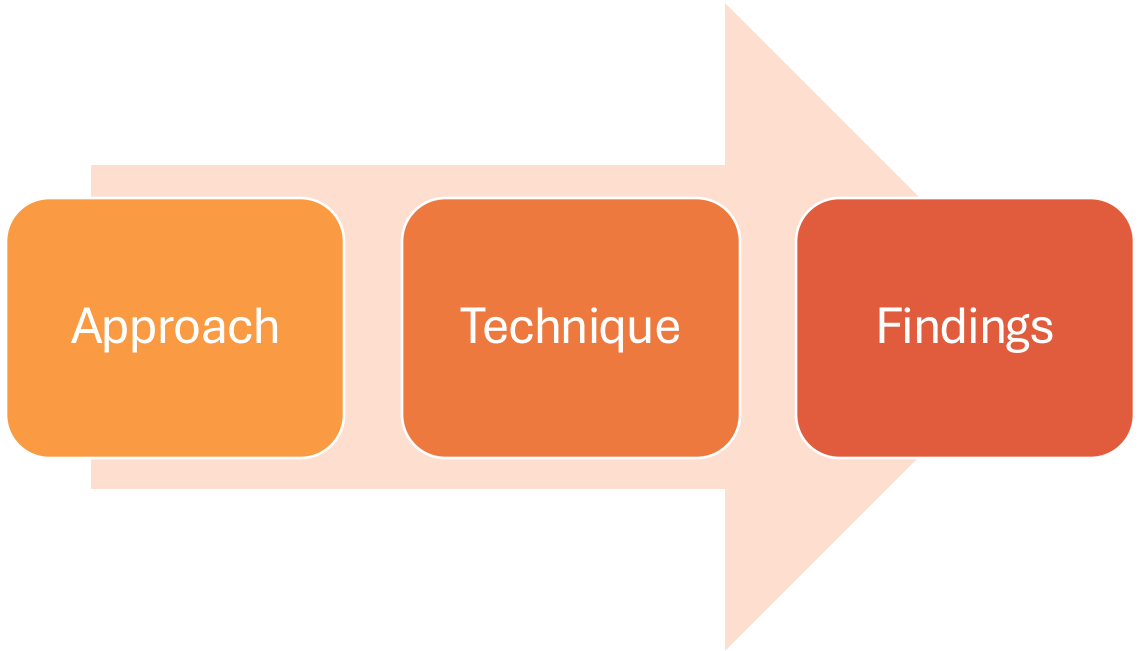
History

1	HISTORY		
	Elicits problem related data, stresses important points, has a well-organised approach.	Excellent	9 -10
	As above but not always well-organised.	Good	7 - 8
	As above but misses a few relevant information.	Satisfactory	5 - 6
	As above but concentrates on data not relevant to problem, misses important information.	Borderline	4
	Approach not well-organised, not problem-related, misses many important information.	Fail	2 - 3
	Irrelevant history or no attempt to answer at all despite prompting.	Bad fail	0 - 1



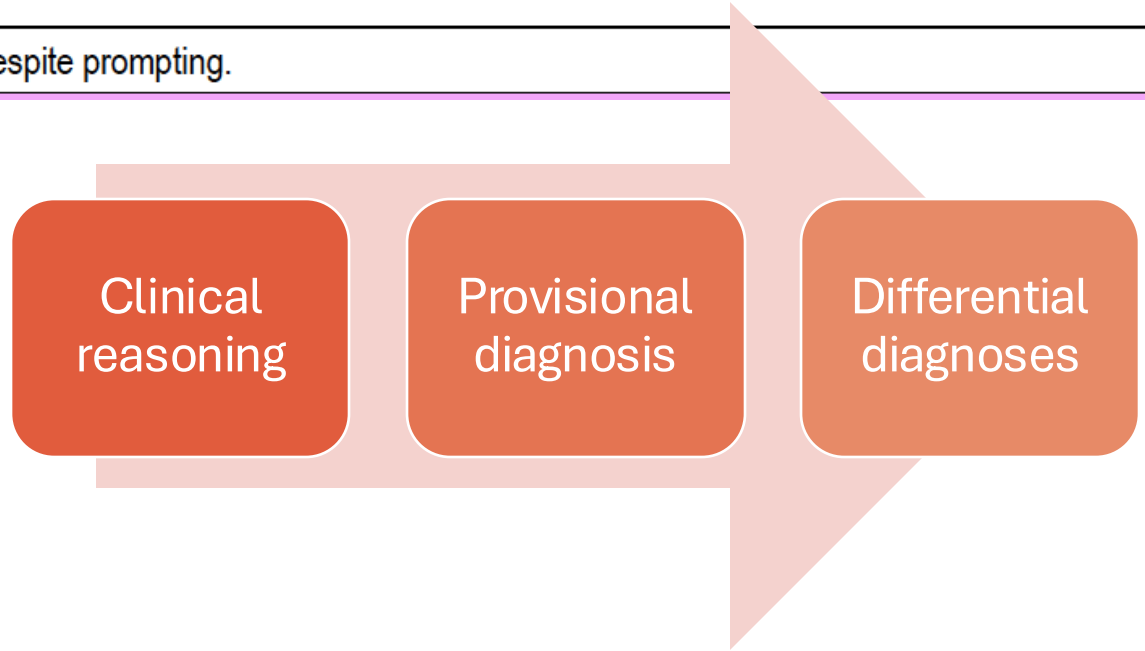
Physical / Mental State Examination

2	PHYSICAL / MENTAL STATE EXAMINATION		
	Elicits and interprets correctly all signs, systematic with good technical and organisational approach.	Excellent	9 -10
	As above with satisfactory organisational approach but less systematic.	Good	7 - 8
	As above but misses a few relevant physical / mental state signs with lack organisational approach.	Satisfactory	5 - 6
	As above but some technical and organisational imperfection, some important data missed, invents signs.	Borderline	4
	Approach technically and organisationally imperfect or unacceptable, important data missed, invents signs.	Fail	2 - 3
	Irrelevant clinical examination or no attempt to answer at all despite prompting.	Bad fail	0 - 1



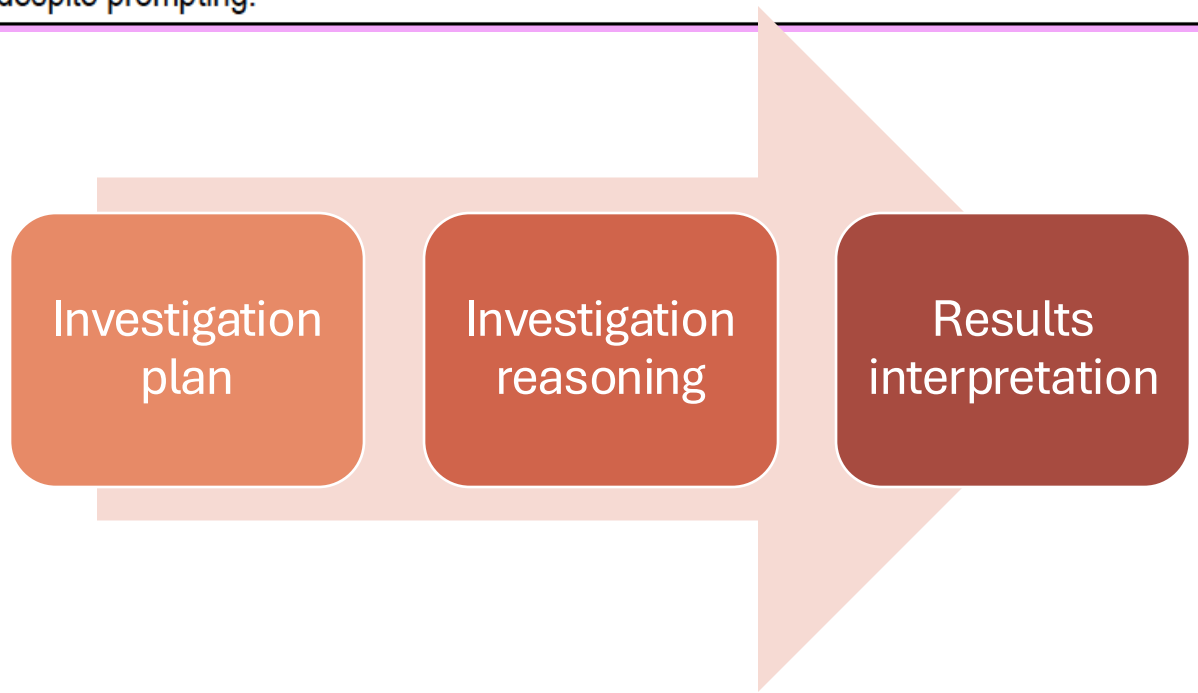
Diagnostic Ability

3	DIAGNOSTIC ABILITY		
	Makes reasoned deduction from available data, able to give correct provisional and differential diagnoses.	Excellent	9 -10
	As above, able to give provisional diagnosis but lack comprehensive differential diagnoses.	Good	7 - 8
	As above but shows minor faulty deductions, able to give correct provisional but not all relevant differential diagnoses.	Satisfactory	5 - 6
	Makes major faulty deductions from available data, give wrong provisional diagnosis, but able to give some differential diagnoses.	Borderline	4
	Does not follow logical approach to deduction from data (haphazard), faulty deduction, give wrong provisional diagnosis and differential diagnoses.	Fail	2 - 3
	No attempt to answer at all despite prompting.	Bad fail	0 - 1



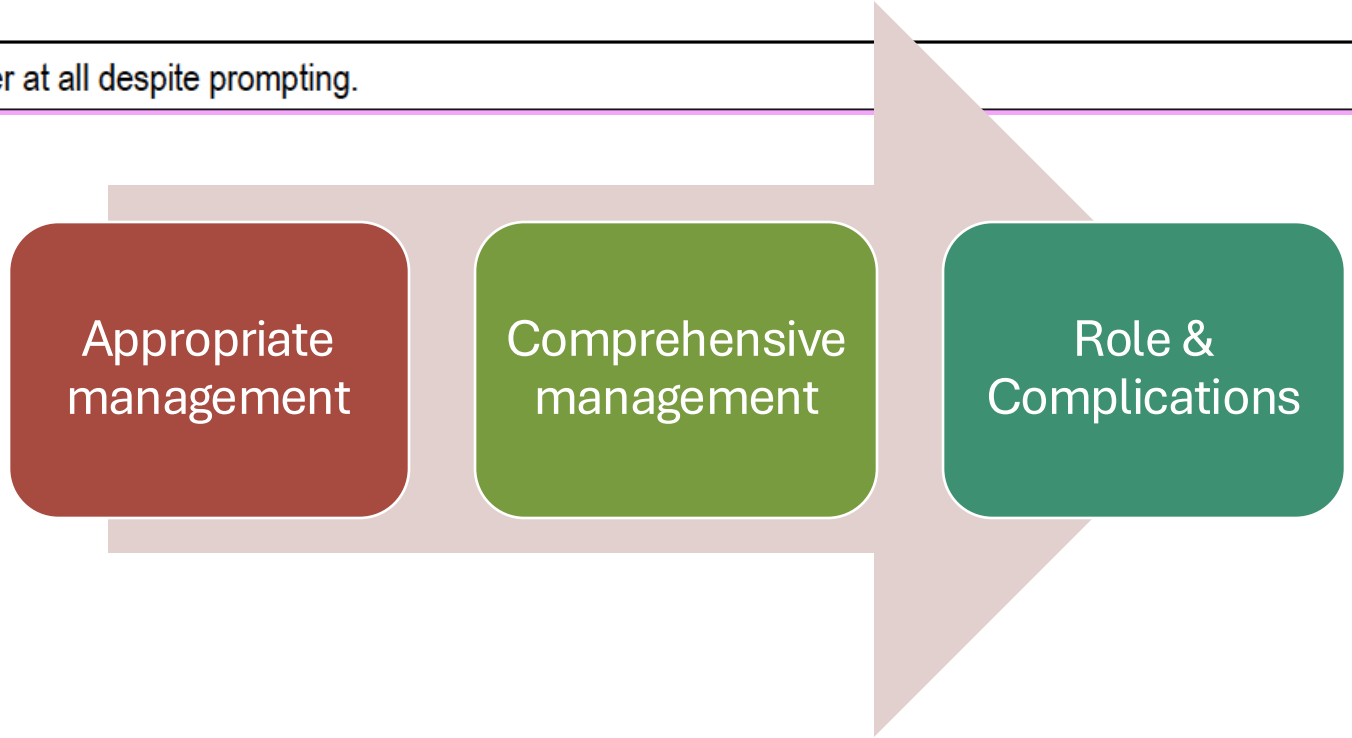
Investigation

4	INVESTIGATION		
	Plans/requests and interprets investigations appropriate to the problem with attention to specificity, reliability, patient safety and comfort, and cost. Able to explain reasons for and nature of investigations.	Excellent	9 -10
	As above, able to explain reasons and nature of investigations but less comprehensive.	Good	7 - 8
	As above but misses a few relevant investigations.	Satisfactory	5 - 6
	Misses a few relevant investigations and suggests investigations not appropriate to the problem.	Borderline	4
	Makes inappropriate decision in ordering investigations, misinterprets data.	Fail	2 - 3
	No attempt to answer at all despite prompting.	Bad fail	0 - 1



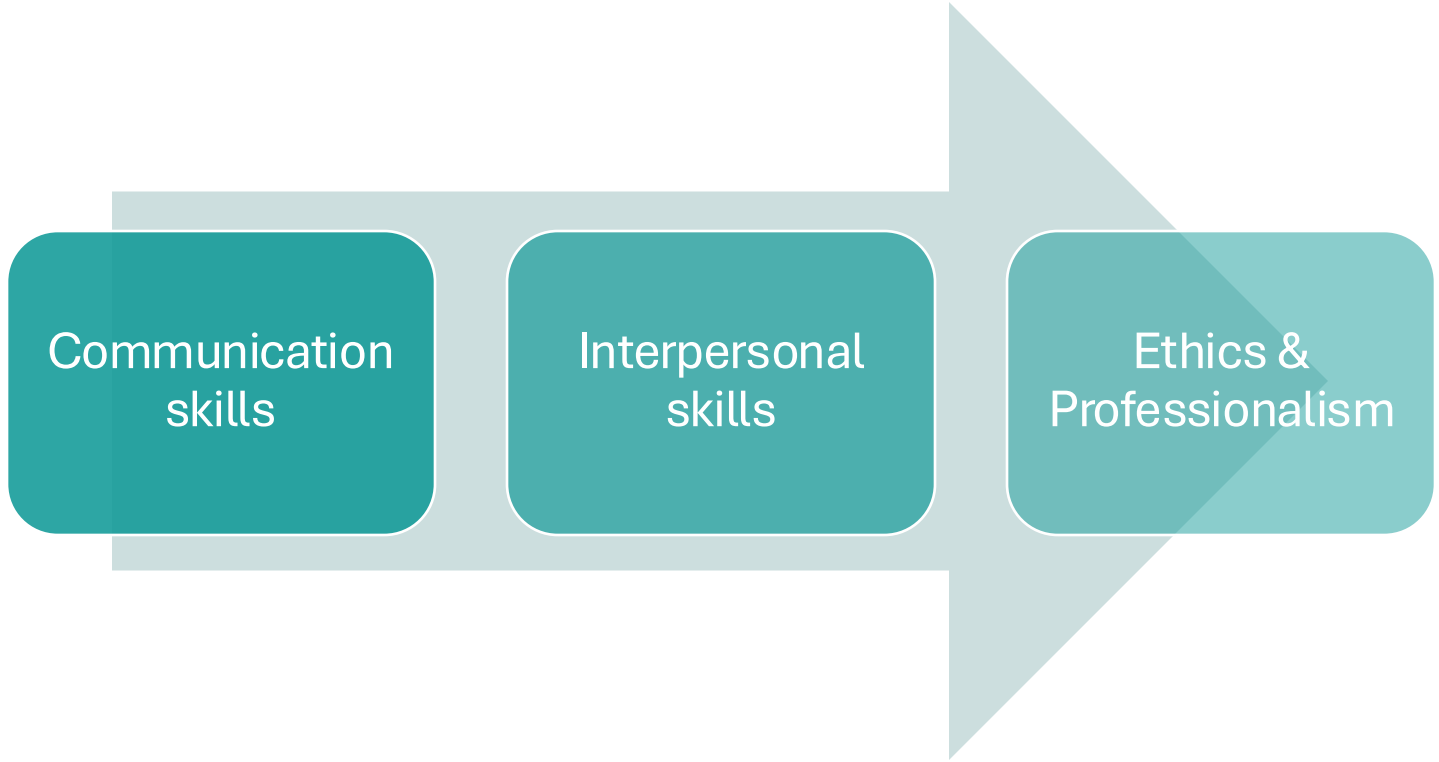
Patient Management

5	PATIENT MANAGEMENT		
	Suggests appropriate comprehensive management, exhibits awareness of the role and possible complications of the proposed intervention.	Excellent	9 -10
	As above but less comprehensive management.	Good	7 - 8
	As above but misses a few relevant points in the comprehensive management suggested.	Satisfactory	5 - 6
	Misses many important points about comprehensive management.	Borderline	4
	Suggests inappropriate management, show lack of awareness of role of proposed interventions and their possible complications.	Fail	2 - 3
	No attempt to answer at all despite prompting.	Bad fail	0 - 1



Professionalism

6	PROFESSIONALISM		
	Excellent rapport with the patient, able to gain the patient's trust and confidence.	Excellent	9 -10
	Good rapport with the patient.	Good	7 - 8
	Acceptable rapport with the patient.	Satisfactory	5 - 6
	Not able to establish rapport at all with the patient.	Borderline	4
	Rough and inconsiderate to the patient.	Fail	2 - 3
	Obviously upsets the patient by being inconsiderate.	Bad fail	0 - 1



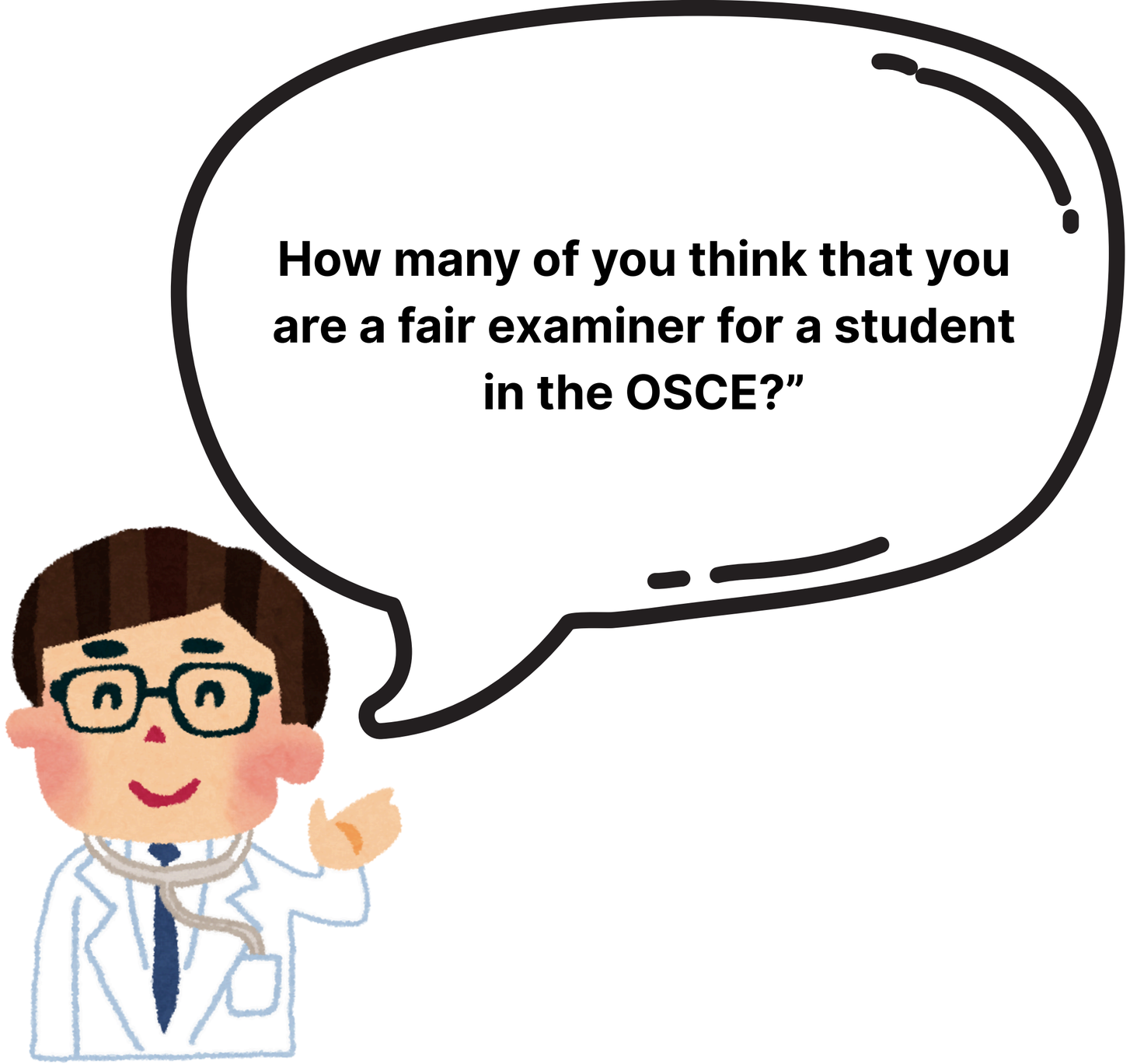
Let’s View the Video..

Thank You

OSCE CALIBRATION

Assoc Prof Dr Nurjasmine Aida Jamani

Department of Family Medicine, KOM, IIUM/ DEAR SASMEC@IIUM


How many of you think that you are a fair examiner for a student in the OSCE?”

WHAT IS CALIBRATION?

Process of training and aligning examiners to ensure consistent, accurate, and fair assessment of candidates' clinical performance.

- Shared understanding among examiners
- Consistency in scoring and judgement
- Standardization of expectations

OBJECTIVE

01

Build clear understanding in the use of the marking rubric

Be clear on the grid marking scales

02

Improve inter-rater reliability between examiners

Assessment goals, judging performance, sources of judgment error

03

Discuss performance standards

What is clear pass/ borderline/ fail?

Why calibration matters?

- Ensures fairness to students
- Enhances reliability of assessment
- Builds confidence in exam outcomes
- Protects institutional credibility

WHY IT MATTERS?

BENEFITS

Fairness

Students feel assessed fairly

Consistent

Faculty gain confidence and consistency

Valid

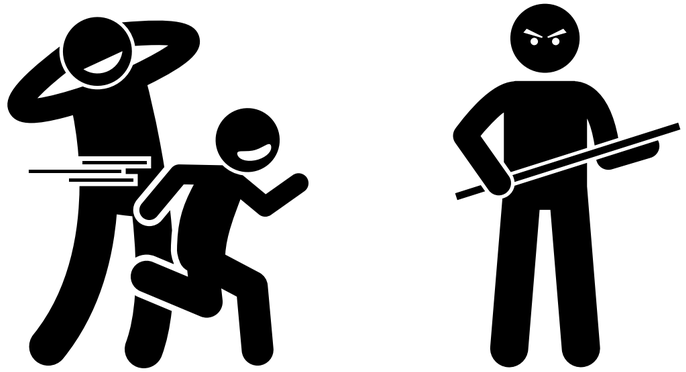
Institution ensures high-stakes exam validity

Accreditation

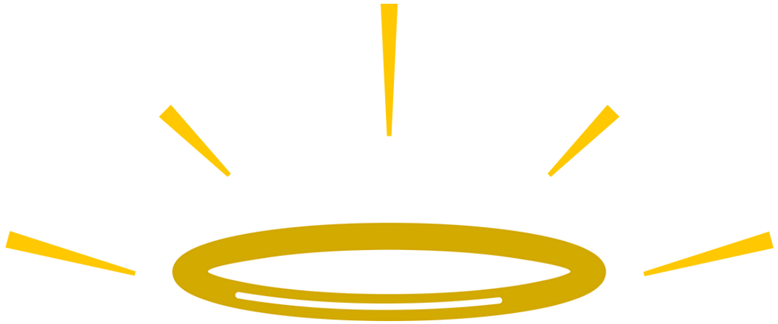
Better preparation for future accreditation visits

COMMON CHALLENGES IN OSCE

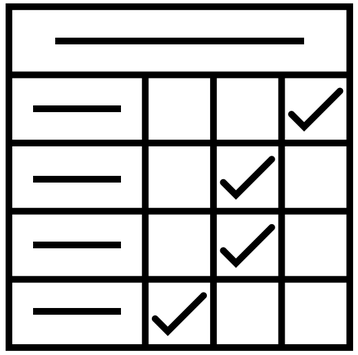
Leniency vs Stringency



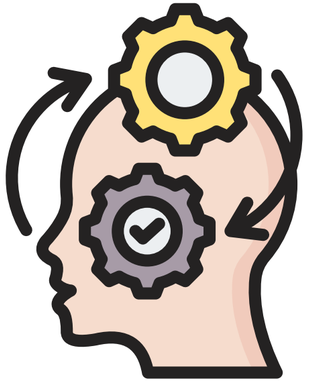
Halo effect



Unclear marking rubrics



Variability in student performance interpretation



ROLE OF EXAMINERS

01 Be objective, not subjective

02 Follow rubric faithfully

03 Observe carefully, avoid assumptions

04 Provide consistent scoring

OSCE

Scenario:

A 24-year-old male was admitted with a close right tibial plateau fracture following an MVA. Around 12 hours post admission the staff nurse in the ward called you, the house officer, to review the patient due to complains of severe pain

Task:

The student is required to do a quick and proper assessment to obtain a diagnosis and manage/treat the problem accordingly.

2. The objective of this station is to evaluate/assess the ability of the student to quickly identify the emergent problem of compartment syndrome through relevant and focused history and physical examination. The ability of the student to apply the immediate management of the problem should also be assessed

3. This is a 10-minute station.

4. Please grade the student for every item listed in the CHECK LIST. Do not leave any of the items unmarked.

5. Score the GLOBAL RATING independently of the numeric score (for medical education unit used)

CHECK LIST MATRIC NUMBER: _____

No	Task Demonstration/ Questions	Not done/ wrongly done	Partially done	Inadequately done	Adequately done	Well done
1.	Introduction: - Introduce self - Explain the purpose of interview - Ensure confidentiality - Gather sociodemographic data	0				1
2.	Duration of symptoms	0				1
3.	History of present illness: -Core: Irritability, elated mood, increase energy, increase in goal directed activities -Grandiosity/ Grandiose delusion -Distractibility -Reduce need for	0	1	2	3	4

GLOBAL RATING

An examiner's overall, holistic assessment of a candidate's performance at a station, typically on a scale from "fail" to "excellent"

> **Global rating scales administered by experts are a more appropriate summative measure when assessing candidates on performance-based examinations**

Regehr G, MacRae H, Reznick RK & Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Academic Medicine 1998; 73(9): 993-997.

GLOBAL RATING

Irrespective of the marks scored in the checklist, please assess the student using the global rating scale below. Please √ in the appropriate box.

	FAIL
	BORDERLINE
	PASS

BORDERLINE STATEMENT FOR MBBS PROGRAM

"The borderline passing graduate of MBBS IIUM should demonstrate adequate fundamental knowledge, safe clinical judgement and decision-making ability, able to work with supervision, has effective communication and upholding professionalism and ethical values incorporating Islamic values."

MEDICAL EDUCATION & QUALITY UNIT

Kulliyyah of MEDICINE كلية الطب

OSCE MARKING EXERCISE

FAMILIARISE THE RUBRIC

Look and understand the objective of the station, the task that students need to know and the rubric marking

WATCH THE VIDEO

Observe the performance of the students in the video

PUT IN YOUR MARKS

Enter your marks in the google sheet and grade the performance of the students

Discussion

TAKE HOME MESSAGE

Calibration = fairness + reliability

Examiners must check bias, stick to rubrics, and practice together

Regular calibration keeps OSCEs trustworthy

“If students prepare for the exam, we must also prepare for our role as examiners.

THANK YOU