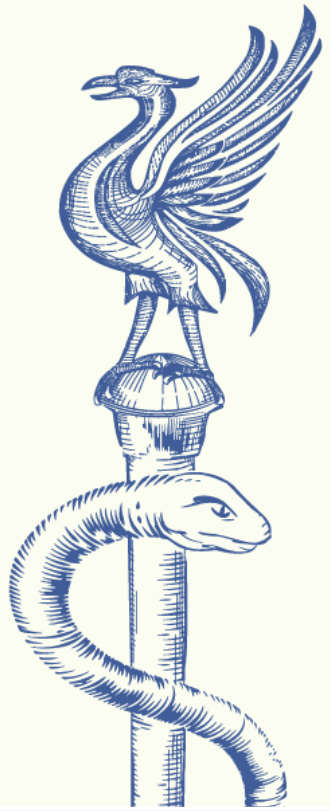
UNIVERSITY OF LIVERPOOL

Assessment the basics

For an undergraduate medical curricula

Task to start with!

- Send you to breakout rooms (5 different rooms)
- Each room will have a different area to think about
- Question is “What is the purpose of assessment?”
- I will give you 5 – 10 mins to discuss the question



What is the purpose of assessment

- Students
- School of Medicine
- University
- GMC
- General population

UNIVERSITY OF LIVERPOOL

What will we cover today

1. What is the purpose of assessment
2. How does assessment align to Liverpool MBChB and beyond
3. How is assessment quality assured
4. Medical Licensing Assessment

Purpose of assessment

- The first question is why do we assess?
- How and when should we assess?
- How do we know the assessment is working?

Assessment strategy

- Why are we testing, and can we evidence this?
  - Is there an assessment strategy with a clear purpose
  - Is it feasible and deliverable (can it be resourced)

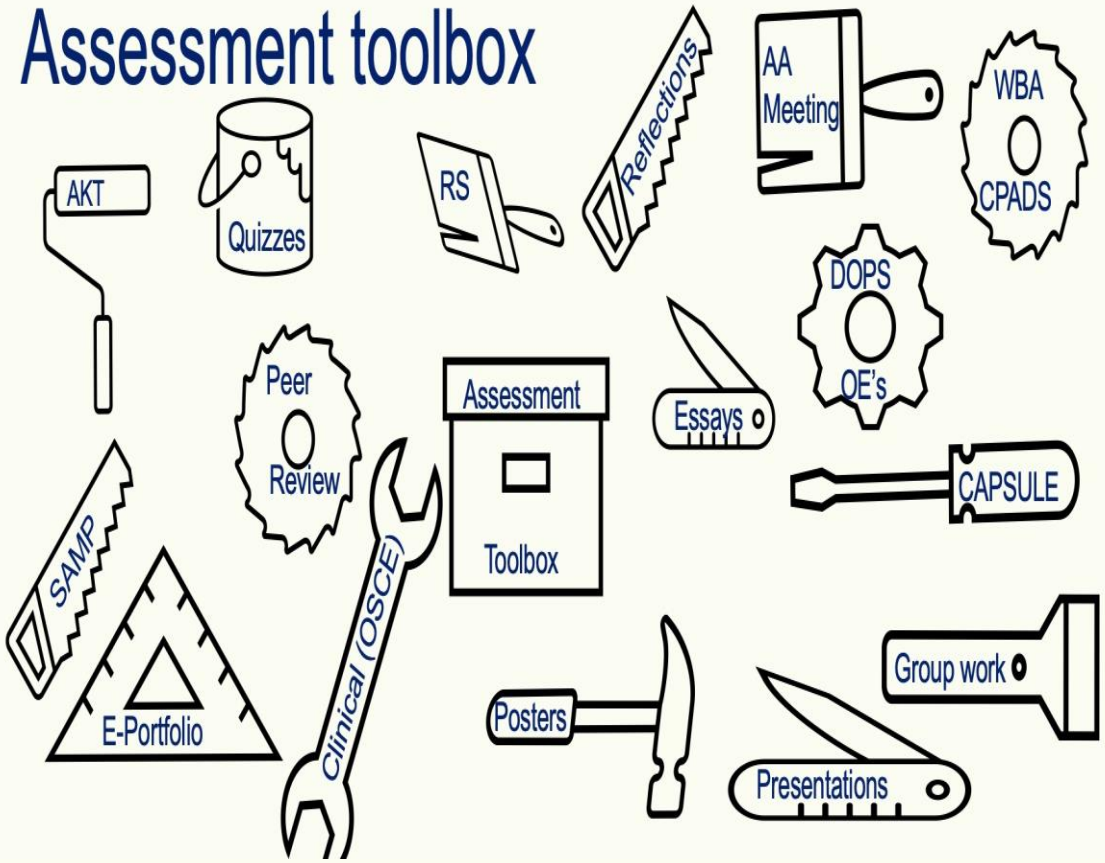
Why we assess

- Students want to show what they have learned
- Tutors want evidence that that students have understood “my” part of the course
- The school needs to demonstrate to the university and GMC that the students have been taught and learned the curriculum

Assessment drives learning !!!

How do we assess

- This is back to feasible and deliverable
- Assessment tools
  - Feasibility – ease to produce, ease of marking, relevance, cost.
- Assessing softer skills
  - professionalism
- Not just end of year assessments



UNIVERSITY OF LIVERPOOL

Assessment development

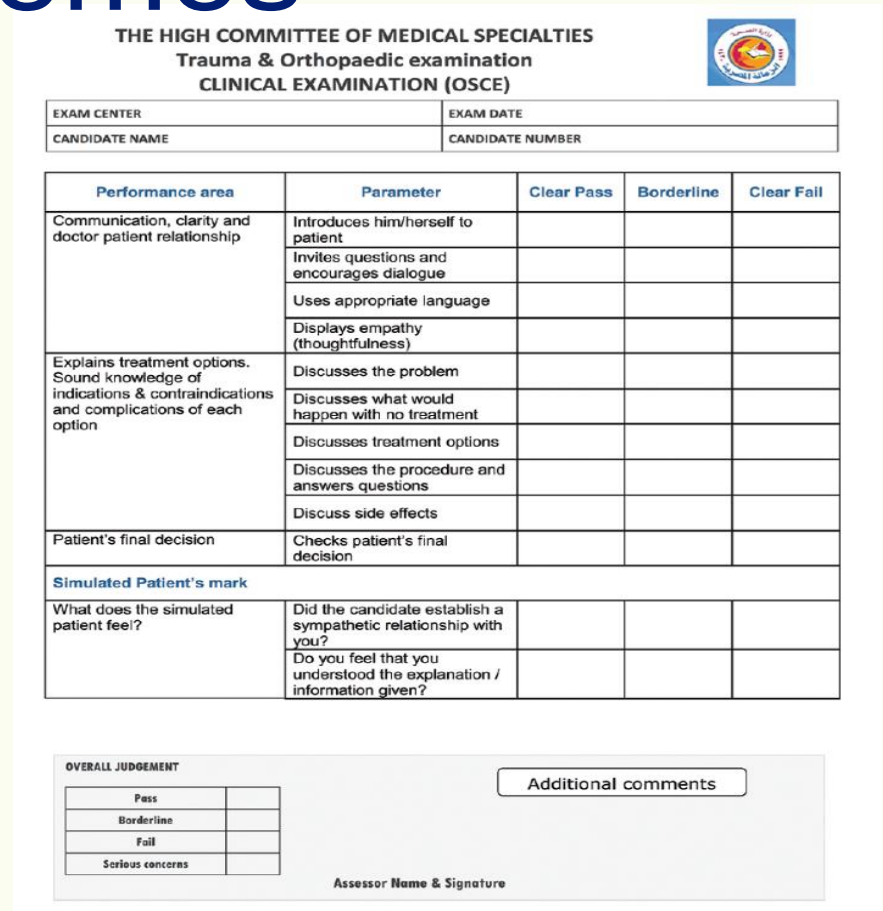
Computer based marking

- Well this makes marking easier
- Sustainability – no paper but using PC's
- Currently using SBA's often 'knowledge' only bias
- National exams are all computer based

- Can we deliver better tests using new formats of computer based marking?

Performance marking schemes

- Does this drive formulaic behaviours?
  - The OSCE way
- What about clinical variation
- Authentic clinical assessment
  - In reality is this possible



THE HIGH COMMITTEE OF MEDICAL SPECIALTIES
Trauma & Orthopaedic examination
CLINICAL EXAMINATION (OSCE)

EXAM CENTER	EXAM DATE
CANDIDATE NAME	CANDIDATE NUMBER

Performance area	Parameter	Clear Pass	Borderline	Clear Fail
Communication, clarity and doctor patient relationship	Introduces him/herself to patient			
	Invites questions and encourages dialogue			
	Uses appropriate language			
	Displays empathy (thoughtfulness)			
Explains treatment options. Sound knowledge of indications & contraindications and complications of each option	Discusses the problem			
	Discusses what would happen with no treatment			
	Discusses treatment options			
	Discusses the procedure and answers questions			
	Discuss side effects			
Patient's final decision	Checks patient's final decision			
Simulated Patient's mark				
What does the simulated patient feel?	Did the candidate establish a sympathetic relationship with you?			
	Do you feel that you understood the explanation / information given?			

OVERALL JUDGEMENT

Pass	
Borderline	
Fail	
Serious concerns	

Additional comments

Assessor Name & Signature

What is the current reality

- 20 end of block tests
- 8 AKTs
- 5 projects
- End of placement evaluations
- Endless WBA (minimum criteria set)
- 6 clinical assessments
- Presentations
- Posters

Why assess?..... Achievement

- For Students

  As a diagnostic indicator of development and progress.

- For the University

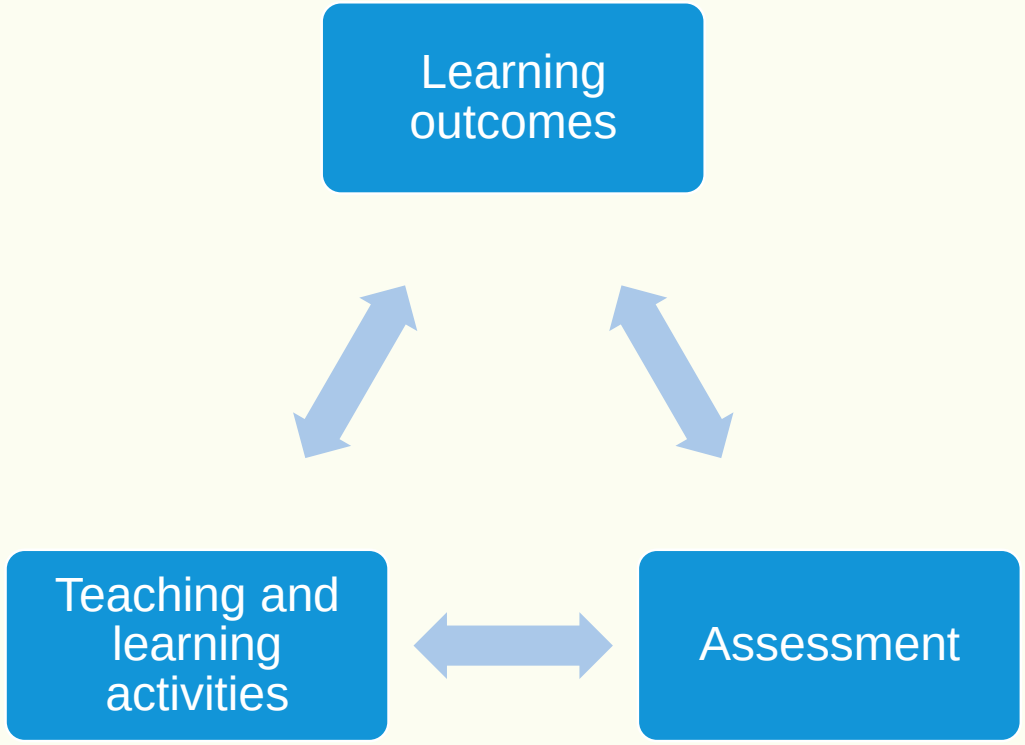
  As an indicator of the quality of curriculum, faculty and students.

- For regulators (GMC) and the public

  As an assurance of competency and fitness to practice?

Coherent and comprehensive

- Constructive alignment
- Developed with the curriculum learning and teaching
- Selection of the right tool depending on the purpose of assessment, consistent with the outcomes.



Briggs & Tang 2003, pg53

Blueprinting

When you have selected an assessment tool the content must align with the learning outcomes.

Sampling across the curriculum including domains and themes

Using lectures, small group teaching (CCP, CBL), clinical skills teaching, clinical placements, HARC the list goes on.

Blueprinting

- We can not test everything
- We need to sample wide enough across the curriculum

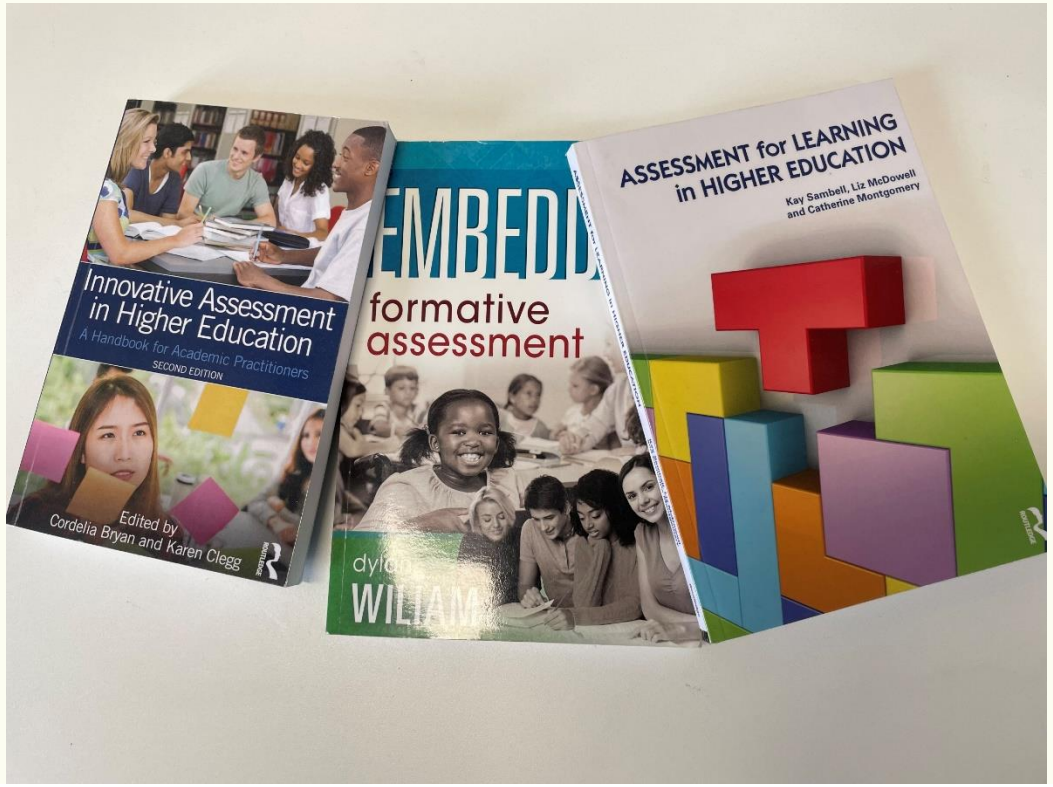
- In year blueprint (vertical) and across the years (horizontal)

- Check the blueprint against the national regulators requirements

Information gathering (History)	Information sharing (explanation)	Integration of data, investigations, risk factors	Physical Examination	Medical knowledge/ Diagnosis / Problem Lists	Planning & Investigations	Management & Shared Decision Making	Risk reduction / health promotion	Recognising, reducing & dealing with error	Ethico-legal aspects of care	Probity & Handling Complaints	Working effectively with colleagues (including senior escalation)	Dealing with uncertainty	Practical Clinical Skills	Record Keeping	Prescribing	Safety	Infection Prevention & Control	Readiness for safe practice	Managing uncertainty	Delivering person-centred care	Demonstrate that candidates can identify and interpret clinical findings.
X	X										X					X					
X				X												X	X	X	X	X	X
	X		X	X													X	X	X	X	X
								x	x				x			x	x	X	X	X	X
	X		X	X									X				X	X			X
																		X	X	X	X
			x	x	x	x	x		x								x	x	x	x	x	x
			x	x	x			x	x			x			x			x	x	x	x	x
			x	x	x		x		x									x	x	x	x	x
			x	x	x	x	x		x		x						x	x	x	x	x	x
								x	x				x		x			x	x	x	
			x		x		x	x	x			x		x	x	x		x	x	x	x	x
			x	x	x	x	x		x	x							x	x	x	x	x	x

Feedback

- The literature talks about feedback that is useful!!
- What does useful feedback look like?
  - Written
  - Verbal
  - Timely
- Who uses the feedback?



Assessment for learning

- Using an assessment tool to support learning
- Provide feedback
  - Not just correct or incorrect
  - Why is it the correct answer (rational)
  - Domains, systems, themes
- Support feedback with review sessions with curriculum leads
- Help students to develop a plan
  - Identification of strengths, weaknesses, reflection on personal performance

Assessment for progression

- Are students at the standard that would be considered safe to move to the next stage of their training?

- Feedback – yes
  - Identification of strengths and weaknesses
  - Reflection
  - Plan
- Passing students often don’t do this they are relieved and happy to move on.

Sequential testing model

- Both Clinical and AKT use this format
- Two parts to a single assessment
  - Sequence 1 – this is a screening assessment that the full cohort takes. The performance of students is reviewed and student who have achieved a higher threshold can be exempt from the second sequence.
  - Sequence 2 – the remaining students will take the second part of the exam. Both sequence 1 and 2 are combined and the pass/fail threshold is set.

Sequential model continued

- Why do we do this?
  - Having more questions (AKT) or stations (clinical) allows wider testing of the students for them to demonstrate their ability.
  - We are not over testing those students
  - Cost effective with expensive clinical assessments
  - Less students requiring to attend a resit attempt, but overall fail rate after resit remains similar across the different years

Sequential challenges

- Language is crucial –
  - Having to sit sequence 2 is **not** a fail or resit
  - Sequence 2 is different questions/stations for a student to demonstrate their ability
  - **Students should all prepare and expect to sit both sequences**, it is a bonus if they are exempt.
  - Both sequence 1 and 2 are in the assessment period

Homer et al 2008; The benefits of sequential testing: Improved diagnostic accuracy and better outcomes for failing students.

UNIVERSITY OF LIVERPOOL

Let’s have a break

What is standard setting?

- Setting the ‘cut score’ or pass mark for an assessment
- Two broad categories:
  - Relative or ‘norm-referenced’
    - Filling limited number of places
    - Set proportion fail
    - Pass mark is set relative to the mean
  - Absolute or criterion-referenced
    - Meet a pre-defined (fixed or variable) standard

Criterion-referenced standard setting

- Fixed standard – ‘Pass mark is always 50 %’
- Variable standard – Pass mark changes

Advantages of criterion-referencing

- Considers difficulty of the assessment
- Avoids false negatives and false positives
- Helps maintain standards
- More defensible standard

Criterion-referenced standard setting

- Relies on judgement of experts
- Test-centred
- Based around the ‘borderline’ or ‘just passing’ candidate

Applied Knowledge Test - Ebel method

- Two stages process
  - Judges consider all items individually
  - Formal meeting to review selected questions
  - Re-rate questions if needed

Initial rating

- Consider:
  - Difficulty of the question for the borderline candidate
    - Easy/Moderate/Hard?
  - Relevance of the question to the curriculum **at the level of the assessment**
    - Essential/Important/Additional?

Formal meeting

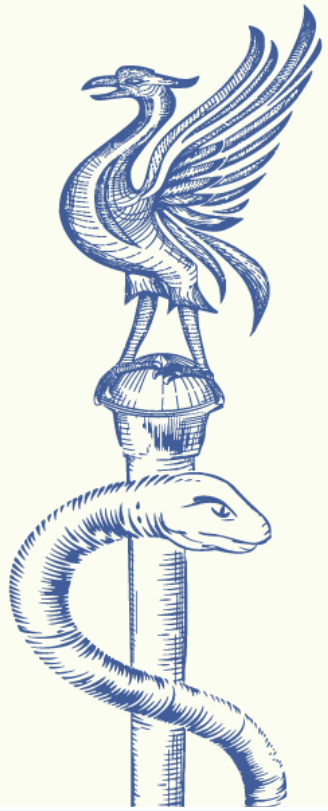
- Meet (on Zoom) and discuss individual questions as necessary
  - Where there is extreme disagreement between categories
  - Problem identified during question review
- Ask judges at extremes to state their views and rationale
- All judges invited to re-rate

- After the meeting academic lead then calculates pass mark

Clinical Assessment

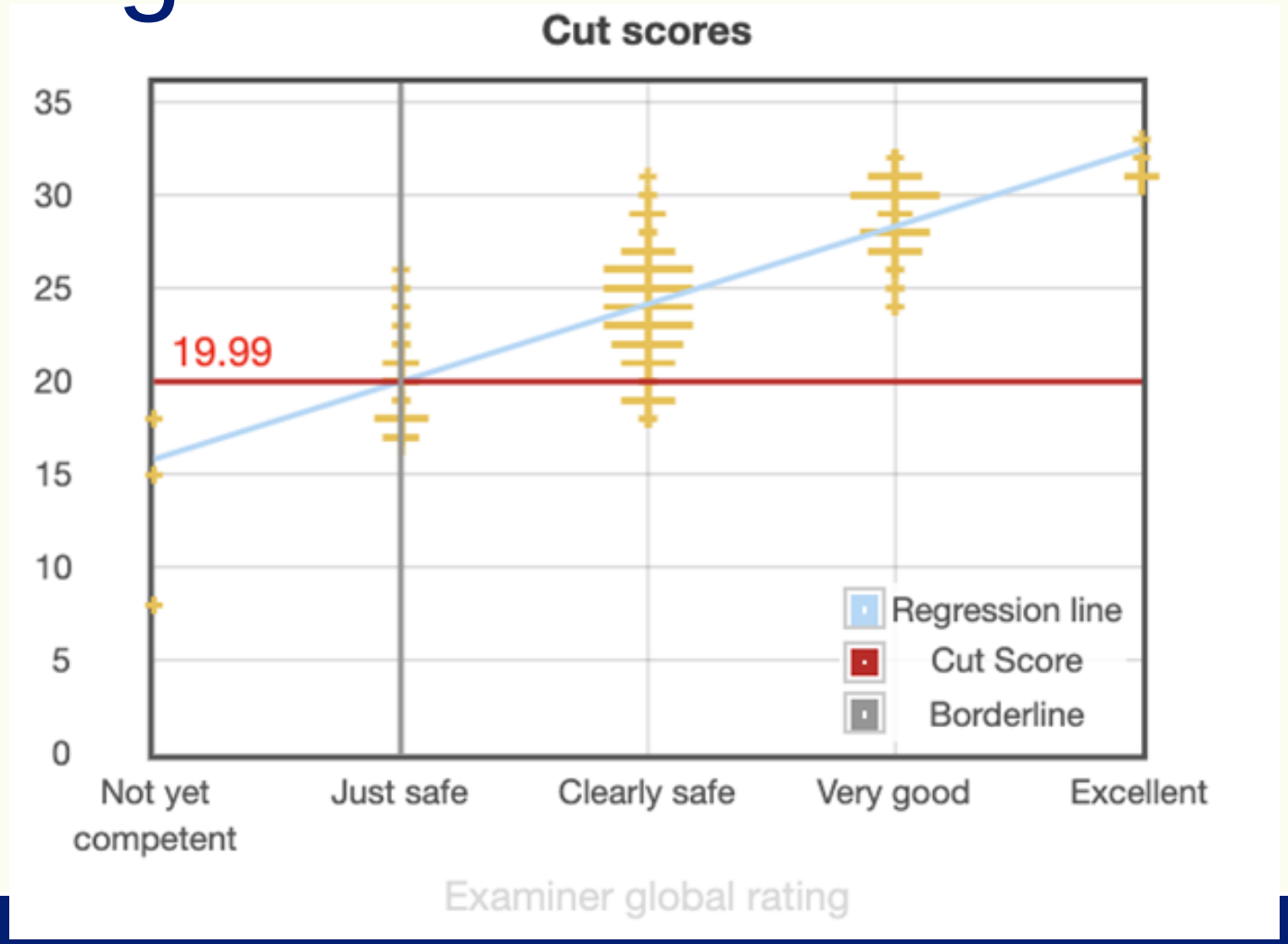
Borderline Regression Standard Setting

- Each marking criteria has a global rating
  - Not yet competent
  - Just safe
  - Clearly Safe
  - Very Good
  - Excellent
- Each examiner will mark the overall performance criteria this is used for standard setting.



UNIVERSITY OF LIVERPOOL

Borderline Regression



Quality Assurance

- Internal & External review
- On the day incident reports
  - Invigilators
  - Students
- QA examiners in the clinical assessment
- Psychometric review
- Post assessment review

UNIVERSITY OF LIVERPOOL

A (very) quick guide to question performance information from Practique

To help you improve question design and interpret feedback on student performance.

Statistics Analysis report - Year 4 Written Assessment for Learning April 2022 - 4 April 2022

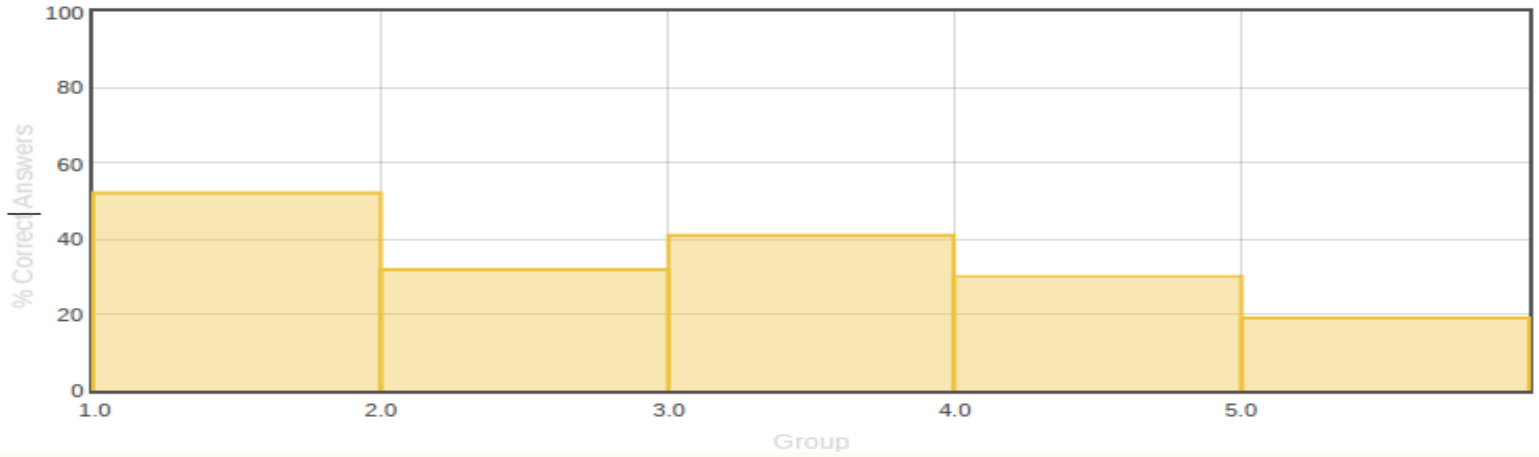
Question: 1 Single best answer question #3 (version 3)

Statistics	
33% Item discrimination:	0.240
Point Biserial:	0.252
Correct Answer:	A
Facility:	0.35

Legend

% candidates who responded a correct answer. There are 5 groups, Group 1 is top 20% of candidates by score, group 2 another 20% with lower score etc.

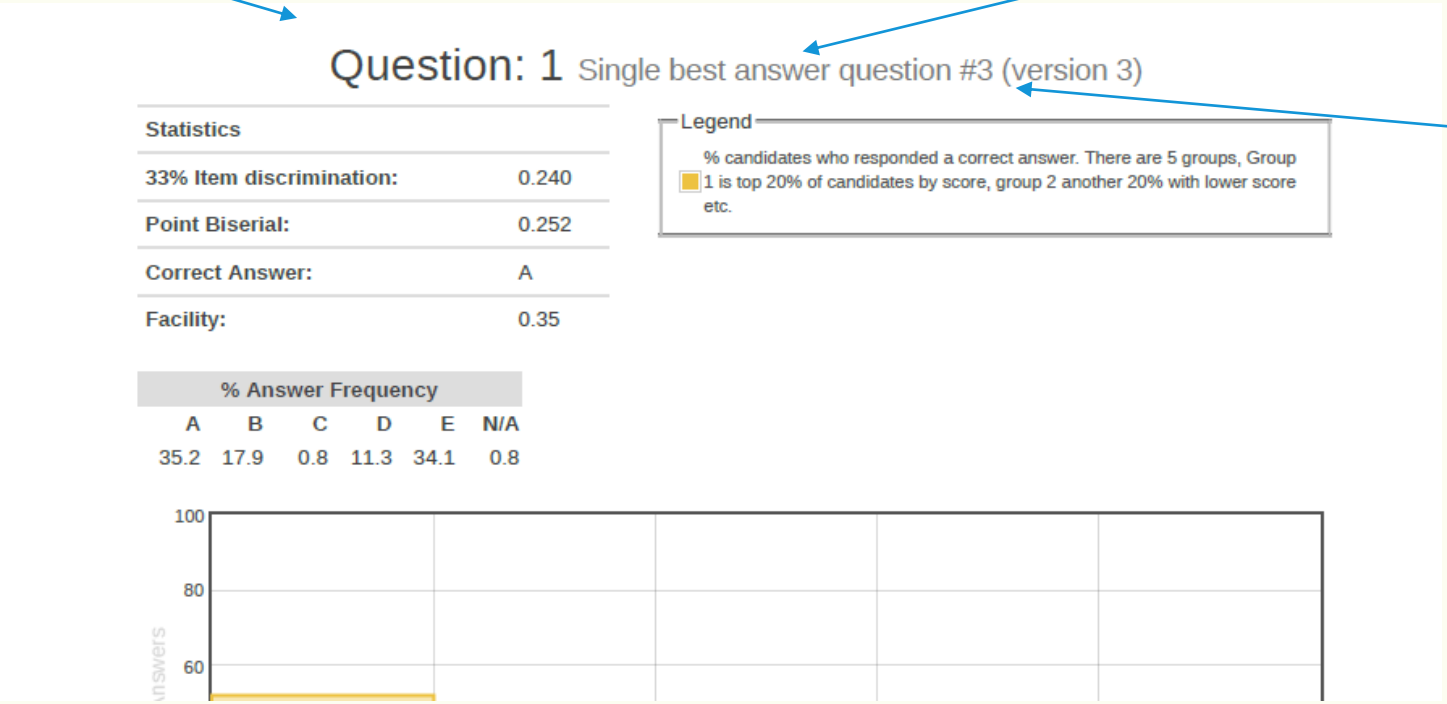
% Answer Frequency					
A	B	C	D	E	N/A
35.2	17.9	0.8	11.3	34.1	0.8



Question number: Corresponds to order of question in test/exam question pdf.

Type of question: Usually Single Best Answer (SBA)

Question identifier: Unique code for question in Practique



Question: 1 Single best answer question #3 (version 3)

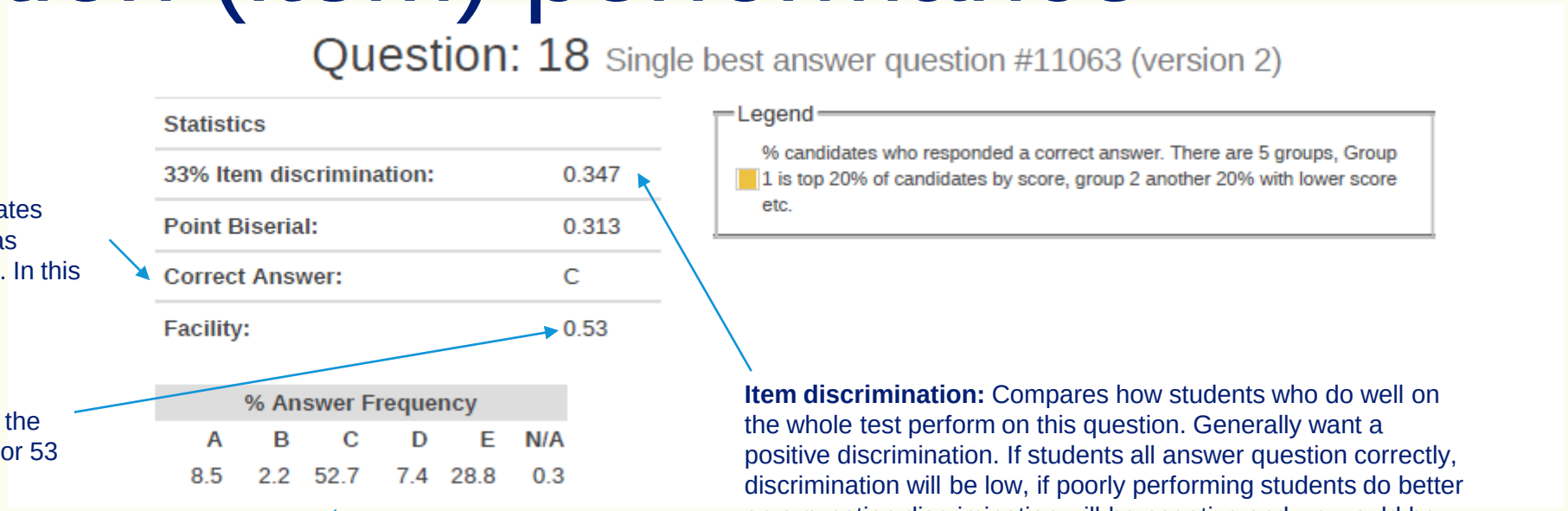
Statistics	
33% Item discrimination:	0.240
Point Biserial:	0.252
Correct Answer:	A
Facility:	0.35

Legend

% candidates who responded a correct answer. There are 5 groups, Group 1 is top 20% of candidates by score, group 2 another 20% with lower score etc.

% Answer Frequency					
A	B	C	D	E	N/A
35.2	17.9	0.8	11.3	34.1	0.8

Question (item) performance



Question: 18 Single best answer question #11063 (version 2)

Statistics	
33% Item discrimination:	0.347
Point Biserial:	0.313
Correct Answer:	C
Facility:	0.53

Legend

% candidates who responded a correct answer. There are 5 groups, Group 1 is top 20% of candidates by score, group 2 another 20% with lower score etc.

% Answer Frequency					
A	B	C	D	E	N/A
8.5	2.2	52.7	7.4	28.8	0.3

Correct answer: Indicates which option is coded as correct for the question. In this case option C

Facility: Proportion of students who answered the question correctly, 0.53 or 53 %

Item discrimination: Compares how students who do well on the whole test perform on this question. Generally want a positive discrimination. If students all answer question correctly, discrimination will be low, if poorly performing students do better on a question discrimination will be negative and we would be concerned and examine the question.

Answer frequency: Proportion of class selecting each option as their answer. We would look at any questions where the wrong answer is more popular than the coded correct

Plot of quintile performance: This gives similar information to the 33 % item discrimination, but considers the class divided into fifths by overall performance. Ideally the ‘staircase’ should be as in this example, showing that more of the students who perform well on the test overall have got this question correct than in the other fifths of the class. If the staircase goes in the other direction we would be concerned.

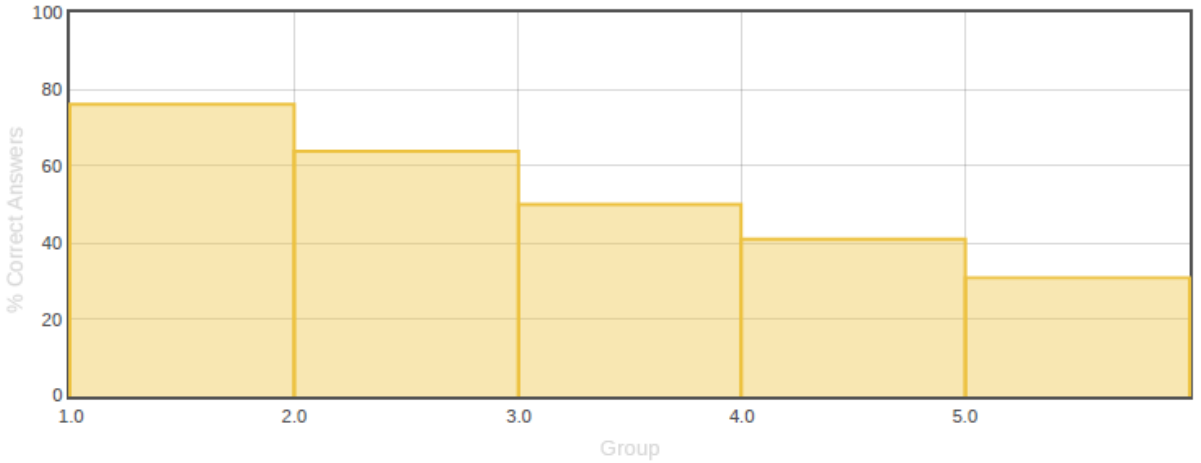
Question: 18 Single best answer question #11063 (version 2)

Statistics	
33% Item discrimination:	0.347
Point Biserial:	0.313
Correct Answer:	C
Facility:	0.53

Legend

% candidates who responded a correct answer. There are 5 groups, Group 1 is top 20% of candidates by score, group 2 another 20% with lower score etc.

% Answer Frequency					
A	B	C	D	E	N/A
8.5	2.2	52.7	7.4	28.8	0.3

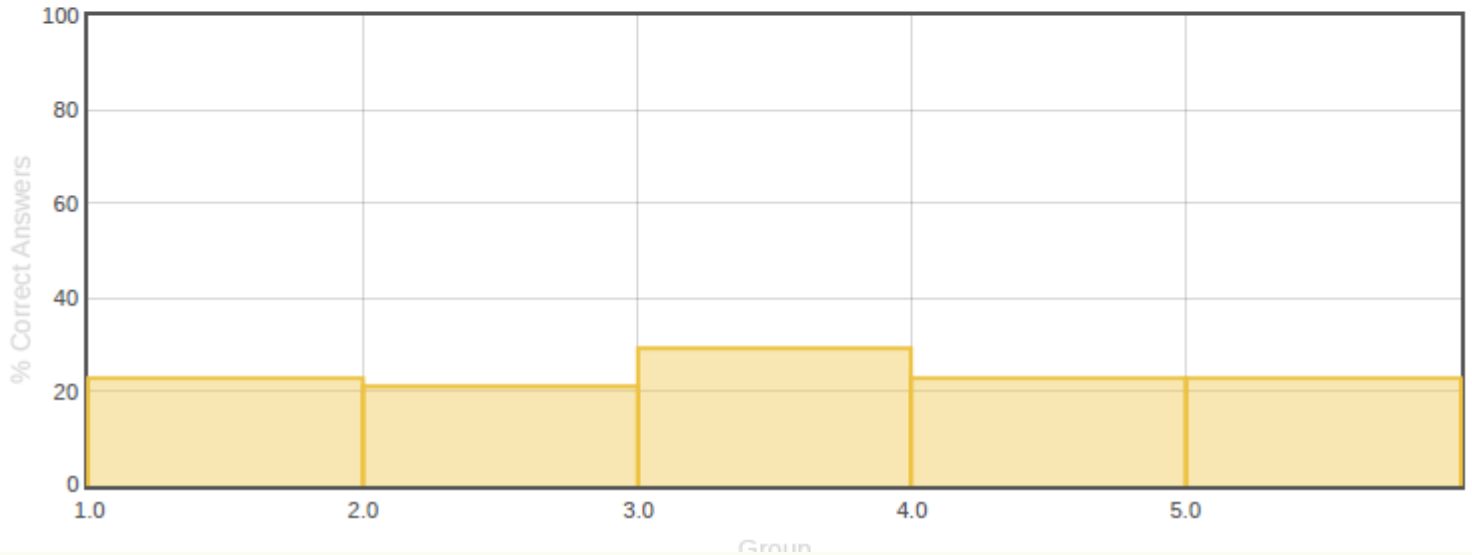


Statistics	
33% Item discrimination:	-0.008
Point Biserial:	-0.019
Correct Answer:	C
Facility:	0.24

Legend

% candidates who responded a correct answer. There are 5 groups, Group 1 is top 20% of candidates by score, group 2 another 20% with lower score etc.

% Answer Frequency					
A	B	C	D	E	N/A
4.7	36.3	24.2	26.9	7.7	0.3

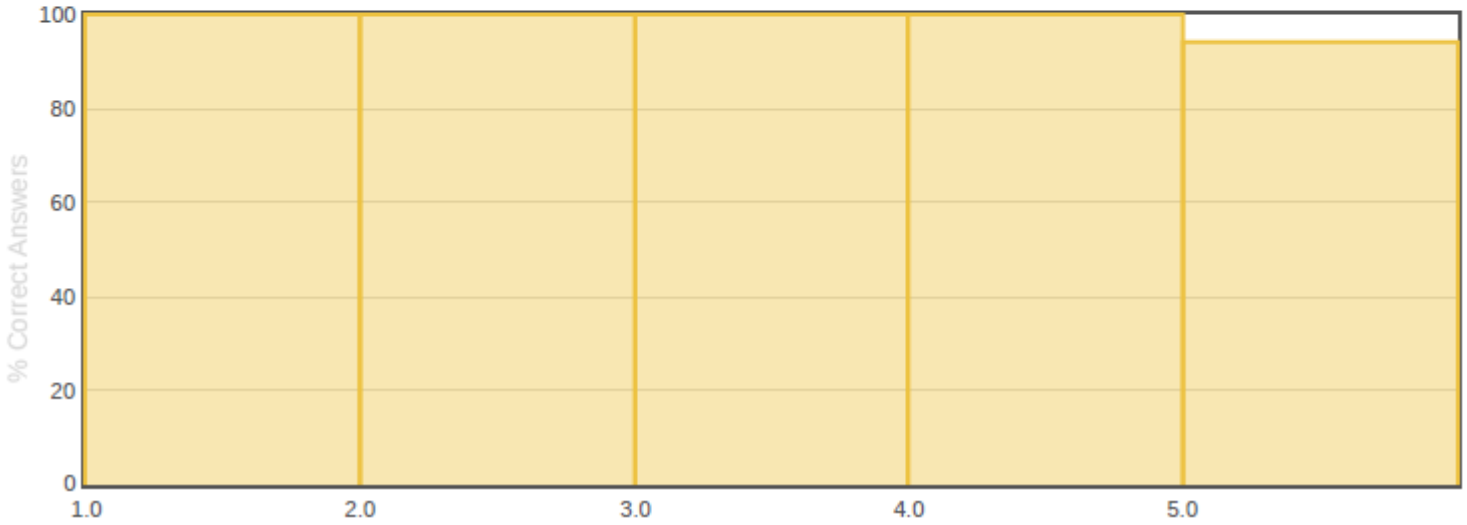


Statistics	
33% Item discrimination:	0.033
Point Biserial:	0.277
Correct Answer:	D
Facility:	0.99

Legend

% candidates who responded a correct answer. There are 5 groups, Group 1 is top 20% of candidates by score, group 2 another 20% with lower score etc.

% Answer Frequency					
A	B	C	D	E	N/A
0.3	0.3	0.0	98.9	0.3	0.3

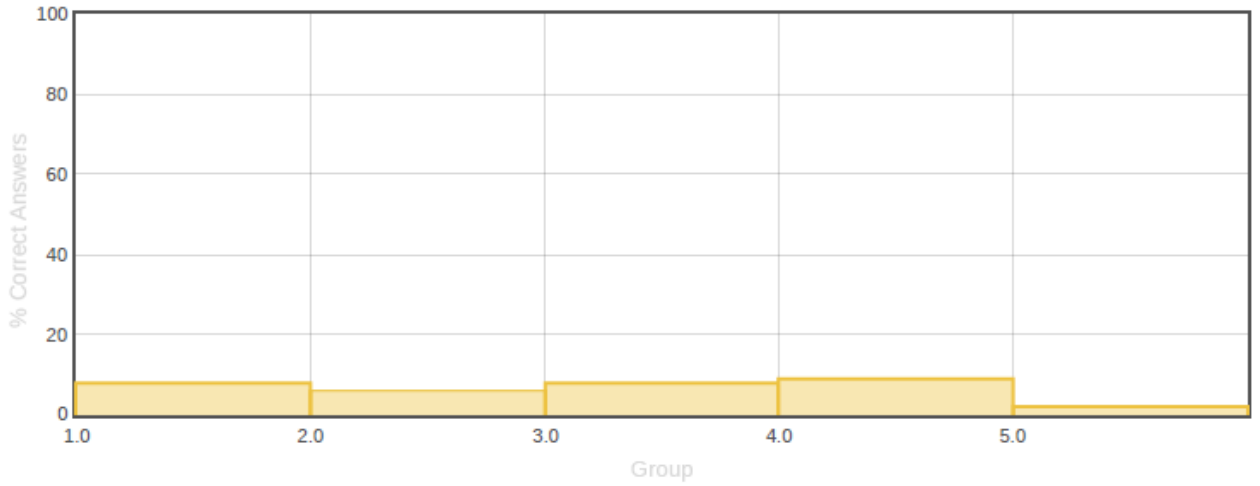


Statistics	
33% Item discrimination:	0.008
Point Biserial:	0.051
Correct Answer:	C
Facility:	0.07

Legend

% candidates who responded a correct answer. There are 5 groups, Group 1 is top 20% of candidates by score, group 2 another 20% with lower score etc.

% Answer Frequency					
A	B	C	D	E	N/A
0.3	1.1	7.1	90.9	0.0	0.5

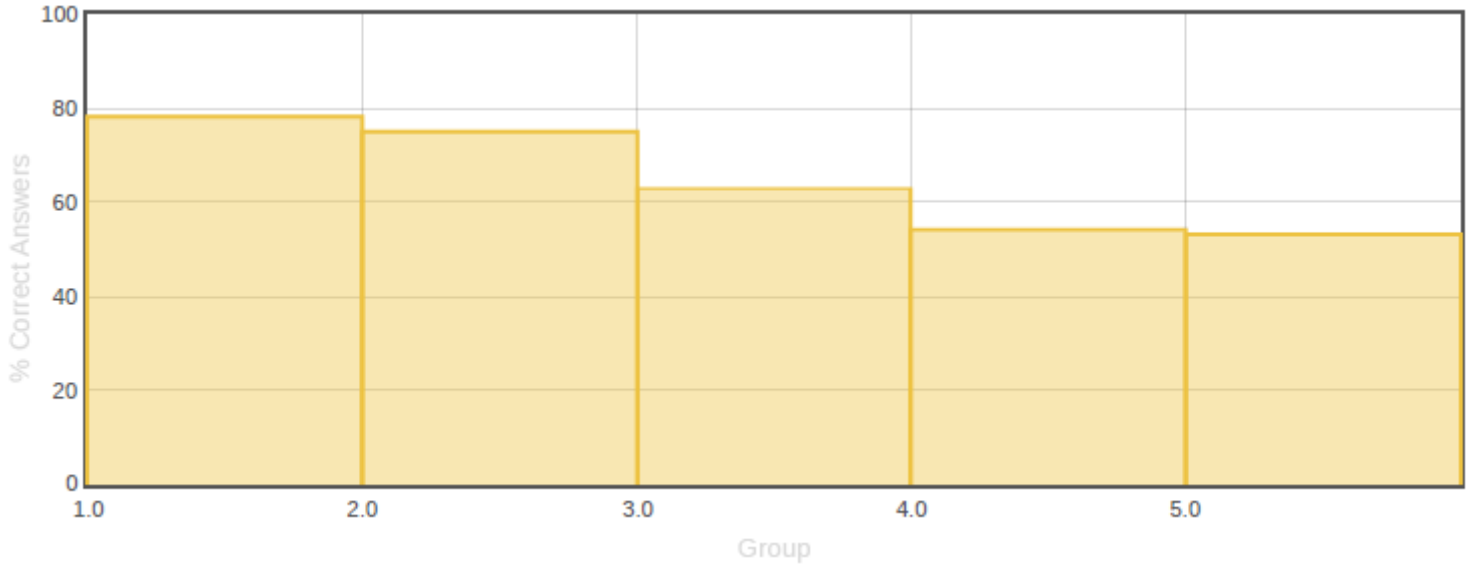


Statistics	
33% Item discrimination:	0.256
Point Biserial:	0.246
Correct Answer:	C
Facility:	0.65

Legend

% candidates who responded a correct answer. There are 5 groups, Group 1 is top 20% of candidates by score, group 2 another 20% with lower score etc.

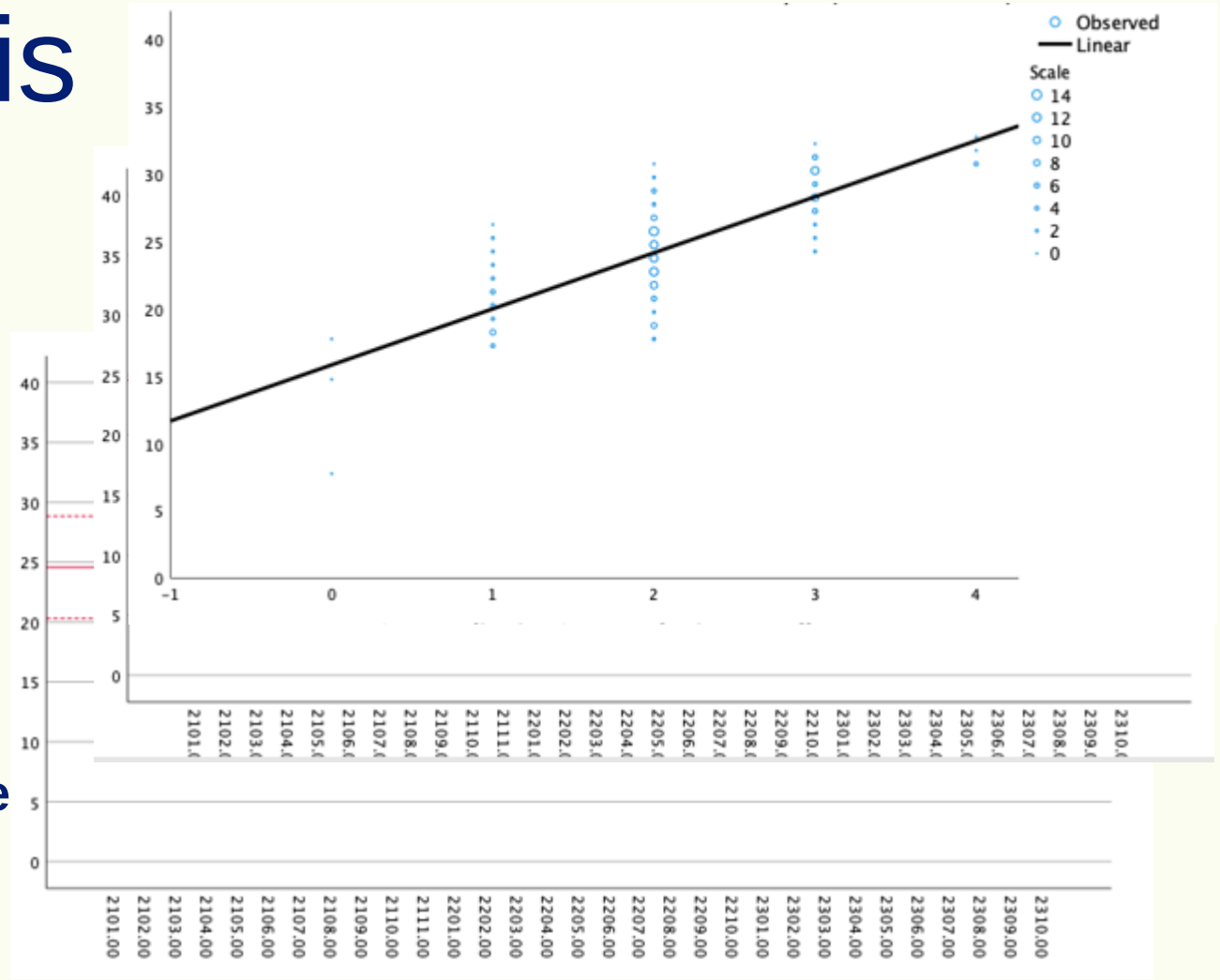
% Answer Frequency					
A	B	C	D	E	N/A
7.1	17.0	65.1	7.4	2.7	0.5



Clinical item analysis

- Overall reliability, and reliability with a station deleted

- Scatter plots – checklist vs. global grade
- R square (coefficient of determination – correlation between checklist & global score)
- Inter-grade discrimination (average checklist mark difference between grades)
- Within station, between-circuit **error variance** (a proxy for error due to assessors)



UNIVERSITY OF LIVERPOOL

Post assessment

Domain	Coho
Communication and interpersonal skills	68.18
Diagnosis, clinical reasoning, knowledge	69.77
Examination Skills	71.71
History taking skills	74.17
Investigations and management of the case/patient	64.64
Medical records and prescribing	58.51
Organised and professional interaction	63.94
Presentation skills and findings (summarising)	70.77
Procedural skills	68.19

Analysis - System Block

System Block	Cohort average
CVS	61.80%
Endo	69.15%
GI	65.85%
MSK	64.01%
Neuro	59.19%
PPGH	55.06%
Respiratory	58.98%
Therapeutics	62.65%
UG	64.46%
Overall	**62.49%**

UNIVERSITY OF LIVERPOOL

Changes with the Medical Licensing Assessment

- Two assessments – AKT and Clinical

- The schools year 5 clinical assessment will be our evidence for the clinical portion of the MLA.
- The school will hold the AKT using Liverpool University assessment regulations (Reasonable Adjustments, Extenuating circumstances, Appeals etc)
- The exam will be delivered to all schools by the MSC

Areas of Clinical Practice

Paper 1: 100 items over 2 hours

50 items

-Cardiovascular
-Respiratory
-Gastrointestinal
-Medicine of older adult

50 items

-Neurosciences
-Ophthalmology
-Endocrine & metabolic
-Renal & urology
-Infection
-Dermatology

Paper 2: 100 items over 2 hours

50 items

-Cancer
-Breast
-Haematology
-Palliative & EoL care
-Peri-op med & anaesthesia
-Musculoskeletal
-Emerg med & intensive care
-Ear, nose & throat

50 items

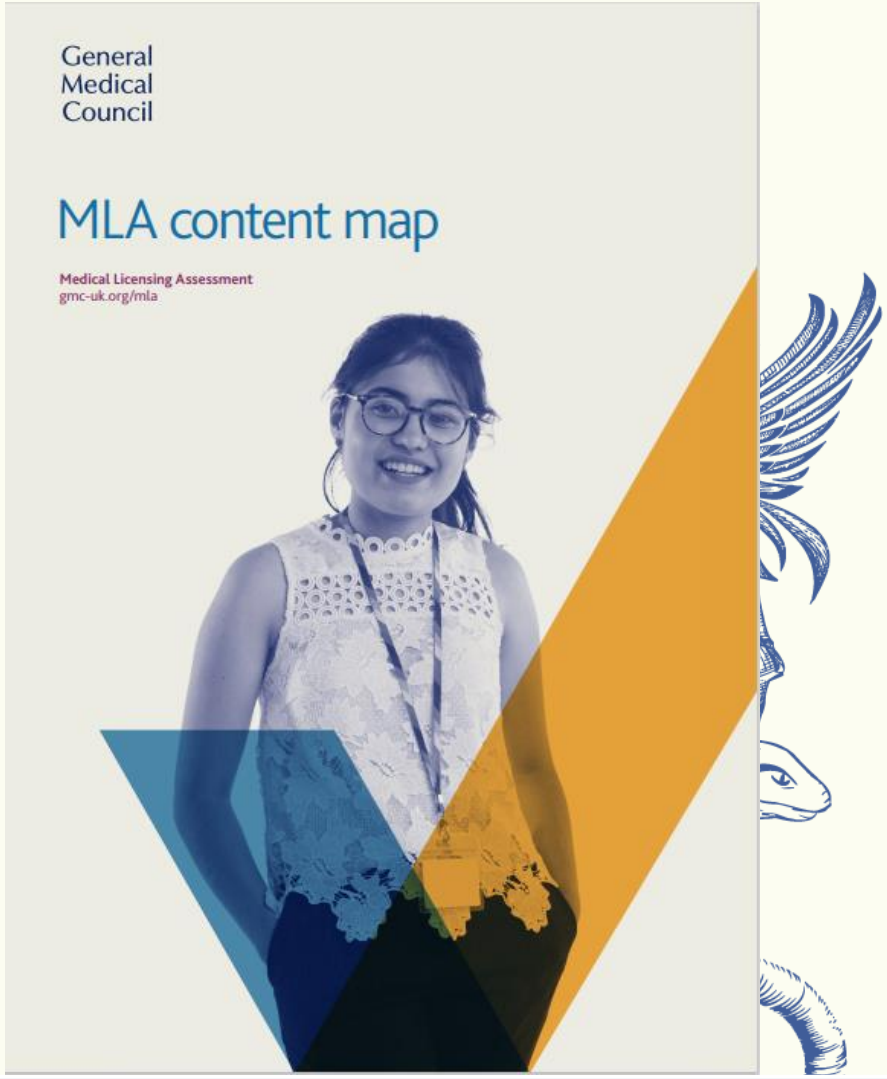
-Child health
-Mental health
-Obstetrics & gynaecology
-Sexual health
-Social/population health & research methods
-Medical ethics & law

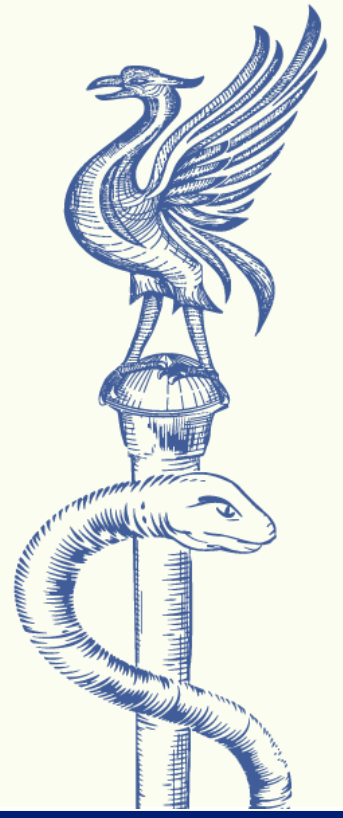
General Practice, Acute Medicine, Surgery & Clinical imaging will be covered across the four clusters.

MLA Content map

Outlines:

- Areas of Clinical Practice
- Areas of professional knowledge
- Clinical and professional capabilities
- Practical skills and procedures
- Patient presentations
- Conditions



UNIVERSITY OF LIVERPOOL

Thank you & any questions