

How to Design and Test Effective Multiple Choice Questions

Peter Devitt
Edward Palmer

Write your own MCQ

The lesion shown in the photograph has been present for six years. The most likely diagnosis is:

- A. Onychogryphosis
- B. Onchocerciasis
- C. Onychophagia
- D. Onychocryptosis
- E. Onychomycosis



Value: \$30

The process at the AMC

- Individual contribution
- Small group analysis
- Large group critique
- New questions run as pilots
- Pilots assessed for suitability of including in item bank
- Manpower:
 - Clinicians, Educationalist, Psychometrician, Statistician, Secretaries, Software

A 56-year-old man presents with a six month history of pain in his right calf. The pain is brought on by walking and eased when he stops. More recently he has experienced cramps in the right leg, particularly at night. He smokes 40 cigarettes a day. On examination his blood pressure is 150/100 mmHg. His right femoral pulse is diminished and the right dorsalis pedis pulse cannot be felt. His ankle-brachial index is 0.3 (normal 1). Nothing abnormal is found on urinalysis. Which one of the following is the most appropriate plan of management?

- A Stop smoking, encourage exercise and review in three months
- B Stop smoking encourage exercise and prescribe pentoxifylline
- C Arrange duplex studies and review
- D Arrange an arteriogram
- E Stop smoking, encourage exercise and prescribe an ACE-inhibitor

%correct: 56%
R Biserial: .23

value: \$3,000

What happened in the interim...

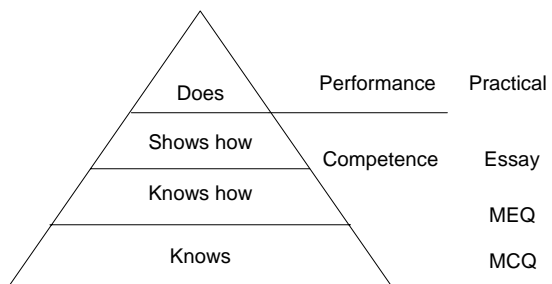
- No pay rise, but:
 - Uniformity of questions
 - Identification of item-writing flaws
 - Testing of higher cognitive skills
 - Statistical rigour
 - Security

What are you trying to test or measure?

- Knowledge:
 - Recall of fact
 - Application of fact
- Ability to reason
- Judgement and decision-making

What test instruments are available?

- Essay
- Modified Essay Question (Short Answer Question)
- Multiple choice question
- Viva voce
- Practicals (OSCE, Mini CEX)



The ideal test instrument...

- Sample the curriculum widely
- Validity
- Reliability
- Fidelity
- Educational impact

Educational impact

- Assessment stimulates learning
- The assessment processes:
 - Diagnostic
 - Formative
 - Summative

Sampling the curriculum

- What do the students know?
- What stimuli can be provided (for learning)?
- What test is best?

Validity

- Content:
 - Does the test measure what it purports to measure?
- Predictive:
 - Does it discriminate between the good and the poor students?

Reliability

Does the test produce similar results on repeat testing?

(assuming similar cohorts of examinees)

Fidelity

Does the test replicate the same challenge as would be presented in the real world?

(eg the reality of clinical medicine)

Educational Impact

- Does the test instrument stimulate learning:
 - Formative
 - Summative

Suitability of the test instruments

	Sampling	Validity	Reliability	Fidelity	Educational Impact
Essay					
MEQ					
MCQ					

Suitability of the test instruments

	Sampling	Validity	Reliability	Fidelity	Educational Impact
Essay					
MEQ					
MCQ	4				

Suitability of the test instruments

	Sampling	Validity	Reliability	Fidelity	Educational Impact
Essay					
MEQ					
MCQ	4	4			

Suitability of the test instruments

	Sampling	Validity	Reliability	Fidelity	Educational Impact
Essay					
MEQ					
MCQ	4	4	4		

Suitability of the test instruments

	Sampling	Validity	Reliability	Fidelity	Educational Impact
Essay				4	
MEQ					
MCQ	4	4	4		

Suitability of the test instruments

	Sampling	Validity	Reliability	Fidelity	Educational Impact
Essay				4	
MEQ					
MCQ	4	4	4		?

Types of MCQ

- Single, best option
- True or false
- Extended matching
- Relationship analysis
- Multiple completion

Content

Structure

Content

- Important topic
- Relevant to professional practice
- Current information
- Reasoning in preference to recall

Structure

- The stem:
 - Complete statement of the problem
 - Clarity
 - Brevity
 - Clear question

Structure

- The responses:
 - One clearly correct answer
 - Distractors:
 - Plausible
 - Less correct
 - Homogeneous content
 - Similar in construction and length

The stem...

- Satisfy the 'cover' test
- Avoid grammatical clues (sentence completion)
- Minimise cueing
- Avoid redundancy

Cueing

- Seeing the correct answer amongst the distractors without understanding the question:
 - Grammatical (only one response follows grammatically)
 - Repetition of word or phrase from the stem
 - Long, specific response
 - Use of key descriptive word or phrase
 - Convergence (all components in one response)

Analysis of examples

Concerning primary hyperaldosteronism (Conn's syndrome):

1. Most patients have a persistent hyperkalaemia of over 6.0 mmol/L
- 2.* The tumours are yellow in colour because they are full of cholesterol
3. The tumours may be localised using ¹³¹I metaiodobenzylguanidine (I¹³¹MIBG)
- 4.* Most of the tumours are benign
- 5.* These tumours can present with hypertension

Example

Rectovaginal fistula may occur in all the following situations EXCEPT:

- A. after prolonged and difficult labour
- B. in a patient with chronic perianal sepsis secondary to Crohn's disease
- C. as a complication of lymphogranuloma venereum
- D. after irradiation for prostatic carcinoma
- E. in a patient with local recurrence of rectal cancer

Example

Central retinal artery occlusion can be caused by giant cell arteritis. Is the ESR raised:

- A. sometimes
- B. always
- C. mostly
- D. occasionally
- E. never

Example

Tissue biopsies are placed in formalin to

- a) enable a frozen section to be performed if necessary
- b) prepare them for paraffin embedding
- c) aid H&E staining
- d) prevent autolysis*
- e) prevent them drying out

Example

A 23 year old man presents with a pigmented lesion on his right calf. This is locally excised and the histology report describes the lesion as a malignant melanoma of depth 0.72 mm, Clark's level II, with tumour extending to 1 mm of the resection margin. What is the optimum treatment of this lesion?

- A. Continued observation, with no further active treatment at this stage
- B. Further excision of the wound site, ensuring a 1 cm margin of clearance
- C. Further excision of the wound site, ensuring a 2 cm margin of clearance
- B. Further excision of the wound site, ensuring a 5 cm margin of clearance and a skin graft
- E. Further excision of the wound site and inguinal lymph node dissection

Example

Which one of the following statements regarding prepatellar pain in adolescence caused by Osgood-Schlatter disease is correct?

- 1. It frequently leads to patellar tendon avulsion.
- 2. It is best treated by cortisone injection.
- 3. It is usually self-limited.
- 4. It frequently requires cessation of sports activities.
- 5. It usually comes on before age 10.

Example

An 18-year-old man is injured playing football and complains of abdominal pain. When seen in the Emergency Department his pulse is 110/min, temperature 37°C and blood pressure on lying of 85/55 mm Hg. He is tender in the left upper quadrant of his abdomen. An intravenous line has been inserted by the paramedical team and an infusion of saline is running. What is the next most appropriate stage in his management?

- A. Perform a spiral CT scan
- B. Give one litre of colloid rapidly and reassess
- C. Transfer to an intensive care unit
- D. Perform an abdominal ultrasound examination
- E. Arrange for an urgent laparotomy

Example

A 74-year-old woman presents to the Emergency Department after passing many stools with gross blood. She describes the blood as dark red with clots. She is hemodynamically stable. Which one of the following is not an appropriate investigation to determine the cause of the bleeding?

1. Flexible colonoscopy.
2. Labeled red cell scan.
3. Gallium scan.
4. Angiogram.
5. Anoscopy and rigid sigmoidoscopy

Example

A patient presents with a history of hematemesis of about 1 litre over the preceding 2 hours and the passage of 2 melena stools. The pulse is 95/minute and the blood pressure 120/80 mmHg. After inserting an intravenous line, which one of the following is the most appropriate next course of management?

1. Start immediate treatment with a proton pump inhibitor.
2. Start therapy for *Helicobacter pylori* eradication.
3. Perform immediate surgery because the risk of uncontrollable rebleeding is high.
4. Make immediate referral for gastroscopy.
5. Begin therapy with a mucosal protective drug such as sucralfate.

Example

A 23-year-old man is resuscitated from multiple traumas following a motorcycle accident. Several days later, he is noted to have increasing dyspnea. Which one of the following statements about the investigation of adult respiratory distress syndrome (ARDS) is correct?

1. A decreased P_aO_2 is the earliest abnormality seen.
2. Early chest x-ray changes are diagnostic.
3. Hypoxemia responds well to increases in the inspired oxygen concentration (FI_{O_2}).
4. An increased P_aCO_2 with a decreased P_aO_2 often occurs early in the course of ARDS.
5. Evidence of systemic sepsis is rarely present on investigation.

Example

Patients known to be hepatitis B carriers are of concern, as they may infect their babies at birth. How could you protect the neonate?

1. Give hepatitis B immunoglobulin to the mother.
2. Give hepatitis B immunoglobulin to the neonate.
3. Give hepatitis B vaccine to the mother.
4. Give hepatitis B immunoglobulin and hepatitis B vaccine to the neonate.
5. Therapy is unnecessary because of transfer of passive immunity from the mother.

Example

A 4-year-old boy presents with a history of constipation from the age of 6 months. His stools, produced every 3 to 4 days, are described as large and hard. Physical examination is normal apart from the rectal examination which reveals a capacious rectum, poor sphincter tone, and stool in the rectal vault. Which one of the following is the most likely diagnosis?

1. Hirschsprung disease.
2. Functional constipation.
3. Hypercalcemia.
4. Intestinal obstruction.
5. Hypothyroidism.

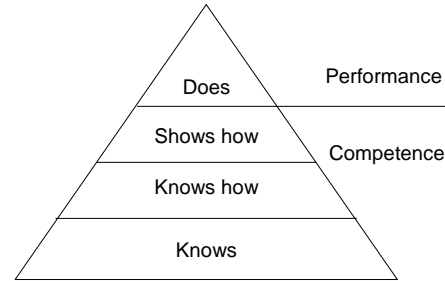
The Challenge...

Design and construct the perfect MCQ

(15 min)

How to protect your item bank

- Produce detailed stems
- Write groups of questions:
 - Same or similar topics
 - Small variations in the stem
 - Identical responses - but slightly different stem to fit each of the five responses
- Avoid using images



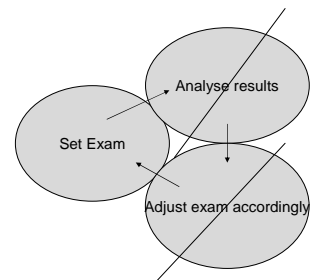
References

- <http://www.tlc.murdoch.edu.au/eddev/evaluation/mcq/mctests.html>
- <http://www.cs.nott.ac.uk/~smx/PGCHE/E-type.htm>

Why not essays or MEQs?

- Time-consuming to mark
- Subjective:
 - Poor inter-rater reliability
- Limited sampling

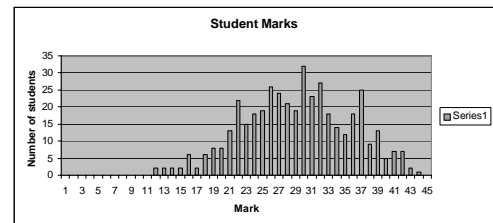
Analysis in practice



Post test

- What is the process for testing the quality of the questions in your area?

Example



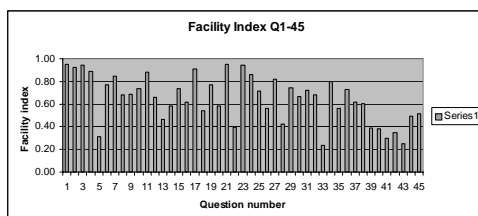
Question Facility / Difficulty

- Facility = number of correct answers/total answers
- Difficulty = 1 – facility
- What is a good value for facility?

Facility

- Values between 0.3 and 0.7 are considered good. Questions with facility outside this range should be examined seriously with the intent of the question considered.

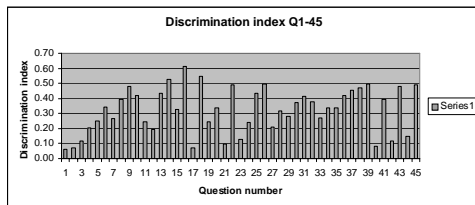
Facility Index



Discrimination

- Do better students outperform weaker students in any given question?
- High performing group (top quartile), has N students
- Out of the top quartile, CH students get the answer correct
- Out of the bottom quarter, CL students get it correct
- Discrimination = $(CH - CL)/N$
- A value of 0.3 is considered good

Discrimination



Distractors

- Are the choices available to the student reasonable?
- Are they too easy or too hard?
- A good distractor should allow good students to realise it is incorrect but will confuse poorer students.

Distractors

- The number of people expected to pick each distractor is:
 - $n = (\text{number answered item wrong}) / \text{number of distractors}$

Item Response Theory

Rasch Analysis

- Considers difficulty of item AND ability of student
- If student ability = difficulty of item then the probability p of getting answer correct is 0.5
- Iterative process
- Single parameter Item Response Theory, parameter is difficulty of item
- Assumes discrimination = 1
- Assumes chance does not play a part

Discrimination

- Rasch analysis assumes discrimination = 1
- Post hoc calculation adjusts this for each item
- Value > 1: Discriminates more than expected for an item of this difficulty

