ITEM WRITING FOR ASSESSMENTS

HANDBOOK

MCQs

EMQs

SAQs

OSCEs

JUNE 2014

AUSTRALASIAN COLLEGE
FOR EMERGENCY MEDICINE

INTRODUCTION

This Handbook has been developed for the use of high quality item writing for MCQs, EMQs, SAQs and OSCEs. For each of these forms of assessment you are provided with:

- **A brief introduction, examples and guidelines.**

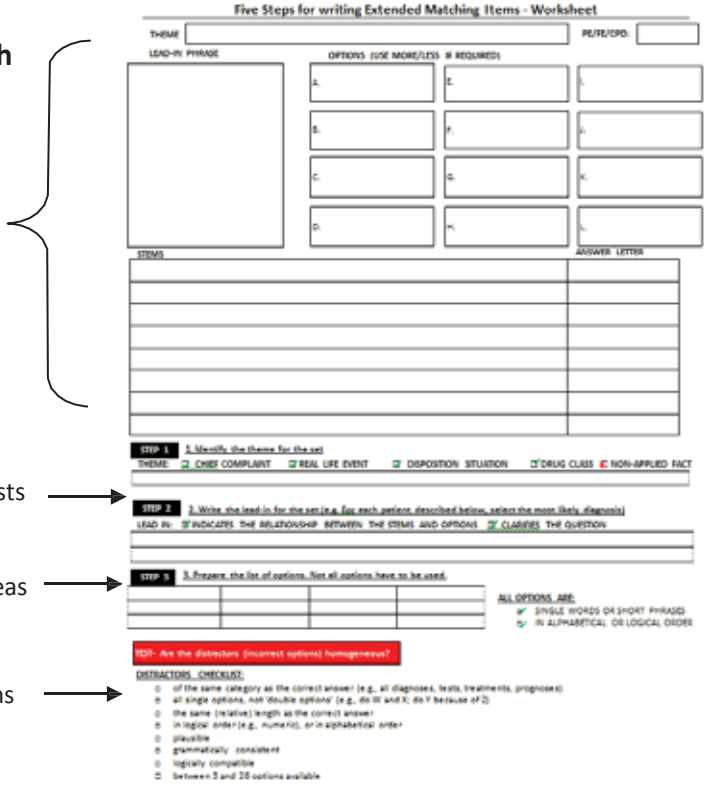
- **A practical worksheet with the following sections:**

An area to write up the draft item – final versions should be written into the electronic template and submitted electronically.

Steps to write items with checklists

An area to work through ideas

Tests for high quality items



- **Excerpts from current item writing literature for reference and further reading.**

- **The college templates (also available electronically)**: Once you have used the worksheet, the question can be entered into the college template. These are available electronically on the ACEM website.

- **Confidentiality and Intellectual Property Statement**

AUSTRALASIAN COLLEGE
FOR EMERGENCY MEDICINE

CONTENTS

Multiple Choice Questions

Reference: **Case, D. & Swanson, S.** 2002, *Constructing Written Test Questions for the Basic & Clinical Sciences,* Third Edition.

Brief Introduction to writing one-best Answer MCQs

MCQs utilise a stem that will ask the examinee to choose the one best answer from four options. The aim is to ensure clinical relevance of items and reduce items testing recall of knowledge. Advantages of their use include that they:

- are more salient to most areas of medicine
- an efficient sampling of knowledge
- machine scoreable

To start writing:

1. Read through the MCQ Writing Guidelines – Five Tests
2. Use the MCQ Worksheet in this pack to practice

When you are familiar with the process use the MCQ Template available on the ACEM website (example template also included within this pack) alongside the Writing Guidelines.

Examples

MCQ Example One

A 32 year old man has a 4 day history of progressive weakness in his extremities. He has been healthy except for an upper respiratory tract infection 10 days ago. His temperature is 37.8C. BP is 130/80 mmHg, Pulse 94bpm, and respiration rate is 42 and shallow. He has symmetric weakness of both sides of his face as well as proximal and distal extremity muscles. Sensation is intact. No tendon reflexes can be elicited and the plantar responses are flexor.

Which of the following is the most likely diagnosis?

A) Guillain-Barré Syndrome
B) Myasthenia gravis
C) Poliomyelitis
D) Polymyositis

Answer: A

MCQ Example Two

What is MOST likely to exacerbate a case of pelvic inflammatory disease?

A) Menstruation
B) Pregnancy
C) Progesterones
D) Urinary tract infection

Answer: A

MCQ Example Three

Following an overdose of a substance, which of the following is the MOST likely ingestion – side effect combination

A) Metoprolol – profound hypotension
B) Metformin – profound hypoglycaemia
C) Propanolol – QRS widening
D) Venlafaxine – hepatic failure

Answer: C

Further reading

- Appendix 1 – **Multiple Choice Questions** from Jolly, B. 2014, *Written Assessment* in Swanwick, T. (ed) *Understanding Medical Education: Evidence, Theory and Practice*, Second Edition, pp. 264-265. (Full chapter available on request)
- Appendix 2 – **Types of written assessments and their primary usages** from Jolly, B. 2014, *Written Assessment* in Swanwick, T. (ed) *Understanding Medical Education: Evidence, Theory and Practice*, Second Edition, pp. 258-260. (Full chapter available on request)
- **Section 2 – Writing One Best-Answer Questions for the Basic and Clinical Sciences** from Case, S. & Swanson, D. 2002, *Constructing Written Test Questions for the Basic and Clinical Sciences*, Third Edition, pp. 31-66

MCQ Writing Guidelines – Five Tests

Reference: Case. S & Swanson D (2001 3rd Edition) *Constructing Written Test Questions for the Basic & Clinical Sciences*

TEST 1

1. Each item should focus on a clinically important concept or problem

- Avoid items that only require recall of isolated facts
- Focus on: - common, serious or potentially catastrophic clinical problems
  - problems that would be encountered in real life Emergency Medical Practice.

EXAMPLE

A six-year-old boy has a one-day history of (*description of presenting complaint, history, physical exam*)? What is
✓ the most appropriate therapy?

A six year old boy has a one-day history of (*description of presenting complaint, history, physical exam*)? What is
✘ the best drug to treat Otitis media?

Regarding Otitis media?

TEST 2

2. Each item should assess the application of knowledge

- Use clinical vignettes
- Focus items on key concepts and principles that are essential information (without access to references) for all examinees to understand.

EXAMPLE

✓ A 62-year-old man develops left-sided limb ataxia, Horner's syndrome, nystagmus, and loss of appreciation of facial pain and temperature sensations. Which artery is most likely to be occluded?

✘ Rotator cuff muscles include?

TEST 3

3. The stem of the item must pose a clear question, and it should be possible to arrive at an answer with the options covered.

- To determine if the question is focused, cover up the options and see if the question is clear and if the examinees can pose an answer based only on the stem.
- Rewrite the stem and/or options if they could not.

EXAMPLE

✓ A patient has (*symptoms and signs*). Which of the following is the most likely explanation for the (*findings*)?

✘ Which of the following is incorrect regarding the trigeminal nerve?

TEST 4

4. All distractors (i.e., incorrect options) should be homogeneous.

- They should fall into the same category as the correct answer (e.g., all diagnoses, tests, treatments, prognoses, disposition alternatives).
- Avoid using 'double options' (e.g., do W and X; do Y because of Z), unless the correct answer and all distractors are double options.
- All distractors should be plausible, grammatically consistent, logically compatible, and of the same (relative) length as the correct answer.
- Order the options in logical order (e.g., numeric), or in alphabetical order.

EXAMPLE

✓ A 65-year-old man has difficulty rising from a seated position and straightening his trunk, but he has no difficulty flexing his leg. Which of the following muscles is most likely to have been injured?

A. Gluteus maximus
B. Gluteus minimus
C. Hamstrings
D. Iliopsoas
E. Obturator internus

✘ Regarding calcium containing solutions?

A. it is recommended to give calcium chloride Intramuscularly
B. calcium chloride is safe to give via rapid intravenous bolus
C. calcium gluconate is the preferred solution for intravenous administration
D. 10ml of 10% calcium gluconate contains the same mmol of calcium ions as 10ml of 10% calcium chloride

TEST 5

5. Avoid technical item flaws that provide special benefit to testwise examinees or that pose irrelevant difficulty.

- Do NOT write any questions of the form 'Which of the following statements is correct?' or 'Each of the following statements is correct EXCEPT.'

EXAMPLE

✓ A (*patient description*) has (*abnormal findings*). Which of the following (*positive laboratory results*) would be expected?

✘ With regard to the duodenum, which is NOT TRUE?

Five Tests for One-Best-Answer MCQs- Worksheet

Reference: **Case, D & Swanson, S** 2002, *Constructing Written Test Questions for the Basic & Clinical Sciences,* Third Edition.

PE/FE/CPD:

Correct answer letter:

QUESTION ITEM STEM

OPTIONS (in alphabetical order)

A.

B.

C.

D.

TEST 1 1.**Each item should focus on an important concept, typically a common or potentially catastrophic clinical problem.**

CONCEPT: ☑ **CLINICAL PROBLEM** ☑ **REAL LIFE EVENT** ☒ **ISOLATED FACT**

TEST 2 2.**Each item should assess the application of knowledge.**

APPLICATION OF KNOWLEDGE: ☑ **CLINICAL VIGNETTE** ☑ **ESSENTIAL INFORMATION** ☒ **ISOLATED FACT**

TEST 3 3.**The stem of the item must pose a clear question, and it should be possible to arrive at an answer with the options covered.**

COVER THE ANSWERS AND WRITE DOWN YOUR OWN, DOES IT MATCH? ☑ **MATCH** ☒ **NO MATCH**

TEST 4 4.**All distractors (ie, incorrect options) should be homogeneous.**

LIST DISTRACTORS

ALL DISTRACTORS ARE:

- ☑ of the same category as the correct answer (e.g., all diagnoses, tests, treatments, prognoses)
- ☑ all single options, not "double options" (e.g., do W and X; do Y because of Z)
- ☑ the same (relative) length as the correct answer
- ☑ in logical order (eg, numeric), or in alphabetical order
- ☑ plausible
- ☑ grammatically consistent
- ☑ logically compatible

TEST 5 5.**Avoid technical item flaws that provide special benefit to testwise examinees or that pose irrelevant difficulty.**

CHECK FOR TECHNICAL ITEM FLAWS

TECHNICAL ITEM FLAWS:

- ☒ Which of the following statements is correct?
- ☒ Which of the following statements is correct EXCEPT?

Extended Matching Questions

Reference: **Case, D. & Swanson, S.** 2002 *Constructing Written Test Questions for the Basic & Clinical Sciences,* Third Edition.

Brief Introduction to Writing a Single-Answer EMQ

Extended Matching Questions (EMQs) are a form of multiple choice questions. You may also hear or read about them referred to as 'R-type' items. They are used extensively in medical education at all levels on an international basis. Advantages of their use include that they:

- can contain minimal cueing (thereby reducing the artefacts produced by testwise students).
- are much less dependent on semantic and grammatical features than other test forms.
- can be written in an easier, more realistic style.
- are good discriminators of students (i.e. they separate poorly and well performing students very efficiently with good reliability).

The format is particularly suited to testing reasoning, application of knowledge to practice and clinical problem solving MCQs, rather than the more regular knowledge-based, single fact recall questions.

To start writing:

1. Read through the EMQ Writing Guidelines – Five Tests
2. Use the EMQ Worksheet in this pack to practice

When you are familiar with the process use the EMQ Template available on the ACEM website (example template also included within this pack) alongside the Writing Guidelines.

Examples

EMQ Example One

Theme: Toxicology

Options:

a) Aspirin
b) Butane
c) Chlorine gas
d) Clonidine
e) Clozapine
f) Colchicine
g) Doxepin
h) Erythromycin
i) Ferrous sulphate
j) Flecainide
k) Fluoxetine
l) Metformin
m) Methanol
n) Morphine
o) Omeprazole
p) Organophosphate
q) Paracetamol
r) Promethazine
s) Propranolol
t) Quetiapine

Lead-in: For each patient admitted to the emergency department with an overdose select the most likely medication or compound ingested:

1. A 19 year old woman with a history of depression. She is vomiting and confused

   Her vital signs are:

HR	120	beats/min (regular)
BP	100/70	mmHg
RR	20	/min
O_2 Saturation	98	% on air

Temperature	37.7	^{o}C	
ECG: sinus tachycardia			
Arterial Blood Gas			Reference range
pH	7.45	mmHg	(7.36-7.44)
pCO_2	24	mmHg	(35-45)
pO_2	100	mmHg	
HCO_3^-	18	mmol	(21-28)
Electrolytes			
Na^+	140	mmol/L	(135-145)
K^+	3.7	mmol/L	(3.2-4.3)
Cl^-	100	mmol/L	(99-109)
BUN	10	mmol/L	(3.0 – 8.0)
Creat	75	mmol/L	(45-85)
BSL	5.0	mmol/L	(3.0 – 5.5)
Lactate	0.8	mmol/L	(0.5 -2.0)

Answer: a) Aspirin

2. A 35 year old man with no significant past medical history. Presents after ingestion of unknown substance. He is confused and agitated, complaining of blurred vision.
His vital signs are:

HR	90	beats/min (regular)
BP	95/70	mmHg
RR	30	/min
O_2 Saturation	98	% on air
Temperature	36.8	^{o}C
ECG: sinus tachycardia		

Arterial Blood Gas			Reference range
pH	7.07	mmHg	(7.36-7.44)
pCO_2	24	mmHg	(35-45)
pO_2	100	mmHg	
HCO_3^-	8	mmol/L	(21-28)
Electrolytes			
Na^+	140	mmol/L	(135-145)
K^+	3.7	mmol/L	(3.2-4.3)
Cl^-	90	mmol/L	(99-109)
BUN	10	mmol/L	(3.0 – 8.0)
Creat	75	mmol/L	(45-85)
Glucose	6.5	mmol/L	(3.0-5.5)
Lactate	3.0	mmol/L	(0.5 – 2.0)

Answer: m) Methanol

EMQ Example Two

Theme: Paediatrics

Options:

a) Acute renal failure
b) Appendicitis
c) Constipation
d) Diabetic ketoacidosis
e) Gastroenteritis
f) Intussusception
g) Malrotation with volvulus
h) Mesenteric adenitis
i) Psychogenic vomiting
j) Pyloric stenosis

Lead-in: For each child presenting with vomiting described select the most likely diagnosis from the options provided:

1. A 5 year old boy with no significant past history complains of abdominal pain and vomiting for 2 days. He has had a recent URTI and has been bed wetting. He looks unwell with deep respirations and is moderately dehydrated.

   His vital signs are:

HR	160	beats/min (regular)
BP	90/60	mmHg
RR	30	/min
O_2 Saturation	100	% on air
Temperature	36.5	^oC

Answer: d) Diabetic ketoacidosis

2. A 1 week old child has been vomiting for 24 hours. The child is breast fed, and has been wetting nappies but had no bowel motions. The child is intermittently grizzly.

   Vital signs are:

HR	150	beats/min (regular)
RR	30	/min
O_2 Saturation	100	% on air
Temperature	36.8	^oC

Answer: g) Malrotation with volvulus

Further reading

- Appendix 1 - **Extended Matching Questions** from Jolly, B. 2014, *Written Assessment* in Swanwick, T. (ed) *Understanding Medical Education: Evidence, Theory and Practice*, Second Edition, pp. 264-265. (Full chapter available on request)
- Appendix - **Types of written assessments and their primary usages** from Jolly, B. 2014, *Written Assessment* in Swanwick, T. (ed) *Understanding Medical Education: Evidence, Theory and Practice*, Second Edition, pp. 258-260. (Full chapter available on request)
- **Section 3 – Extended Matching Items** from Case, S. & Swanson, D. 2002, *Constructing Written Test Questions for the Basic and Clinical Sciences*, Third Edition, pp. 69-103
- Case, S. & Swanson, D. 1993 **Extended-matching items: a practical alternative to free-response questions**, *Teaching and Learning in Medicine: An International Journal, 5:2,* pp:107-115.

EMQ Writing Guidelines – Five Tests

Reference: Case, S & Swanson, D 2002 *Constructing Written Test Questions for the Basic & Clinical Sciences,* Third Edition

TEST 1

1. Each item should focus on a clinically important concept or problem.

- Avoid items that only require recall of isolated facts.
- Focus on: - common, serious or potentially catastrophic clinical problems.
  - problems that would be encountered in real life emergency medical practice.

EXAMPLE

✓ Lead-In: For each patient with fatigue, select the most likely diagnosis.
Stem: A 15 year old girl has a 2 week history of fatigue and back pain. She has widespread bruising, pallor, and tenderness over the vertebrae and both femurs. A complete blood count shows hemoglobin concentration of 7.0 g/dL, leukocyte count of 2000/mm3, and platelet count of 15,000/mm3.
Option: Acute leukemia

TEST 2

2. Each item should assess the application of knowledge

- Use clinical vignettes.
- Focus items on key concepts and principles that are essential information (without access to references) for all examinees to understand.

EXAMPLE

✓ For each of the following patients select the most appropriate therapy
✘ For each of the following patients select the best drug to treat Otitis media.

TEST 3

3. The stem of the item must pose a clear question, and it should be possible to arrive at an answer with the options covered.

- To determine if the stem is focused, cover up the options and see if the examinees can pose an answer based only on the stem.
- Rewrite the stem and/or options if they could not.
- A lead-in must be specified for each set.
- Include lead-ins specifying the relationship between the items and the options.

EXAMPLE

For each of the following patients with [chief complaint], select the most likely diagnosis.

✓ - the item will be a patient with a chief complaint

TEST 4

4. All distractors (i.e., incorrect options) should be homogeneous.

- They should fall into the same category as the correct answer (e.g., all diagnoses, tests, treatments, prognoses, disposition alternatives).
- Avoid using 'double options' (e.g., do W and X; do Y because of Z) unless the correct answer and all distractors are double options.
- All distractors should be plausible, grammatically consistent, logically compatible, and of the same (relative) length as the correct answer.
- Order the options in logical order (e.g., numeric), or in alphabetical order.
- They should include between 3 and 26 options -- all those that require an appropriate level of discrimination.

✓
A. Ankylosing spondylitis
B. Osteoporosis
C. Intervertebral disc infection
D. Spinal stenosis
E. Multiple myeloma
F. Myofascial pain

✘
A. is motion sickness
B. have no effects on people
C. is a reduction in visibility
D. cause death
E. esthetics, economics, health
F. are completely controlled

TEST 5

5. Avoid technical item flaws that provide special benefit to testwise examinees or that pose irrelevant difficulty.

- Sets without lead-ins (or with nonspecific lead-ins, such as 'Match each item with the best option') should NOT be used, because they generally pose inconsistent or ambiguous tasks for examinees.

Five Steps for writing Extended Matching Questions - Worksheet

THEME | **PE/FE/CPD:**

LEAD-IN PHRASE

OPTIONS (IN ALPHABETICAL ORDER - USE MORE/LESS AS REQUIRED)

A.	E.	I.
B.	F.	J.
C.	G.	K.
D.	H.	L.

STEMS	ANSWER LETTER

STEP 1 **1. Identify the theme for the set**

THEME: ☑ **CHIEF COMPLAINT** ☑ **REAL LIFE EVENT** ☑ **DISPOSITION SITUATION** ☑ **DRUG CLASS** ☒ **NON-APPLIED FACT**

STEP 2 **2. Write the lead-in for the set (e.g. For each patient described below, select the most likely diagnosis)**

LEAD IN: ☑ **INDICATES THE RELATIONSHIP BETWEEN THE STEMS AND OPTIONS** ☑ **CLARIFIES THE QUESTION**

STEP 3 **3. Prepare the list of options. Not all options have to be used.**

ALL OPTIONS ARE:

- ☑ SINGLE WORDS OR SHORT PHRASES
- ☑ IN ALPHABETICAL OR LOGICAL ORDER

TEST- Are the distractors (incorrect options) homogeneous?

DISTRACTORS CHECKLIST:

- ☑ of the same category as the correct answer (e.g., all diagnoses, tests, treatments, prognoses)
- ☑ all single options, not 'double options' (e.g., do W and X; do Y because of Z)
- ☑ the same (relative) length as the correct answer
- ☑ in logical order (e.g., numeric), or in alphabetical order
- ☑ plausible
- ☑ grammatically consistent
- ☑ logically compatible
- ☑ between 3 and 26 options available

STEP 4 **4. Write the stems (Not all stems have to be used in every exam)**

ARE THE STEMS:

- ☑ WITHIN A SET SIMILAR IN STRUCTURE
- ☑ PATIENT VIGNETTES (IF APPROPRIATE)
- ☑ KEY CONCEPTS AND PRINCIPLES THAT ARE ESSENTIAL INFORMATION

TEST- Are the stems of the item focused and pose a clear question?

COVER THE OPTIONS AND WRITE DOWN YOUR OWN, DOES IT MATCH?

☑ **MATCH**

☒ **NO MATCH** → **REWRITE THE STEM/OPTIONS**

STEP 5 **5. Review the items**

IS THERE ONLY A SINGLE "BEST" ANSWER FOR EACH STEM ☐ **YES** ☐ **NO**

AT LEAST FOUR REASONABLE DISTRACTORS FOR EACH STEM ☐ **YES** ☐ **NO**

TEST 1- Check for technical item flaws that provide special benefit to testwise examinees or that pose irrelevant difficulty

TECHNICAL ITEM FLAWS:

- ☒ Sets without lead-ins
- ☒ Non-specific lead-ins e.g. "Match each item with the best option…"
- ☒ Grammatical cues or inconsistency

TEST 2- Peer review

CAN A PEER DETERMINE THE CORRECT ANSWER?

☑ **YES** → **A SUCCESSFUL EMQ!**

☒ **NO** → **MODIFY OPTIONS OR ITEM TO ELIMINATE AMBIGUITY**

Short Answer Questions

Reference: **Jolly, B** 2014, *Written Assessment* in Swanwick, T. (ed) *Understanding Medical Education: Evidence, Theory and Practice,* Second Edition, pp. 293-305

Brief introduction to writing an SAQ

Short answer questions are used to ascertain factual knowledge or understanding. Various forms exist including completion of sentences, supplying a missing line, giving a short descriptive or analytical answer or annotating diagrams. The response therefore is varied; from one or a few words, a paragraph to extensive writing. Advantages of their use include that they:

- avoid cueing as examinees have to construct an answer
- are easier to mark than essay questions
- versatile in usage

To start writing:

1. Read through the SAQ Writing Guidelines – Five Tests.
2. Use the SAQ Worksheet in this pack to practice.

When you are familiar with the process use the SAQ Template available on the ACEM website (example template also included within this pack) alongside the writing guidelines.

Examples

SAQ Example One

A 3 year old child presents with their parents who are concerned that she has aspirated a toy.

The child is not distressed.

Vital signs are:

HR	100	beats/min (regular)
BP	95/70	mmHg
O_2 Saturation	98	% on air
Temperature	37.3	^oC

a. What are three indications for bronchoscopy in a child with a possible inhaled foreign body? (3 marks)

Answer: historical choking / coughing or persistent cough post choke, unilateral wheeze or atelectasis, or hyperinflation on expiration on X-ray

b. In acute total upper airway obstruction in a child, list the initial treatment steps in sequential order?
(4 marks)

Answer: BLS with back blows and chest thrusts, attempted ventilation with BVM, direct visualization possibly attempted removal, ETT to advance obstruction,)

c. What features in history are most relevant when assessing a child for possible foreign body aspiration?
(3 marks)

Answer: Age, opportunity, cough/ choking, sudden onset, no prodrome

SAQ Example Two

You are the duty Consultant in a northern Australian emergency department during the summer months. You receive a 35 year old female surfer who has been being "dragged" from the water and brought in by car. She is extremely distressed by leg pain.

Vital signs are:

HR	140	beats/min (regular)
BP	150/90	mmHg
RR	26	/min
O_2Saturation	92	% on air
GCS	15	

A photo of her right leg is shown.

1. What is your initial interpretation of the wounds? (2 marks)

Answer: Extensive Chironex fleckeri (Box jelly fish) stings to the leg with potential for systemic envenomation

2. List three possible causes of her hypoxia (3 marks)

Answer: Possible near drowning, Pulmonary oedema post envenomation, chest injury sustained in the water > contusion/PTx, (medical cause such as asthma)

3. What is the recommended initial treatment of the local leg injury? (3 marks)

Answer: Vinegar should be liberally applied if not already done so. Analgesia: IV narcotic, large doses often required. Remove tentacles with gloves if any remain attached.

What would be the indications for anti-venom in this case? (2 marks)

Answer: If patient develops Cardiovascular instability/cardiac arrest (usually occurs soon after sting). Ongoing severe local pain not controlled by IV narcotics

SAQ Example Three

A member of the police force presents to your urban district hospital emergency department after being stabbed in the shoulder with a freshly used hypodermic needle whilst working. Assessment of the patient shows no acute wound repair is needed.

1. What is the approximate relative risk in this setting of the infection transmission of the following?
   (a) HIV
   (b) HCV
   (c) HBV

2. List 4 key issues that should be addressed in the assessment of this patient.

Answer: Patient anxiety, infection counselling, follow up with appropriate work services, confidentiality.

3. 3. List 3 blood tests that are essential to aid in the assessment of this patient.

	Blood Test	Clinical Meaning
1.		
2.		
3.		

4. What 4 specific discharge counselling topics should be provided to this patient?

Answer: Follow up results, further serological testing, referral to appropriate services, explanation of infection risk in layman's terms.

5. What are 3 ethical and legal principles involved when giving discharge counselling?

Answer: Confidential labelling of specimens, notification of appropriate work authorities, work cover certificate)

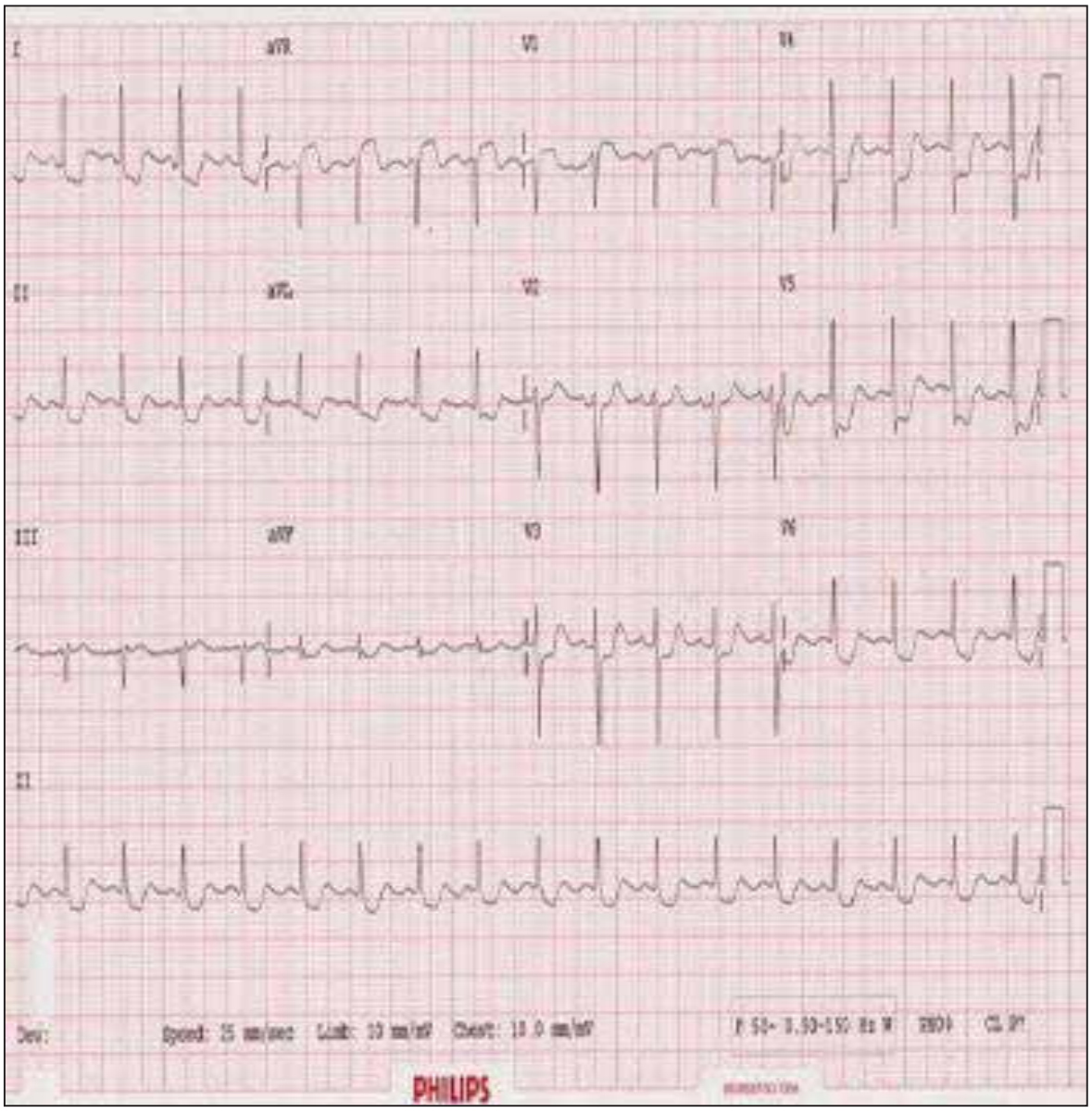
SAQ Example Four

A 50 year old man presents to the emergency department with severe crushing chest pain of 5 hours duration.

Vital signs are:

Temperature	37.4	^{o}C
BP	70/50	mmHg
RR	20	/min
O_2Saturation	91	% on air
GCS	15	
Weight	76	Kg

The following ECG is obtained:



1. What is the diagnosis? (1 mark)

Answer: Acute STEMI

2. What is the likely cardiac pathology? (1 mark)

Answer: Likely proximal LAD or left main artery occlusion

3. What ECG changes support the diagnosis? (1 mark)

Answer: ST elevation AVR and V1, with marked widespread ST depression lead I, II and antero lateral

4. List and justify your immediate emergency department management priorities for this patient (excluding investigations). Include any drug doses (7 marks)

Management	Justification	Dose

17

Answer:

Management	*Justification*	*Dose*
Oxygen by mask	*Currently hypoxic, ? LVF*	*To have O_2 sat >95%*
Aspirin	*Acute STEMI*	*150-300mg*
Pain control: Opiates	*BP a problem currently – may need fluids first, nitrates not indicated with hypotension*	*Titrated to BP and pain* *1-2mg morphine aliquots* *or* *25 – 50 mcg Fentanyl aliquots*
IV fluid bolus	*Evidence of circulatory compromise, and likely RV involvement*	*500ml N Saline then review*
Thrombosis prevention	*Prevent reocclusion*	*Clopidogrel 600mg,or similar* *Heparin 5000u bolus then approx. 1000u/hr*
Urgent cardiology notification for definitive management.	*Urgent angiography/PCI*	
Inotropes likely required	*cardiogenic shock /BP support until definitive Rx*	*Reasonable pressor/ inotrope Rx option*

Further reading

- Appendix 1 – **Short Answer Question** from Jolly, B. 2014, *Written Assessment* in Swanwick, T. (ed) *Understanding Medical Education: Evidence, Theory and Practice*, Second Edition, pp. 263-264. (Full chapter available on request)
- Appendix 2 – **Types of written assessments and their primary usages** from Jolly, B. 2014, *Written Assessment* in Swanwick, T. (ed) *Understanding Medical Education: Evidence, Theory and Practice*, Second Edition, pp. 258-260. (Full chapter available on request)
- Case, D. & Swanson, S. 2002 *Constructing Written Questions for the Basic and Clinical Sciences,* Third Edition.

SAQ Writing Guidelines

QUESTIONS

1. **Each question should be clinically relevant and important to assess**

    - Can be answered realistically in a few words or phrases – avoid essays
    - Visual aids (such as X-rays, ECGs, photos, etc.) can enhance the relevance of the questions – make sure they contribute to the assessment.

2. **Each question should be clear and unambiguous**

    - Should have one focus and one interpretation
    - Restrict the length of the answer by using precise wording to define the task

        - Ask direct questions: "What is . . . "
        - Use actions verbs such as: "List", "Name"
        - Do not use words like "Discuss", "Describe", "Outline" unless you limit the length of the answer or specify a limiting restriction such as "Describe THREE specific methods of …"

    - Length of vignettes should be proportional to the value of the question.
    - Do not provide cues in text.

ANSWERS

1. **Simple and detailed marking scheme is required.**

    - Ensures the questions can be marked objectively and consistently
    - Marking scheme should be reviewed by others before implementation of question to ensure validity

2. **Model answers should be as comprehensive and clear as possible**

    - Ensures little subjectivity by marker
      Clearly outlines the expectations for:

        - Content expected
        - Length of answer
        - Mark allocation

Short Answer Questions - Worksheet

Reference: **Jolly, B**. 2014 *Written Assessment* in Swanwick, T. (ed) *Understanding Medical Education: Evidence, Theory and Practice,* Second Edition, pp. 293-305

PE/FE/CPD: | **MARK ALLOCATION:**

QUESTION ITEM STEM

MARKING SCHEME

MODEL ANSWER

STEP 1 **1. Select the specific learning objectives**

OBJECTIVE: ☐ **FACTUAL RECALL** ☐ **COMPREHENSION** ☐ **APPLICATION** ☐ **ANALYSIS**

STEP 2 **2. Select the most appropriate SAQ format for the objective**

FORMAT: ☐ **COMPLETION ITEM** ☐ **OPEN-ONE WORD PHRASE OR ANSWER** ☐ **SERIES OF ANSWERS OR SHORT PARAGRAPH**

STEP 3 **3. Write a clear and unambiguous stem**

STEM:

STEMS:

- ☑ CLINICALLY RELEVANT
- ☑ IMPORTANT TO ASSESS
- ☑ ONE FOCUS AND ONE INTERPRETATION
- ☑ LENGTH OF VIGNETTE PROPORTIONATE TO QUESTION VALUE
- ☑ "What is…"
- ☑ ACTION VERBS: "List", "Name"
- ☑ CAN USE VISUAL AIDS
- ☒ GRAMMATICAL CUES
- ☒ "Discuss", "Describe", "Outline" unless with limiting restriction:
- ☑ Describe THREE specific methods of…

TEST- Is the stem focused and length of the answer limited?

IS THE WORDING PRECISE TO DEFINE THE TASK?

☑ **YES**

☒ **NO** → Amend the stem

WRITE YOUR OWN ANSWER TO THIS STEM. IS YOUR ANSWER LIMITED IN LENGTH?

☑ **YES**

☒ **NO** → Amend the stem

STEP 4 **4. Provide a simple and detailed marking scheme**

MARKING SCHEME: ☑ **ENSURES OBJECTIVE AND CONSISTENT MARKING**

TEST- PEER REVIEW

CAN A PEER DETERMINE THE CORRECT ANSWER?

☑ **YES** → **A SUCCESSFUL MARKING SCHEME**

☒ **NO** → **MODIFY STEM OR MARKING SCHEME IF APPROPRIATE**

STEP 5 **5. Write a model answer**

MODEL ANSWER: ☑ **COMPREHENSIVE AND CLEAR** ☒ **MARKER SUBJECTIVITY**

TEST- ARE EXPECTATIONS CLEARLY OUTLINED?

MODEL ANSWER OUTLINES EXPECTATIONS FOR:

- ☑ CONTENT EXPECTED
- ☑ LENGTH OF ANSWER
- ☑ MARK ALLOCATION

Objective Structured Clinical Examinations

Reference: **Boursicot, K., Roberts ,T. & Burdick, W. 2014** *Structured assessments of clinical competence* in Swanwick, T. (ed) *Understanding Medical Education: Evidence, Theory and Practice*, Second edition, pp. 293-305

Brief introduction to writing OSCEs

Objective structured clinical examinations are becoming an increasingly utilised form of assessment in clinical competence. They are blueprinted to the learning outcomes of the curriculum and involve rotating around a series of structured cases and undertaking specific tasks usually involving a clinical skill e.g. history taking or examination.

An advantage of their use is that OSCEs are a fair and reliable method of assessing clinical skills.

To start writing:

1. Read through the OSCE Writing Guidelines – Five Tests
2. Use the OSCE Worksheet in this pack to practice

When you are familiar with the process use the OSCE Template available on the ACEM website (example template also included within this pack) alongside the Writing Guidelines.

Example 1

SUBJECT AND CURRICULUM REFERENCE

Difficult airway management

Medical Expertise

Teamwork and Collaboration

Prioritisation and Decision Making

CLINICAL SCENARIO STEM

A 50 year old man is brought in by ambulance with an IV in situ. The patient has been assaulted with a cricket bat. He has isolated head and face injuries. On arrival his vital signs are as follows:

HR	90	beats/min	
RR	8	/min	
BP	150/80	/mmHg	
O_2 Saturation	90	%	15L oxygen via non rebreather mask.
GCS	4	with equal and reactive pupils	

On examination he has obvious extensive midface fractures bilaterally, with blood coming out of his mouth, and gurgling respirations.

INSTRUCTIONS

Candidate:

The scenario is in the Resuscitation Room. There is a high fidelity mannequin that will respond as a live patient. Your registrar has already prepared the appropriate drugs and difficult airway trolley.
You have an assistant who is a competent nurse, but requires instruction. The registrar has been called away. You have been called by your registrar for an anticipated difficult airway and are required to manage the patient's airway.

Role player - nurse assistant:

You are an experienced airway nurse. You will follow the candidate's instructions promptly, efficiently and competently. You will not prompt him with regards to patient management. You will alert him to significant deterioration in the patient's condition (e.g. when the oxygen saturations deteriorate, or if the patient becomes bradycardic).

Examiner:

This scenario requires an advanced mannequin, difficult airway and resuscitation equipment and a competent nurse to assist the candidate.

The mannequin is impossible to intubate.

The candidate must first recognise the need to secure the airway and attempt intubation using RSI.
Once intubation fails, the candidate must commence an appropriate difficult airway algorithm (e.g. reattempt intubation with some changes such as repositioning head, use of a bougie, etc., assistant providing laryngeal manipulation, etc.), then LMA insertion, BVM ventilation.

If the candidate inserts an oral airway and ventilates competently via bag-valve-mask, he is initially able to oxygenate and ventilate the person for approximately 30 seconds. However over that period of time, the patient's oxygen saturations deteriorate markedly. The candidate must now recognise a "can't intubate, can't oxygenate" scenario and proceed to a surgical airway.

NOTE: If by 5 minutes the candidate has not proceeded to a surgical airway, at that point the patient becomes peri-arrest: HR 20 /min, apnoeaic, O2 Saturation 60%.

Assessment criteria

- Recognition of need to secure airway and intubate patient.
- Attempt intubation with appropriate drugs, equipment and technique.
- Once first intubation attempt fails, commence difficult airway algorithm.
- Maximum of 3 attempts at intubation, and at each attempt must alter some factor to increase likelihood of intubation (e.g. Reposition patient, or use of a bougie, or having assistance provide external laryngeal manipulation)
- LMA insertion – must have competent technique and insert successfully, however is unable to ventilate patient adequately via LMA.
- Maintain oxygenation between attempts at intubation/LMA by competent ventilation with bag-valve-mask
- NOTE: After at least 3 attempts of establishing a definitive airway (ETT and/or LMA) candidate is able to ventilate and oxygenate patient via BVM (if performed competently). However after approx. 30 seconds, situation deteriorates into a "can't intubate can't oxygenate" scenario.
- Candidate must recognise and verbalise that this is a "can't intubate, can't oxygenate" scenario.
- Candidate must successfully establish a surgical airway.
- Candidate must successfully ventilate patient via the surgical airway, using an appropriate technique.

Example 2

SUBJECT AND CURRICULUM REFERENCE

History taking - diarrhoea

Medical Expertise

Communication

CLINICAL SCENARIO STEM

A 25 year old previously healthy man presents to the emergency department complaining of 2 weeks of diarrhoea.

INSTRUCTIONS

Candidate:

A 25 year old previously healthy man presents to the emergency department complaining of 2 weeks of diarrhoea. Take a history. After 6 minutes you will be asked to summarize the findings and give a differential diagnosis.

You will not be required to examine the patient. Vital sign are normal. The abdomen is soft, non-tender with no distension.

Role player - patient:

You are 25 and called Michael Jones. You started to notice that your stools were more runny than normal two weeks ago, and since then you have been opening your bowels with increasing frequency up to 10 times per day. This is very unusual for you. For the last five days you have noticed blood in the stool, and this morning the toilet pan was filled with blood, which alarmed you and made you come to the emergency department. You have no other previous visits to doctors or the hospital. You are not short of breath. If asked, say that you have felt more tired than usual recently. No nausea or vomiting and no shortness of breath. If asked about pain, say that you have had one or two twinges of pain in your stomach during attacks of the diarrhoea but otherwise not. You drink 5 pints of beer most Saturday nights, do not smoke and have no allergies. You have two sisters who are both healthy and your parents are both healthy and in their 50s. Your appetite is usually good but you have been less hungry than normal for the past two weeks. Your jeans are looser than normal. You have had no recent overseas travel.

Examiner:

Observation only. After 6 minutes ask the candidate to give a differential diagnosis.

Assessment criteria

- Confirms that reason for attendance is diarrhoea.
- Duration of symptoms.
- Amount and frequency of stools.
- Any blood – alone or mixed with stool – bright red or dark.
- Nausea, vomiting, haematemesis (coffee grounds), abdominal pain.
- Weight loss? Fever? Lethargy? Dizziness.
- PMH? Meds (NSAIDS, anticoagulants, ASA) Family history? Diet. Travel history. Joint problems, eye problems, skin problems?
- Candidate should include inflammatory bowel disease in his differential diagnosis.

Example 3

SUBJECT AND CURRICULUM REFERENCE

Asthma management

Leadership and Management

Medical Expertise

CLINICAL SCENARIO STEM

You are working in a rural district hospital. You have immediately available an experienced emergency department nurse and emergency department registrar.

A 26 year old man is brought into your emergency department having a severe asthma attack. Initial arterial gas on high flow oxygen is as follows:

pH	7.10
pO_2	54
pCO_2	120
$HCO3^-$	18
Base	21.0
Na	135
K	4.6
Cl	99
Gluc	6.4
Lactate	4

INSTRUCTIONS

Candidate:

You are required to describe the blood gas result to the staff and commence therapy for this patient with severe asthma. You may assume that all drugs and equipment usually available in a Resuscitation Room are available. You will need to provide detailed instructions to the nurse and registrar including drug dose and administration route.

You will have a high fidelity resuscitation mannequin and two staff members; an emergency department registrar and an emergency department nurse who will assist by carrying out your instructions but will not initiate any treatment unless requested. They may alert you to major changes in the patient's clinical status.

Role player – ED nurse:

You are an experienced ED nurse and will follow instructions as given by the candidate promptly and efficiently. You are able to note any significant changes to the clinical status of the patient. You are not to initiate therapy without being asked, nor prompt regarding patient management. You are able to seek clarification if any instruction is unclear. If asked to administer a drug you will need to be provided with the dose and route of administration. If asked to ventilate or connect to a ventilator you will expect to be provided with clear instruction, and can prompt the candidate if required.

Role player – ED registrar:

You will follow instructions as given by the candidate. You are able to note any significant changes to the clinical status of the patient. You are not to initiate therapy without being asked, nor prompt regarding management. You are able to seek clarification if any instruction is unclear. If asked to administer a drug you will need to be provided with the dose and route of administration. If asked to ventilate or connect to a ventilator you will expect to be provided with clear instruction, and can prompt the candidate if required.

Examiner:

The candidate should efficiently describe the blood gas result to the staff members and then immediately proceed with resuscitation.

The candidate should assume the team leader role, and communicate with and utilise team members appropriately.

The candidate should commence reasonable therapy for severe asthma including intravenous drugs with doses and prepare for immediate intubation and ventilation.

Once intubated, the candidate should aim for adequate oxygenation and accept hypercarbia, with prolonged expiratory time.

Assessment criteria

- Accurately identify severe acute respiratory failure.
- Manage drug therapy for severe asthma appropriately.
- Identify need for intubation and proceed to intubation using appropriate drugs for RSI.
- Intubation technique.
- Appropriate post intubation checks and ventilation goals in a patient with severe asthma.

See Appendix 4 for Full Set of OSCE Stations

Further reading

Appendix 3 – **Structured assessments of clinical competence** from Boursicot K., Roberts T., & Burdick W. 2014 in Swanwick, T. (ed) *Understanding Medical Education: Evidence, Theory and Practice*, Second edition, pp. 293-305. Full chapter available on request.

Newble, D. 2004 *Techniques for measuring clinical competence: objective structured clinical examinations.* Medical Education.

OSCE Writing Guidelines

Newble D. 2004 *Techniques for measuring clinical competence: objective structured clinical examinations.* Medical Education.

Define the purpose of the station

STEP 1

- Ensure focus on an authentic clinical problem that is sufficiently realistic.
- Avoid assessing competencies that are not included in the OSCE blueprint.

Develop the clinical scenario

STEP 2

- Can use the stem to frame a task and make it appropriate for the given time limit
  - if assessing ability, frame the stem with relevant patient information.
  - if assessing judgement as well as ability, consider limited patient information in the stem.
- May use a standardised patient.
- May use written of oral questions presented to a candidate.

Develop the station instructions

STEP 3

Candidate instructions

- Specify venue and other circumstances e.g. have they met the patient before.
- Provide explicit instructions about the task to be performed (e.g. take history, examine, explain etc).
- Include any other information required e.g. data.

Examiner instructions

- Provide explicit instructions to examiner to enable them to run the station effectively
  - if prompting will detract from candidate's performance, ask examiner not to prompt
  - if a particular task requires examiner intervention, state what he/she should do
- Give guidance on timing for the station.

Patient instructions

- Provide background information for the patient (where relevant) including name, age, employment, social circumstances, gender, ethnicity, behavioural characteristics, presenting history, past medical history, family history, understanding of illness, attitude to illness.
- Specify answers or questions to ask during the scenario.

CHECK for congruence between:

- the candidate's instructions
- the examiner's instructions
- the simulated patient role

Develop Scoring Sheet for Station

STEP 4

- List the specific items which are important in the performance of the task. Check the items correspond with what is being assessed.
- Be realistic about minutiae (is this really necessary?) and is time available.
- If prompting/questioning required by the examiner, indicate where this should occur.
- Ensure the scoring sheet and clinical scenario stem are compatible.

Objective Structured Clinical Examinations- Worksheet

Reference: **Boursicot K., Roberts T., and Burdick W.** *Structured assessments of clinical competence* in Swanwick, T. (ed) *Understanding Medical Education: Evidence, Theory and Practice*, Second edition, 2014: pp. 293-305

FE/CPD:

PURPOSE OF STATION:

CLINICAL SCENARIO STEM

CANDIDATE INSTRUCTIONS

EXAMINER INSTRUCTIONS

PATIENT INSTRUCTIONS

SCORING SHEET

•	•
•	•
•	•
•	•

STEP 1 **1. Define the purpose of the station**

PURPOSE: ☑ **AUTHENTIC CLINICAL PROBLEM** ☑ **REALISTIC** ☑ **COMPETENCIES INCLUDED IN OSCE BLUEPRINT**

This station tests…

STEP 2 **2. Develop the clinical scenario**

STEM: ☑ **FRAMES THE TASK** ☑ **MAY USE STANDARDISED PATIENT** ☑ **WRITTEN QUESTIONS** ☑ **ORAL QUESTIONS**

TEST- IS THE STEM APPROPRIATE FOR THE GIVEN TIME LIMIT?

IF ASSESSING ABILITY → STEM CONTAINS RELEVANT PATIENT INFO? ☑ YES / ☒ NO → **AMEND STEM**

IF ASSESSING JUDGEMENT AND ABILITY → IS PATIENT INFORMATION SUCCESSFULLY LIMITED? ☑ YES / ☒ NO → **AMEND STEM**

STEP 3 **3. Develop the station instructions**

CANDIDATES: ☑ **VENUE AND CIRCUMSTANCES** ☑ **EXPLICIT TASK INSTRUCTIONS** ☑ **INCLUDE OTHER INFO REQUIRED**

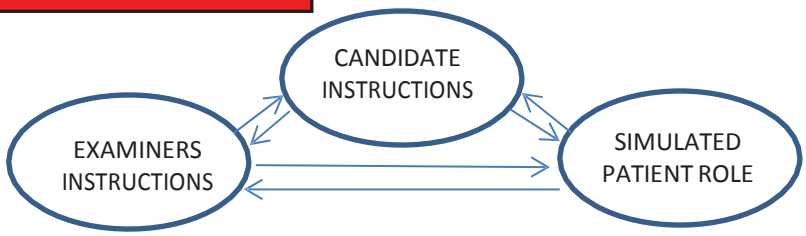
EXAMINERS: ☑ **INSTRUCTIONS EXPLICIT** ☑ **DETAILS PROMPT/NO PROMPT/INTERVENTION** ☑ **INCLUDE TIMING INFO**

PATIENTS: ☑ **BACKGROUND INFO** ☑ **SPECIFY QUESTION/ANSWERS DURING SCENARIO:**

Include relevant background info such as:

- ☑ Name
- ☑ Age
- ☑ Employment
- ☑ Social circumstances
- ☑ Gender
- ☑ Ethnicity
- ☑ Behavioural characteristics
- ☑ Presenting history, past medical history, family history
- ☑ Understanding of illness
- ☑ Attitude to illness

TEST- IS THERE CONGRUENCE BETWEEN:



STEP 4 **4. Develop the Scoring Sheet**

SCORING SHEET: ☑ **ITEMS CORRESPOND TO WHAT IS BEING ASSESSED** ☑ **REALISTIC MINUTIAE AND TIME**

List specific items important to task performance:

TEST- COMPATIBILITY BETWEEN:

APPENDIX 1

In this example, computer delivery, or physical removal of the answers, first, to Question 1, and then 2, before giving the information about anaemia and asking candidates Questions 3 and 4, would be required to avoid both backward and forward cueing. Q1 and Q2 test broad knowledge of such clinical presentations and initial diagnostic strategy. They require understanding of the clinical significance of the scenario. Question 3 tests linkage between data from investigation (that may not have been initially considered by a test taker) and subsequent questioning. Question 4 is vague and open to misinterpretation – for example, Question 1 asks for three likely diagnoses – which one does question 4 refer to? Is the test taker supposed to assume certain positive or negative outcomes from their examination of signs in Question 3? What does 'information' mean in Q4? What does 'affect' mean? What is the designer's rationale for asking Q3 after the delivery of the information about the blood test? Would this information be better after Q3?

Where's the evidence for MEQs?

Psychometric studies done on the MEQ in the 1980s showed that reliabilities ranged between 0.43 and 0.90 (Cronbach's alpha) for a 60-item test, depending on the content area.(38) However, one study suggested that over 50 per cent of MEQ items in a general test for undergraduates in medicine and surgery tested nothing more than factual recall.(39) This contrasts with the rationale for MEQs that emphasises their ability to reflect analysis, interpretation and clinical decision-making. A more recent study in the same institution has resulted in the removal of the MEQ from the undergraduate assessment programme.(40)

Short-answer question

Many educators use short-answer questions (SAQs) in some form. Frequently, they are used as means of gauging students' factual knowledge or understanding – for example, in lectures and ward rounds. In the verbal form they tend to be quite short, asking for one word or a few alternative answers, within a specific context, as in the following example:

- What is the most common feature of diabetic retinopathy we are likely to see in this patient?

The other major use of short-answer questions is in assessments. Various forms exist requiring the test-taker to complete the sentence or supply a missing line (a 'cloze test'), give short descriptive or analytical answers, or annotate diagrams. Such questions can demand a wide range of responses, from one or several words, a paragraph, to more than a page. The different forms of SAQ provide for great versatility in usage, but make classification difficult. An individual question can be used to assess a specific objective and unlike multiple-choice questions, short-answer questions have the advantage of requiring students to construct an answer, rather than choosing (or guessing) from provided options, so avoiding cueing (at least when SAQs are used sparingly).

SAQs are easier to mark than essay questions and usually involve a structured marking sheet that indicates all possible answers, and ones that should or should not get credit. Marking sheets should also indicate whether spelling needs to be perfect or which common misspellings are acceptable. One-word answers are computer scoreable. Currently programs are being developed for scoring that involves longer answers.(41)

Items should be marked with assessors blind to the identity of candidates and different markers allocated to different questions or sets of questions. In this way, examiner bias is diluted for each candidate. Some assessors report that having the marking done at one time in a large room with all examiners able to talk to each other as unexpected responses are discovered is beneficial to efficient and equitable scoring. Test designers need to be prepared to accept answers not on the score sheet, some of which may or may not have been predicted. There will need to be a system for referring these to the test convenor or committee – do not allow discretion at the marker level, some markers may be unable to make this judgement.

Marking poses the major difficulties with this form of assessment, although there is variability between markers on most constructed response types of question. Increasing the number of markers and number of questions can ameliorate the problem, but is frequently impractical.(42,43) Many educators allege that SAQs reduce the likelihood that students will look for the relations between objectives or sections of the subject whilst studying, and that complex issues cannot always be satisfactorily addressed in short answers. However, there is little empirical evidence for these assertions.

Constructing an SAQ

- Identify the specific learning objectives the item will cover. These are generally in the area of factual recall, comprehension, application or analysis. Higher levels such as evaluation or synthesis will probably require a longer test format, such as a modified essay.
- Choose the most appropriate SAQ format for the objective – a cloze or completion item, an open one-word or phrase answer, a series of answers or a question that requires a short paragraph.
- State the item concisely in clear, unambiguous simple language. A good SAQ tests factual knowledge or capacity to analyse and clinically interpret a scenario. Introducing an element of the test taker's ability to make sense of the question introduces construct-irrelevant variance into the assessment.
- Look at the draft item from a number of different perspectives – mentally try out adequate and inadequate responses. Ideally, an item aimed at one fact should have just one answer, and one aimed at alternatives (e.g. differential diagnoses) should have as many as are appropriate. However, what you may think of as a clear, straightforward question may frequently be answered in multiple ways, depending on how the reader reads it.
- It is good practice to give the test taker an indication of the length of answer required.
- Indicate how many marks are available for the question.
- Some research suggests that items asking for positive perspectives (e.g. knowing the best method, describing

good practice or identifying the most relevant facts) have greater educational significance (e.g. in terms of capacity to measure objectives) than knowing the poorest method, unsatisfactory practice or the least relevant issues. However, clinical science sometimes depends on the capacity to rule out rare or unlikely occurrences, so research done in general educational settings may not always apply in the health context. If you have to word an item negatively, it is advisable to use some form of emphasis for the negative words (e.g. 'What is *not* an appropriate management option in this situation?'), using italics, bold type or underlining to stress this for the test taker.

- Try to avoid grammatical cues to the answer or providing answer spaces that are equal or proportional to the lengths of the required responses.
- Where a numerical answer has to be supplied, for example from a calculation based on clinical data, indicate that both;
  a) the degree of precision expected (e.g. give your answer to one decimal place and answers within 5% of the correct value will be given credit) and
  b) the appropriate units must be indicated.

Not doing this will result in uncertainty for markers about whether the answers supplied are acceptable or not.

Where's the evidence for SAQs?

There is very little research on short-answer questions, particularly in medicine. However, there is some evidence from secondary education that constructed response short-answer questions measure exactly the same thing as MCQ items, as long as the stems are the same.(44) In other words, the cognitive task set to the test taker is more important than the response format. However, once the task diverges, even in the same content domain, the correlation between the two forms falls off. Also, SAQs are more reliably scored than essays,(45,46) largely because the pitfalls of scoring lengthy answers are avoided, and because SAQs can sample more widely in a given time. In addition, using SAQs may reduce the reported differences between men and women, and black and white racial groups on propensity to omit items in MCQ tests.(47) In medicine SAQs have successfully been used as a reliable alternative to MCQ items in a progress test in the Netherlands.(48) One study showed that SAQ tests could produce better retention of information over time, as long as the delayed test was a short-answer test. There was no difference between groups if the test was an MCQ test.(49)

Selected response formats

Multiple-choice questions: Multiple true/false formats

Multiple-choice testing was once seen as an enduring solution to the reliable and valid measurement of knowledge in 'knowledge rich' or knowledge-dependent environments such as medicine, bioscience and engineering. Invented in 1914 by Frederick Kelly, head of a training school in Kansas, USA, by 1926 the multiple-choice test had become the rite of passage for entering post-secondary education in the USA. The MCQ was developed into several forms. One of these is the multiple true/false item, called an 'X type' item in North America, which has become a significant feature of assessment of knowledge in medicine and many other professions over the last 50 years.

BOX 19.4 Example of a multiple true/false item

Stem	Options	
The following present as chronic (>3 months) airspace disease on a chest radiograph.	A. Streptococcal pneumonia	T/F
	B. Adult respiratory distress syndrome	T/F
	C. Pulmonary oedema	T/F
	D. Asbestosis	T/F

In essence, an MCQ is a question that proffers several answers from which the correct one or ones must be chosen. In multiple true/false types a set of options, usually four to six, is given of which each can be either true or false, and the candidate is required to indicate which is correct for each option. An example is shown in Box 19.4.

Over the last few years the multiple true/false item has received a good deal of critical attention. Many examining bodies (for example, the National Board of Medical Examiners (NBME) in the USA) have given up using it altogether. The main reasons have been elucidated with a good deal of empirical evidence.(26) In brief, Case and Swanson(26) state that:

- the distinction between 'true' and 'false' is not always clear, and it is not uncommon for subsequent reviewers to alter the answer key
- reviewers rewrite or discard true/false items far more frequently than items written in other formats
- some ambiguities can be clarified, but others cannot
- to avoid ambiguity, item writers are pushed toward assessing recall of an isolated fact, which is not desirable in most testing situations
- application of knowledge, integration, synthesis and judgement questions can better be assessed by one-best-answer questions.

It is also the case that using true/false restricts the choice of answers, as discussed in the NBME guidance,(26) to a sub-set that can best be classified as completely true all of the time or completely false all of the time. For this reason this author will strongly recommend not using this type of item.

Multiple choice questions: Single-best answer

In single-best-answer questions a stem question asks the test taker to choose the one best answer typically from a set of four-to-five options. An example, taken from the NBME guidance,(26) is given in Box 19.5.

MCQ items are usually scored optically or directly by computer. There are standard programs for marking and analysing test data straight from a scanner. The answer 'key' – a line of data containing the correct option for each

BOX 19.5 Example of a one best answer item. (26, p. 38)

Stem	Options
A 65-year-old man has difficulty rising from a seated position and straightening his trunk, but he has no difficulty flexing his leg. Which of the following muscles is most likely to have been injured?	A. Gluteus maximus B. Gluteus minimus C. Hamstrings D. Iliopsoas E. Obturator internus

item – is used in this process and should be double', or even triple-, checked before use. The most common reason for problems at the marking or item analysis stage is a key that contains wrong answers. This may be because the answer has been wrongly transcribed from the item writer's design or (and this not as rare as it should be) because he/she has not provided the best answer.

Most MCQs are scored 1 or 0 for correct or incorrect answers, respectively. Weighting is not necessary for best-answer items; it has very little impact on rankings of students and can reduce reliability. A so-called 'correction for guessing' need not be used.(50)

How to construct a single-best-answer question

Writing multiple-choice items involves following a series of basic rules that, for the most part, apply to all types. A sensible approach to item construction is to have item-writing workshops that force item writers to work in small groups(2,3) and then have their items immediately reviewed by a larger group. During the workshops the rules are as follows and *should* be applied as a test to each of the items that you construct. Each item should pass *all* the rules.

- Focus on an important (non-trivial) concept, typically a common or potentially serious clinical problem. Avoid trivial, 'tricky', or overly complex questions.
- Focus on how knowledge is applied to a clinical situation, not on recognising an isolated fact or association between concept and exemplar.
- The stem must state a clear question, and it should be possible to arrive at an answer with the options covered (the cover test). To determine if the question is clearly focused, cover up the options and read the stem to make sure it is lucid and that other item designers can supply an answer dependent only on reading the stem.
- All distractors and the correct answer should be homogeneous, that is, they should fall into the same category. For example, in an anatomy question all answers should be the same type of structure – bones, vessels, nerves, etc. In a clinical item they should all be diagnoses, tests, treatments, prognoses, and so on.
- All distractors should be salient and plausible. Order the options in numeric, or in alphabetical order (*see* Box 19.5).
- Try to write questions of moderate difficulty – if any of the item constructors have a problem with the item it is probably too difficult. Make sure the correct answer has a sufficiently different degree of correctness, when compared to the distractors across all the conditions identified in the stem. For example, let's assume we are testing knowledge of a condition that affects men, usually in later life. If an incorrect option (distractor) is a diagnosis that does sometimes occur in the age group that the question stem has identified, and the correct answer is one of the rarer diagnoses, the two options may not be far enough apart to make a distinction clear.
- Avoid technical item flaws. For example, all items and options should be grammatically consistent, logically compatible, and of the same (relative) length as the correct answer.
- Writing questions of the form 'Which of the following is correct?' followed by a set of brief, possibly unrelated postulates, one of which is correct, is not advisable. It is basically a true/false item masquerading as a one best answer. Furthermore, it is a good way of ensuring that the questions do focus on trivia. These questions will not be directed at course objectives in a coherent fashion and will contain multiple heterogeneous options.

Where's the evidence for MCQs?

There is far too much research on MCQs to summarise in this chapter. The interested reader should look at recent evidence based guidelines by Wood,(15) Downing(51) and Haladyna *et al.*(52) for comprehensive treatments of many issues. One interesting fact to emerge is that the number of options to use in a one-best-answer item for maximal reliability is more likely to be 2 or 3 than 4 or 5. There is long-standing theoretical and empirical evidence to support this position.(53) This is because this effect is generated when the additional distractors, usually put in to make up or provide a standard number of options are not performing adequately. In items where the 4 or 5 distractors are operating effectively, the item tends to have increased reliability, but this situation is unusual. There is also contradictory evidence that extended matching question distractors, usually a naturally occurring fairly large set of 10–15 (*see* below), may operate more effectively than 3 or 4 pre-selected ones.(26,54) We would hope that the processes that students use to answer one-best-answer MCQ items are at least analytical and at best reasoning-rich. However, evidence suggests that 'the problem with multiple-choice items is not that they are mere exercises in recognition, but that we are unable to predict the processes that will be evoked'.(55, p. S9)

Extended matching questions

Extended matching questions (EMQs) were developed in the early 1990s.(26) However, the kernel of the idea was first conceived by Sue Case in her PhD thesis as early as 1983.(56) She and David Swanson are credited with most of the development work on this format, whilst at the National Board of Medical Examiners. An EMQ is a selected response item in which the item stem has been extended, usually, to a short clinical vignette or scenario and the

BOX 19.6 Example of an extended matching item

Area: Abdominal pain – Diagnosis

Options

A. Abdominal aneurysm	K. Kidney stone
B. Appendicitis	L. Mesenteric adenitis
C. Bowel obstruction	M. Mesenteric artery thrombosis
D. Cholecystitis	N. Ovarian cyst – ruptured
E. Colon cancer	O. Pancreatitis
F. Constipation	P. Pelvic inflammatory disease
G. Diverticulitis	Q. Peptic ulcer disease
H. Ectopic pregnancy – ruptured	R. Perforated peptic ulcer
I. Endometriosis	S. Pyelonephritis
J. Hernia	T. Torsion

Lead In: For each patient with abdominal pain described below, select the most likely diagnosis.

Scenario/Stem: A 25-year-old woman has sudden onset of persistent right lower abdominal pain that is increasing in severity. She has nausea without vomiting. She had a normal bowel movement just before onset of pain. Examination shows exquisite deep tenderness to palpation in right lower abdomen with guarding but no rebound; bowel sounds are present. Pelvic examination shows a 7-cm, exquisitely tender right-sided mass. Hematocrit is 32%. Leukocyte count is 18,000/mm3. Serum amylase activity is within normal limits. Test of the stool for occult blood is negative.

Answer: __

[Next scenario in the domain (diagnosis of abdominal pain) would appear here.]

Source: National Board of Medical Examiners(26)

choices have been extended to include all potentially acceptable ones for the clinical problem or issue that is being addressed by the item. This format was originally targeted towards the application of clinical knowledge to diagnostic and management problems, but has been extended to other areas such as basic science. In the example, in Box 19.6 there are 20 options pertaining to the theme of diagnosis of abdominal pain. This is followed by one or more clinical vignettes. The options are all causes of abdominal pain. It is usual in such items to attempt to make all the questions and options homogenous in this way, so other issues concerning abdominal pain, such as initial management or investigations, are not included. An item should focus on a specific area of clinical cognitive activity that pertains to a specific phase of the clinical process – in this case diagnosis.

EMQs are usually scored 1 for a correct response and 0 for an incorrect one. It is sometimes possible to have more than one best answer – for example, when two (or more) diagnoses are equally likely, given the information in the vignette. However, the scoring of these requires more attention during the scanning process, as for optical scoring two (or more) passes of the score sheets are necessary with each correct answer keyed on each pass. Unless it is clinically important to be able to recognise both potential diagnoses from the same vignette, such multiple responses are probably best avoided (for example, by removing one of the options from the list).

Investigations of the reliability and construct validity of EMQs suggest that they have good measurement properties, and correlate well with other measures of recall, recognition and cognitive functioning.(57) Recently, studies of experts and novices who talk aloud whilst trying to complete EMQ items have strongly suggested that EMQs have good construct validity for, and can be reliably used to assess, clinical reasoning,(58) even though novice and experts approach the same item with different strategies (backwards vs. forwards reasoning, respectively).(59) Furthermore, when used in pathology EMQs are more reliable, better discriminate the competent from the borderline student, and can be written definitively to test core content.(60)

EMQs seem to be easier to write than true/false or other types of one-best-answer items, because in that style of item the convolutions that writers go through to reduce the item set to a smaller number where each is true or false, or there is clearly one best answer, are not needed.(61) Clinicians from some disciplines such as public health, epidemiology and statistics have suggested that EMQs are difficult to write for these content domains, but recent articles suggest they have been adopted or are being developed in some of these hitherto unexplored areas (e.g. psychiatry(62)).

How to construct an EMQ

It is best to write these items first by considering the area or domain of the assessment blueprint for which items need to be written (e.g. abdominal pain in Box 19.6). Then a general question is posed, followed by all the possible answers to that question (e.g. 'What are the causes of abdominal pain in adults?'). After these have been identified, scenarios that pertain to one or more of the answers are constructed. Ideally, create items (particularly the stems/scenarios) in pairs of writers, at workshops of about 8–12 people in total (4–6 pairs), with review every two hours or so in a larger group. This is an effective, and in most examiners' experience, an enjoyable way of generating items. The stages are as follows.

- ***Identify the domain or subject for the set.*** The domain is an area of cognitive activity (e.g. diagnosis, management planning). The subject can be a presenting complaint in a body system or systems (e.g. abdominal pain, so that diagnosis is the focus of the item), or a pre-diagnosed condition (e.g. community-acquired pneumonia, so that management is the focus of the item). Sometimes it might be appropriate to move directly from a non-diagnosed presenting complaint (e.g. abdominal pain) to an investigative option (e.g. ultrasound) or management plan (e.g. restricted diet). However, the more cognitive steps involved in moving between the presenting complaint and the focus of the item (e.g. asking about management), the less will be known about why an

examinee might have answered the item incorrectly. For example, the examinee might have thought a patient with ulcerative colitis had appendicitis and ordered surgical intervention.

- ***Write the lead-in for the set*** (e.g. *for each patient described below, select the most likely diagnosis* (Box 19.6)). The lead-in indicates the relationship between the stems and options. It must be a clear question for examinees. It is an essential component of an extended-matching set. Sometimes two lead-ins can be written at the same time – for example, one based on diagnosis and one on indications for investigations or management. Subsequent scenarios can be used, usually with only minor modification, with either lead-in. In summary, the lead-in should consist of a single, clearly formulated task so that the examinee could, if necessary, create an answer without looking at the options.
- ***Prepare the list of options.*** The list of options should be single words or very short phrases. This list is best developed in a whole-group format. It will be generated in a fairly random order, but the options should be rearranged in alphabetical order for the final item presentation. For example, the initial list for Box 19.6 should contain all the likely causes of abdominal pain as options. Sometimes there are specific causes that occur only or predominantly in a particular subset of the population – for example, in women (e.g. ectopic pregnancy), in men (testicular torsion), in the elderly (dementia). Such options can sometimes become 'zebras',(26) which stand out as so obviously applying to one subgroup of patients that their inclusion is ill advised. In Box 19.6 there are some such options, but there are also sufficient important differential diagnoses in the list to warrant their inclusion for the given scenario.
- ***Write the stems.*** The stems (items) within a set should be similar in structure. Most often, patient vignettes are appropriate. The scenario should contain all the information that one would normally expect to be available from any conscious patient: the presenting problem, the history (including duration of signs and symptoms), the physical findings, and then the results of any immediate diagnostic tests carried out. Sometimes, for a complex case, further data pertaining to development of symptoms over several days might also be given – for example, initial management and subsequent clinical changes. Scenarios can include a smaller set of information, but it is unwise to exclude the information that would normally be collected by or available to the test taker in the real clinical context at the time they were seeing this patient. Specifying this information in a standardised order makes shorter reading time and hence allows more items to be delivered in a given time.
- ***Review the items.*** Make sure there is only one 'best' answer for each question. Having two right answers is possible, but entails more marking effort than it is usually worth. Also make sure that there are at least four reasonable distractors for each item to minimise guessing effects. Evaluate the extent to which the lead-in clearly formulated task. See if the other examiners can create an answer without looking at the options. Satisfying this 'cover-the-options' rule is an important feature of a good question because, if the examinee cannot do that, it means the question is too vague, is not appropriately targeted to the skills the examination is testing, or exhibits some other flaw of test writing. As a final check, review the items (without the correct answer indicated) across other pairs in the writing group. If the pair has difficulty determining the correct answer, modify the option list or the item to eliminate the ambiguity.

Where's the evidence for EMQs?

There is evidence that some MCQs, of which the EMQ type seems to be the most suited to clinical tasks, can involve substantially more than recognition of knowledge learnt by rote memorisation.(55) These authors nevertheless suggest that unfocused items or those with negatively worded stems, as sometimes, of necessity, occur in the typical true/false or best of five types, do not appear to provoke problem-solving skills and forward-reasoning. Whilst the evidence for the link between item type and cognitive response is being developed, they suggest concentrating on items that are low-fidelity simulations of clinical situations with examinee tasks that are relevant for them (e.g. diagnosis and management). EMQs are ideally suited to this role. EMQs also substantially reduce the likelihood of obtaining a correct answer by chance alone.

Although this area is fraught with controversy, and not all of the options provided for any one EMQ stem will be active for that item, modelling suggests that EMQs with between 7 and 12 active distractors will provide good insulation against the need to be concerned about the so called 'guessing' factor in multiple-choice tests.(63) Research on EMQ formats has shown that a reduction in the length of item option lists, from the 15–20 previously thought necessary, is possible without much, if any, deleterious effects on item quality.(64,65) Eight options seem to be a reasonable minimum number. In general, items with more options are more difficult, require more time to complete, and nevertheless have similar discriminating properties to items with eight options. Reducing the whole list to a 'shortlist' of eight or so can be done by carefully constructing physician panels to select the most appropriate set. Moreover, providing the panel with item response statistics from the long item does not seem to improve option selection. The use of a smaller number of options reduces time spent on each item by candidates and therefore increases the number of items that can be used in a set time.(64,65) Swanson *et al.* advised, 'We plan to begin advising [examiners] to reduce the number of options included on option lists in order to make more efficient use of testing time'.(64, p. S95) That advice may now be prudent to implement, as student numbers in medical schools have increased dramatically in a number of countries in the last few years.

Script concordance items

Over the last 15 years interest has developed in constructing a multiple-choice test that can reflect clinicians' capacity to weigh evidence in a clinical encounter. This work has its foundations in a clearer understanding of how clinicians approach the diagnostic task and how this information is

BOX 19.1 Types of written assessments and their primary usages. From Epstein(27)

Method	Domain Usage/Response Mode	Design Factors	Limitations	Strengths
Constructed Response Formats				
1a Essay – Traditional. The typical university essay, either seen or unseen, where the writer is required to describe, discuss and propose new perspectives on one or more issues. The answer may or may not be predetermined.	Any situation where lengthy explanation or detail is required. Detailed synthesis of information; interpretation of literature, evaluation of management options. Context frequently provided by the candidate.	Traditionally questions can vary from the blindingly obvious to the very obscure. Large number of dimensions to the constructed response. Getting questions right takes time. Model answers or protocols help marking. High marking workload.	Can be usurped into provision of lists, e.g. for treatments; can become memory dumps. Can be misinterpreted. Long testing time per topic, so limited coverage possible. Reliability variable and susceptible to rater and candidate bias	Total flexibility in question setting. Can avoid cueing. Regarded as using higher-order cognitive processes.
1b Modified Essay – Specifically developed for medicine – mostly used in general practice. Highly structured case vignette followed by questions on any aspect. Focused on candidates' management of a case or cases. Answer(s) usually predetermined.	Clinical management issues. Some cue identification and reasoning required to link, e.g. signs and symptoms to investigations and management. Context provided by the question.	Can move from one stage of clinical management to another easily, by using slightly different cases to address issues, e.g. patient management in one case, and ethics in a similar one. More efficient sampling of a wide area of knowledge possible	Needs careful design to avoid cueing. As a result can be patchy in sampling knowledge across cases.	Can avoid cueing. Context is controllable by question setter. Can demand wide range of cognitive processes. May be machine scoreable in next 5–10 years.
2a Short Answer – Traditional. A short question that asks for a constructed specific answer, usually requiring one word, a short phrase, or a line or two of text. Answers mostly predetermined.	Recall of specific facts or statements about biomedical or clinical processes. Context provided by the question.	Deceptively simple to construct. Can sample widely different domains of knowledge easily.	Very wide variety of formats and little research on their use and psychometric properties. Can lead to cueing across items. Context provided by question.	Scoring by machine becoming a reality. Can replace MCQs where recall is thought to be vital (e.g. decisions based on core knowledge and experience)
2b Short Answer – Extended. A question that asks for an extended answer, usually requiring a paragraph or two, that may address different aspects, or an extension, of the issue. Answers may be predetermined.	Recall of related groups of concepts or relatively short explanations. Context provided by the question.	Deceptively simple to construct. Can sample widely different domains of knowledge but in more depth than short answer.	As above. Scoring more difficult as depends on multiple attributes of answers involved in essay construction. Machine scoring not possible. Context provided by question. Recent research on analysis of answers can give more insight into level of functioning of candidate.	Total flexibility in question setting.

Selected Response Modes				
1a MCQ – True/False. Typically a short statement or brief paragraph followed by several (3–6) options. Candidates are asked to identify which options are true and which ones false in relation to the initial statement.	Recognition of consonance between two facts, attributes or concepts. Can test recognition of clear-cut knowledge in many domains. Complex items requiring calculations or problem-solving have been used.	Requires all options to be absolutely true or false. Can test knowledge of contraindications through the 'false' option.	Difficult to write. The number used in most assessments can lead to cross-cueing. Can involve silly or irrelevant options due to lack of absolute falsehoods. Getting statements into an absolute true/false mode sometimes requires convolutions such as double-negatives. Extreme controversy over 'correction for guessing' as random choice of options results in 50 per cent score. Rapidly waning in popularity	Can test range of knowledge in limited testing time. Machine scoreable. True/False requirement restricts applicability and engenders artificiality.
1b MCQ – 1 from N. Typically a short statement or brief paragraph followed by several (3–6) options. Candidates are asked to identify the option that best fits with or is the best outcome for the initial statement.	Recognition of consonance between two facts, attributes or concepts. Can reflect basic clinical decisions, basic science or hypothesis generation.	Easier to write than MCQ T/F. Choosing one best answer is more salient to most areas of medicine.	The number used in most assessments can lead to cross cuing. Need not involve a correction for guessing.	Efficient sampling of knowledge. Allow more subtle distinctions than T/F types. Machine scoreable.
1c MCO – Extended matching. Typically a topic area (e.g. Headache), followed by many (6–26) options homogenous to a clinical grouping (e.g. diagnosis). There is a linked question asking candidates to choose the most likely diagnosis. Then one or more paragraphs each comprising a clinical case vignette, including e.g. headache presentation at various stages of progression each of which may indicate different 'best' diagnoses.	Recognition of consonance between (typically) clinical presentations and their underlying pathology, investigation and outcome; diagnoses, prognoses, tests, pharmacology etc Items appear to involve basic clinical reasoning. Students report fidelity to 'real' medicine.	*Relatively easy to generate first* drafts. Salient to most areas of medicine that depend on a clinical context.	Not easy to write in some areas of medicine, especially non-clinical ones, e.g. epidemiology. Some argue that the 'extended' list of options is not as useful as first thought – many options are redundant.	Seem to be more reliable than one-best-answer MCQs and T/F MCQs, due to increased difficulty. No corrections for guessing needed. Good discriminators at higher levels of ability.

(*Continued*)

BOX 19.1 *(Continued)*

Method	Domain Usage/Response Mode	Design Factors	Limitations	Strengths
Selected Response Modes				
2d MCQ – Script Concordance Typically a case vignette followed, for example in items on diagnosis, by statements that give an additional sign or symptom and a question that asks whether a specific diagnosis would be more or less likely if such an attribute were present in the case (*see* example in text) . For example, given a description of a 67-year-old man with chest pain, if pain radiating down the left arm were present, would the likelihood of myocardial infarction be 'strongly increased, ... , strongly decreased'.	Recognition of relationship between, and agreement with an expert group on, attributes of case presentations that are predictive of diagnoses, prognoses, findings on investigation, etc. Appears to involve basic clinical reasoning and personal probabilities.	New type of item, limited experience available of construction. Scoring generated by expert group. May have more than one answer that scores marks. Appears to discriminate effectively between experts and non-experts in some specialties.	Probably limited to diagnostic and prognostic decisions	More research needed, but does show high construct validity for clinical experience. Writing protocols and rules still in development.
Constructed and Selected Response				
1. Short Answer – Key Features. Usually a short case vignette followed by between one and three questions that investigate the taker's knowledge of the main aspects of the case. Answers may be constructed or selected, usually requiring words or short phrases.	Answers that attempt to focus **only** on the critical aspects of clinical cases e.g. key decisions and the factors underpinning those. Developed (1990's) to counter arguments that short answers led to isolated recall of facts and trivialisation. Context provided by the question.	Strict rules for design, done usually by a small team. Items may involve some **selected** responses as well as **constructed** ones. Can explore wide variety of cases. Can match response mode to attributes of the context – e.g. selecting the most important features in clinical investigation results. Shares some properties of Modified Essay Questions.	Scoring and standard-setting can be complex. For example, although single word answers are common, there may be several answers to one question each differentially weighted. There may be totally inappropriate or dangerous answers given by test takers. Can be challenging to avoid cueing between different parts of the item.	Partial scoring by machine is now becoming a reality. Can replace MCQ style questions where recall is thought to be important (e.g. decisions based on core knowledge and experience). Has embedded quickly into assessment technology in medicine.

examiner bias and variations in examiner stringency is a major argument against the long case. Reliability is further compromised when there is little prior agreement between pairs of examiners as to what constitutes acceptable competence. Unstructured questioning and global marking without anchor statements compounds the problem. Reliability in the long case encounter is diminished by variability in degree and details of information disclosure by the patient, as well as variability in patients' demeanour, comfort and health. Furthermore, some patients' illnesses may be straightforward, whereas others may be extremely complex. Examinees' clinical skills also vary significantly across tasks (i.e. task or case specificity),(7) so that assessing examinees on one patient will not provide generalisable estimates of their overall ability.(6,8,9)

While the authenticity of a long-case examination is one of the strengths of the genre, inferring examinees' true clinical skills in the time-constrained environment of actual clinical practice from a 1-hour long-case encounter is debatable. Additionally, given the evidence of the importance of history taking in achieving a diagnosis(10) and the need for students to demonstrate good patient communication skills, the omission of direct observation in this process is a significant weakness.

Objective Structured Long Case Examination Record

onstrated that assessments using structured long cases could be highly reliable (predicted Cronbach's alpha of 0.84),(14) but this required 10 separate cases and 20 examiners raising major issues of practicality.

Short cases

In traditional tests of clinical competence, candidates undertook a series of (usually three to six) short cases. In this type of test, they were taken to a number of patients with widely differing conditions, and asked to examine individual systems or areas and give differential diagnoses of their findings, or to demonstrate abnormal clinical signs or produce spot diagnoses. Although in some ways similar to an OSCE in that they provided a wider range of cases on which the examiner was able to base his or her opinion of the student's ability, there are important differences. Different candidates rarely saw the same set of patients, cases often differed greatly in their complexity and the same two assessors examined the candidate at each case. These cases were not designed to test communication skills, but instead concentrated on clinical examination skills, with communication with the patients merely incidental. The examination was not structured and the examiners were free to ask any questions they wanted. Like the long case there was no attempt to standardise the expected level of performance. For all these reasons OSCEs have superseded this type of assessment.

Objective Structured Clinical Examinations

The remainder of this chapter relates to the OSCE, an assessment format in which the candidates rotate sequentially around a series of structured cases located in 'stations', at each of which specific tasks have to be performed. The tasks usually involve a clinical skill, such as history taking, examination of a patient or a practical skill. The marking scheme for each station is structured and determined in advance. There is a time limit for each station, after which the candidates have to move on to the next task.

The basic structure of an OSCE may be varied in the timing for each station, the use of a checklist or rating scale for scoring, the use of a clinician or standardised patient as examiner and the use of real patients or manikins, but the fundamental principle is that every candidate has to complete the same assignments in the same amount of time and is marked according to a structured marking schedule.

The terminology associated with the OSCE format can vary – in the undergraduate arena they are more consistently referred to as OSCEs, but in the postgraduate setting a variety of terminology exists. For example, in the UK, the Royal College of Physicians' membership clinical examination is called the Practical Assessment of Clinical Examination Skills (PACES), while the Royal College of General Practitioners' membership examination is called the Clinical Skills Assessment (CSA).

Rationale

The use of OSCEs in the quantitative assessment of competence has become widespread in the field of undergraduate and postgraduate medical education(15–19) since they were originally described,(20) mainly due to the improved reliability of this assessment format. This has resulted in a fairer test of candidates' clinical abilities, since the score has become less dependent on who is examining the candidate and which patient is selected for the encounter. The criteria used to evaluate any assessment method are well described(21) and summarised in Chapter 13. We will examine these criteria – reliability, validity, educational impact, cost-efficiency and acceptability – in some detail, as they relate to OSCE design.

Reliability

Essentially, the OSCE was developed to address the inherent unreliability of classical long and short cases. OSCEs are more reliable than unstructured observations in four main ways:

- Structured marking schedules allow for more consistent scoring by examiners according to predetermined criteria; hence reliability is improved.
- Candidates have to perform a number of different tasks across clinical, practical and communication skill domains – this wider sampling across different cases and skills results in a more reliable picture of a candidate's overall competence. The more stations or cases each candidate has to complete, the more generalisable the test is.
- The reliability of the total test score increases with increasing number and increasing homogeneity of stations or cases. Reliability of sub-scores must be carefully reviewed before reporting.
- As the candidates move through all the stations, each is examined by a number of different examiners, so multiple independent observations are collated. Individual examiner bias is thus attenuated.

It is worth bearing in mind that sampling across different cases makes the most important contribution to reliability; the more stations in an OSCE, the more reliable it will be. However, increasing the number of stations has to be balanced with the practicability of an OSCE exercise. Practically, to enhance reliability it is better to have more stations with one examiner per station than fewer stations with two examiners per station.(22,23)

Validity

Validity assessment asks the question 'What is the degree to which evidence supports the inference(s) made from the test results?' Each separate inference or conclusion from a test may require different supporting evidence. Note that it is the inferences that are validated, not the test itself.(24,25)

Inferences about ability to apply clinical knowledge to bedside data gathering and reasoning, and to effectively use interpersonal skills, are most relevant to the OSCE model. Inferences about knowledge, rather than clinically relevant application of knowledge, or clinical and practical skills, are less well supported by this method.(26)

Types of validity evidence include *content validity* and *construct validity*. Content validity (sometimes referred to as *direct validity*) of an OSCE is determined by how well the sampling of skills matches the learning objectives of the course or degree for which that OSCE is designed.(26,27) The sampling should be representative of the whole testable domain for that examination purpose. The best way to ensure an adequate spread of sampling is to use a *blueprint* method, which we will describe later in the chapter.

Construct validity (sometimes referred to as *indirect validity*) of an OSCE would be the implication that those who performed better at this test had better clinical skills than those who did not perform as well. In an OSCE testing situation, we can only make inferences about a candidate's clinical skills in actual practice, as the OSCE is an artificial situation.

To enhance the validity of inferences from an OSCE, the length of any station should be best fitted to the task to achieve the best authenticity possible. Thus, for example, a station in which blood pressure measurement is tested would authentically be achieved in 5 minutes, whereas taking a history of chest pain or examining the neurological status of a patient's legs would be more authentically achievable in 10 minutes.(28)

Educational impact

The impact on students' learning resulting from a testing process is sometimes referred to as *consequential validity*. The design of an assessment system can reinforce or augment learning, or undermine learning:(29,30) it is a well-recognised phenomenon that students focus on their assessments rather than the learning objectives of the course. Explicit, clear learning objectives allied with clinical skills assessment content and format can be a very effective way of encouraging students to learn the desired clinical competencies. Objectives that include action verbs like 'demonstrate' or 'perform', which are then linked to OSCEs that measure ability to demonstrate or perform certain skills, will encourage students to practise these skills. By contrast, an assessment system that measures students' ability to answer multiple-choice questions about clinical skills will encourage students to focus on knowledge acquisition. Neither approach is wrong – they simply demonstrate that assessment drives education and that assessment methods need to be thoughtfully applied. There is a danger in using detailed checklists as this may encourage students to memorise the steps in a checklist rather than learn and practise the skill. Rating scale marking schedules encourage students to learn and practise skills more holistically.

OSCEs may be used for formative or summative assessment. When teaching and improvement are a major goal of an OSCE, time should be built into the schedule at the end of each station to allow the examiner to give feedback to the student on their performance, providing a very powerful opportunity for student learning. For summative certification examinations, expected competencies should be clearly communicated to the candidates so they have the opportunity to learn the skills prior to taking such examinations.

Cost-efficiency

OSCEs can be very complex to organise. They require meticulous and detailed forward planning, and engagement of considerable numbers of examiners, real patients, simulated patients, and administrative and technical staff to prepare the circuits and station materials and manage the examination. It is therefore most cost-effective to use OSCEs to test clinical competencies and not knowledge, which can be tested more efficiently in a different examination format. Effective implementation of OSCEs requires thoughtful deployment of resources, with attention to production of examination material, timing of sittings, suitability of facilities, catering, and collating and processing of results. Other critical logistics include examiner and standardised patient recruitment and training. This is possible even in resource-limited environments.(31)

Acceptability

The increased reliability of the OSCE format over other formats of clinical testing and its perceived fairness by candidates has helped to engender the widespread acceptability of OSCEs among test takers and testing bodies.

Since Harden's original description in 1979,(20) the use of OSCEs has become widespread in the undergraduate level of testing of clinical competence,(17,18,32,33) as well as increasingly in postgraduate assessment.(16–19,34–39) More recently, OSCEs have been used to replace traditional interviews in recruitment processes in both undergraduate and postgraduate settings.(40,41) For example, for recruitment to general practice training schemes in the UK, candidates go through an OSCE format of scenarios in assessment centres where different exercises are assessed by trained assessors, who observe various job-related competencies, including communication skills, team involvement and problem-solving ability.

In North America, clinical skills assessment has been accepted on a massive scale. In 1992, the Medical Council of Canada (MCC) added a standardised patient component to its national licensing examination because of the perception that important competencies expected of licensed physicians were not being assessed.(40) Since inception, approximately 2500 candidates per year have been tested at multiple sites at fixed periods of time during the year throughout Canada. The MCC clinical skills examination uses physicians at each station to score the encounter.

In the US, the Educational Commission for Foreign Medical Graduates (ECFMG) instituted a performance-based examination in 1998 to assess bedside data gathering, clinical reasoning, interpersonal skills and spoken English communication skills of foreign medical graduates seeking to enter residency training programmes. From 1998 to 2004, when it was incorporated into the United States Medical Licensing Examination (USMLE), there were 43642 administrations, including 37930 first-time takers, making it at the time, the largest high-stakes clinical skills examination in the world.(42) The 11 scored encounters had a standardised format, with each requiring the candidate to elicit a medical history, perform a physical examination, communicate in spoken English with a patient in a clinical setting and generate a written record of the encounter. In each station, the candidate encountered a unique standardised patient – a lay person recruited and trained to give a realistic portrayal of a patient with a standardised medical and psychosocial history, and standardised findings on physical examination. Each case had a case-specific checklist containing the elements of medical history and physical examination considered pertinent to that particular case. Simulated patients were trained to recognise appropriate queries and/or physical examination manoeuvres, including acceptable equivalents or variants, and to document each checklist item achieved by the candidate. Simulated patients also evaluated each candidate's interpersonal skills and spoken English proficiency. After each encounter, the candidate generated a patient note on which the pertinent positive and negative elements of history and physical examination were recorded, a differential diagnosis constructed and a diagnostic work-up plan proposed. Performance was evaluated by averaging scores across all encounters and determining the mean for the integrated clinical encounter (data gathering combined with the patient note score) and communication (interpersonal skills and spoken English). Generalisability coefficients for the two conjunctively scored components of CSA were approximately 0.70–0.90.(43)

In 2004, the USMLE adopted the ECFMG clinical skills assessment model and began testing all US medical graduates in addition to foreign medical graduates seeking ECFMG certification.(44) Additional computer and standardised patient training infrastructure was included to ensure comparability across all centres.

The USMLE Step 2 (Clinical Skills) uses 12 standardised patient encounters, each 15 minutes in length followed by 10 minutes to write a patient note. As in the ECFMG CSA examination, standardised patients document the items asked in the history and performed in the physical examination to specified criteria, and evaluate interpersonal skills and spoken English skills, while physician raters score the patient note. Approximately 35000 administrations take place each year.

OSCE design

We turn now to the elements of good OSCE design, which are summarised in Box 21.1.

Blueprinting

For any particular OSCE, the content – clinical tasks chosen for the stations – should map onto the learning objectives of the course and the candidates' level of learning. It is only reasonable to test candidates on what they have been taught.(27)

To map the assessment to the learning objectives, the categories of skill to be tested should be mapped on one axis and the elements of the course being tested should be mapped on the other. Usually in OSCEs, the skills domains are categorised into clinical examination skills, practical skills and communication skills, which can be further subgrouped into history-taking skills and other doctor–patient/colleague interactions. The subject content of the

OSCE will be determined to a certain extent by how the elements of the course are categorised, that is, by subject discipline or systems.

Blueprinting is a powerful tool that helps to focus the OSCE designers on the exact nature of what they wish to test and relate this to the teaching. Once this blueprint or framework for an OSCE is agreed, the individual stations can be planned and classified according to this blueprint. This ensures adequate sampling across subject area and skill, in terms of numbers of stations covering each skill and the spread over the subjects/systems of the course being tested.

The feasibility of testing a particular task also needs to be considered. Real patients with clinical signs can be used to test clinical examination skills, while simulated patients are best for testing communication skills. Simulated patients can also simulate a number of clinical signs (e.g. loss of visual field, localised abdominal pain). Healthy volunteers can be used when testing the technical process of a clinical examination. There are many manikins on the market for testing invasive practical skills, e.g. intravenous cannulation, urethral catheterisation and arterial blood gas sampling.

BOX 21.1 Elements of Objective Structure Clinical Examination design

Blueprinting

Ensuring the test content maps across the learning objectives of the course

Station development and piloting

Writing stations that function well

Examiner training

Engaging examiners, ensuring consistency of marking contributes to reliability

Simulated patient training

Consistent performance ensures each candidate is presented with the same challenge

Organisation

Making detailed plans well in advance. Be prepared!

It is essential to use a blueprint to plan the content of an OSCE as this helps to ensure that different domains of skill are tested equitably and that the balance of subject areas tested is fairly decided. An example is provided in Box 21.2.

Station development

It is important to write out station specifications well in advance of the examination date so the stations can be reviewed and trialled prior to the actual assessment. Sometimes stations that seem like a good idea at the time of writing may turn out to be unfeasible in practice. When writing a station specification, the following aspects should be considered:

- *Construct:* a statement of what that station is supposedly testing, e.g. this station tests the candidate's ability to examine the peripheral vascular system.
- *Clear instructions for the candidate:* to inform the candidate exactly what task they should perform at that station.
- *Clear instructions for the examiners:* including a copy of the candidate instructions, to assist the examiner at that station to understand his or her role and conduct the station properly.
- *List of equipment required.*
- *Personnel requirements:* whether the station requires a real patient or a simulated patient and the details of such individuals (e.g. age, gender, ethnicity).
- *Simulated patient scenario:* if the station requires a particular role to be played.
- *Marking schedule:* this should include the important aspects of the skill being tested, a marking scheme for each item and how long the station should last. The marking schedule may be either a checklist or a rating scale as there is good evidence that, despite the apparent objectivity of structured checklists, global rating scales have been shown to be equally as reliable (*see* Box 20.2). Items can be grouped into the broad categories of process skills, content skills and clinical management skills.

Process skills

For clinical examination stations with a real or simulated patient, these could include introduction and orientation, rapport, professional manner and communicating with the patient appropriately during examination.

BOX 21.2 Example of a system-based blueprint

	History	Explanation	Examination	Procedures
Cardiovascular	Chest pain	Discharge drugs	Cardiac	BP
Respiratory	Haemoptysis		Respiratory	Peak flow
Gastrointestinal	Abdominal pain	Gastroscopy	Abdominal	PR
Reproductive	Amenorrhoea	Abnormal smear	Cervical smear	
Nervous	Headache		Eyes	Ophthalmoscopy
Musculoskeletal	Backache		Hip	
Generic	Pre-op assessment	Consent for post-mortem		IV cannulation

THIS IS A 10-MINUTE STATION	Good	Adequate	Not done/ Inadequate
1 Introduction and orientation (name and role, explains purpose of examination, confirms patient's agreement)	[ ]	[ ]	[ ]
2 Rapport (shows interest, respect and concern, appropriate body language)	[ ]	[ ]	[ ]
3 Appropriately exposes the patient and positions them at 45 degrees		[ ]	[ ]
4 Looks at hands, commenting on peripheral stigmata (i.e. cyanosis, clubbing, splinter haemorrhages, etc.)		[ ]	[ ]
5 Checks the radial pulse, commenting on the rate and rhythm		[ ]	[ ]
6 Asks for patient's blood pressure ***Examiner please give correct BP***		[ ]	[ ]
7 Looks for central stigmata of cardiovascular disease (i.e. anaemia, central cyanosis, hyperlipidaemia)		[ ]	[ ]
8 Examines the JVP correctly (positions the patient's chin and neck; assesses the waveform in the correct area) and comments on findings		[ ]	[ ]
9 Palpates the carotid or brachial pulses, commenting on the character		[ ]	[ ]
Inspects and palpates the praecordium: 10 Localises the apex beat, commenting on the position		[ ]	[ ]
11 Examines for RVH		[ ]	[ ]
12 Auscultates in all four cardiac areas		[ ]	[ ]
13 Moves patient to left side and sits patient forward in expiration	[ ]	[ ]	[ ]
14 Comments on heart sounds and times heart sounds against central pulse	[ ]	[ ]	[ ]
15 Comments on any murmurs		[ ]	[ ]
16 Listens to the lung bases		[ ]	[ ]
17 Candidate attempts to assess peripheral pulses ***Examiner please stop candidate***		[ ]	[ ]
18 Checks for ankle/sacral oedema		[ ]	[ ]
19 Presents a brief summary and conclusions	[ ]	[ ]	[ ]
20 Communicates with patient appropriately during examination (explain what they are doing, gains patient's co-operation)	[ ]	[ ]	[ ]
21 Examines patient in a professional manner (gentle, watches for pain, maintains dignity and privacy)	[ ]	[ ]	[ ]
22 Closure (thanks patient, leaves patient comfortable)	[ ]	[ ]	[ ]
23 Candidate cleans hands after examination		[ ]	[ ]

Figure 21.1 Example checklist mark sheet for cardiovascular examination.

For history-taking stations, these could include introduction and orientation, listening skills, questioning skills, demonstration of empathy and appropriate closure.

For explanation stations, these could include introduction and orientation, rapport, establishing what the patient knows/understands, demonstration of empathy, appropriate organisation of explanation, checking the patient's understanding and using clear language, avoiding jargon.

Content skills

These include appropriate technical steps or aspects of the task or skill being tested.

Clinical management skills

It may be appropriate to ask the candidate some set questions in relation to the specific case.

Figures 21.1 and 21.2 provide examples illustrating the checklist and rating scale marking schedules, respectively.

Piloting

Ideally, stations should be piloted before they are used in examinations to ensure that all stations are functional in terms of the following:

- *Timing:* can the candidates realistically perform the task in the time allotted?

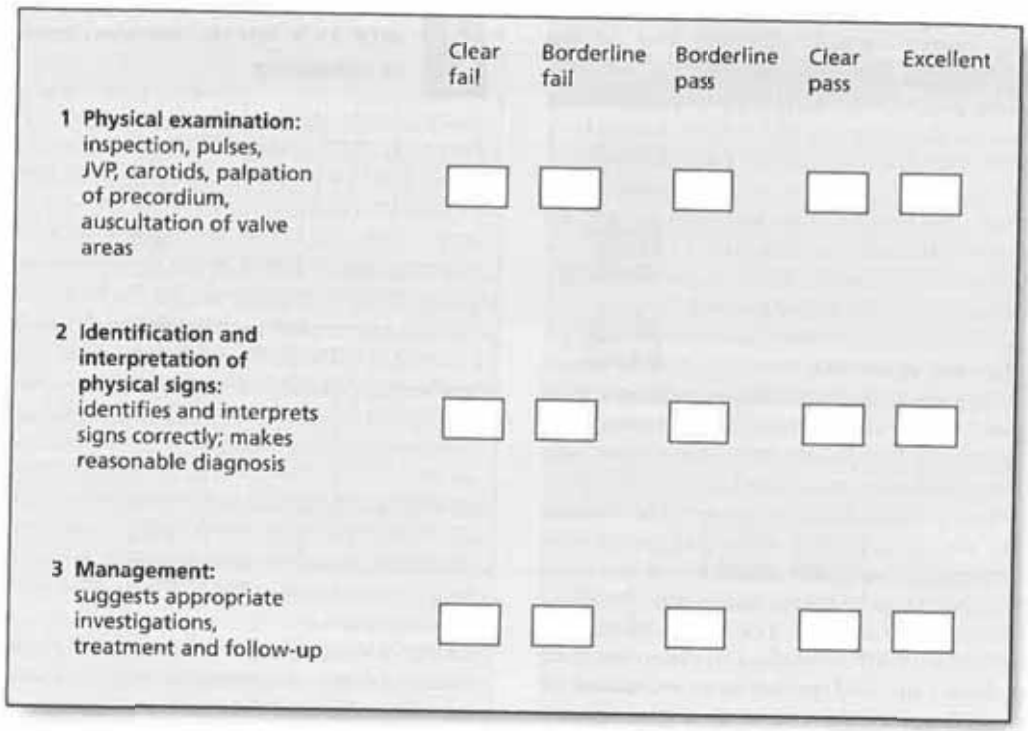


Figure 21.2 Example rating scale mark sheet for cardiovascular examination.

- *Difficulty:* how difficult is the station?
- *Equipment:* is all the equipment required available and on the list?
- *Is an additional helper required* to assist the examiner, e.g. for catherisation, suturing stations?
- *Candidate instructions:* do the instructions tell the candidate exactly what the task is?
- *Examiner instructions:* do the instructions tell the examiner how to conduct the station? Does the examiner know what the candidate has been told to do?
- *Real patient specifications:* are the medical conditions specified?
- *Simulated patient scenario:* is the age/gender/ethnicity specified? Is there enough information for the simulated patient to learn and play the part effectively?
- *Construct validity:* is the station testing what it is meant to test? Does the marking schedule reflect the elements of the task appropriately?

Simulated patient training

For consistent performances, particularly at communication skills stations, it is best to use well-trained simulated patients. Depending on location, it may be possible to organise a database of actors who assist in the teaching as well as assessment of communication skills. It is desirable to have people across a range of ages and ethnicities, as well as a balanced gender mix. Training and monitoring simulated patients is essential to ensure consistent performance – a significant factor in the reliability of the examination. The simulated patients should be sent their scenarios in advance and then asked to go through their roles with other simulated patients playing the same role, while being supervised by a communication skills teacher and/or a clinician, to develop the role to a suitable standard.

Examiner training

OSCEs require large numbers of examiners. This can be a strength, as candidates are observed and scored by clinicians, but it is also one of the potential weaknesses of OCSEs, as inconsistency between examiners will reduce fairness and reliability.

Considerable resources are devoted to examiner training. Structured face-to-face training sessions are good for introducing new examiners to OSCEs and scoring processes. The programme for these events is interactive and very much acknowledges the inherent expertise that experienced clinicians bring to the assessment process. These training sessions cover:

- principles of OSCEs
- role of examiners (i.e. to assess not to teach; to conduct vivas, adhere to marking schedules and respect the role of the simulated patient)
- marking video-recorded OSCE stations, followed by assessment with the clinicians of their marking and getting them to think through their mark allocation
- marking 'live stations' with group members playing the candidate, the assessor and the simulated patient. This demonstrates how stressful this assessment is for the candidate and how difficult it can be to play the part of a good simulated patient

- standard-setting procedure used. This can be crucial when using a student-centred approach, and all the examiners are integral to the standard-setting process. The more the assessors understand their vital role in this process the more likely they are to do it in a satisfactory way. The use of non-clinicians in assessment is discussed in Box 21.3.

Once examiners have had initial training, it may be helpful to refresh examiners' scoring and standards via interactive on-line courses, with videos of candidate performance and feedback on examiner scoring.

Working with real patients

Patients do not always give the same history each time they are asked to repeat it; they can become tired or unwell and they may develop new signs and symptoms to the ones they originally reported; they may even lose previous clinical findings. However, they can be a most valuable resource and need to be treated as such. Using 'real' patients in OSCEs adds greatly to the validity of the assessment. Ideally, patients should be used to assess the detection of common chronic clinical signs. For each clinical sign assessed several patients will be needed and even the most stoical patient should not be expected to be examined by more than 10 students in the course of a day. Ideally, patients should be swapped in and out of the station to allow them to have sufficient rest time.

Practical considerations

The smooth running of OSCEs is highly dependent on the detail of the practical arrangements made in advance and it is worth putting some effort into this to ensure a tolerable day of examinations. There are many aspects to consider.

Prior to the OSCE

Suitable venue

Depending on the number of stations and candidates, more than one circuit may need to be conducted simultaneously. There are advantages (less noise, more privacy for patients) to conducting each station in a separate room (e.g. in an outpatient department), but larger halls divided up with soundproofed partitions can also be suitable. Venues may need to be booked well in advance of examination dates. Appropriate adjacent rooms to the OSCE circuits are required for the gathering of the students, where they can be registered and briefed prior to the examination. Rooms may be required for patients to rest in between each examination. A floor plan of the stations and rest rooms is an invaluable aid to planning.

Video cameras to record encounters may be useful for quality assurance, training of simulated patients and examiners, and standard setting. This is particularly important when attempting to standardise encounters and examiners across different sites.

Recruitment of examiners

Busy clinicians and other teachers will need advance notice to enable them to attend and play the vital role of assessors at each station. It is helpful to send out a grid of dates and

BOX 21.3 FOCUS ON: Simulated patients as assessors

Scoring standardised patient examinations can be done by third-party observers (usually physicians) or by the standardised patients themselves. Physician examiners enhance the validity of the assessment because they can apply holistic judgements and integrate subdomains of sequence, logic and other factors that may be difficult for a non-professional completing a binary checklist to capture. Boulet *et al.*(45) however, demonstrated that holistic judgements from physician examiners are similar to aggregate scores from trained standardised patients, at least in assessing a general, entry-level physician. Physician evaluator models using holistic scoring models may have greater utility in capturing higher levels of expertise – something that checklist models may not be able to do. Any examiner, whether simulated patient or physician, must be thoroughly trained and then monitored to ensure consistent use of the score scale, since variability diminishes reliability.(46)

Ratings of interpersonal and communication skills provide a unique challenge in determining who is best able to provide the ratings. Although the assessment of doctor–patient communication skills can be accomplished by a physician or other observers, and can be done 'live' or via video-taped reviews, it is unclear whether someone watching the interplay between a doctor and a patient can adequately measure the complex, multi-dimensional nature of the communication. Many aspects of this communication, especially those that are non-verbal, are best assessed by the patient or the person trained to be the patient.(44)

Spoken English is another domain that might be better scored by non-physicians. The generalisability coefficient of this component of the Educational Commission for Foreign Medical Graduates Clinical Skills Assessment (CSA), scored by standardised patients, was 0.94.(47)

From a logistical and cost perspective, the examinee volume for the CSA and now the United States Medical Licensing Examination Step 2 (CS) (approximately 35 000 per year) makes it effectively impossible to entertain using physician examiners. Cost analysis also needs to account for training time and quality assurance for standardised patients or physician raters as well as the different nature of the training needed by each group. It may also be harder to standardise a large number of highly educated, typically independently thinking physicians across five test centres in a year-round testing model.

Currently in the UK, at both the undergraduate and postgraduate levels, examiners are clinicians or other health-care professionals. It would probably need a considerable shift in cultural acceptance to move to standardised patients as the sole assessors of clinical competence.

Station	Skill	Scenario	Expected response	Curriculum Reference
1. Physical Examination	Upper limb examination	A 34 year old dance instructor presents complaining of right shoulder pain and difficulty moving it. There is no history of trauma. Perform a full examination of the upper limbs.	Candidates are required to correctly examine the patient and present findings to the examiner. Marking will include examiner assessment and patient feedback.	Medical Expertise
2. Procedure 3. (Double Station)	Difficult airway	A 50 year old man with severe facial injury after being assaulted requires airway management. You have been asked by the Team Leader to manage his airway. The scenario will be on a high fidelity mannequin. You have an experienced airway nurse as an assistant.	Candidates are required to demonstrate appropriate airway management of difficult airway, where intubation is not possible. They are expected to attempt ventilation with RSI, then difficult airway algorithm and ultimately including "can't ventilate, can't intubate" situation. *See Specific detailed OSCE sample - Airway Management*	Medical Expertise Teamwork and Collaboration Prioritisation and Decision Making
4. History taking	Cardiac history	A 34 year old man presents to the emergency department with sudden onset of palpitations that have now resolved. This is his ECG. (The ECG is normal) Take a focussed history, risk stratify him, and explain likely causes to this patient.	Candidates are required to take an accurate focused history with regard to palpitations and risk factors, consider and explain the likely possible causes to the patient. Marking will include examiner assessment and patient feedback	Medical Expertise Communication

Station	Skill	Scenario	Expected response	Curriculum Reference
5. History taking	Sexual and drug history	A 32 year old woman, who is a known IV drug user, presents with an offensive purulent vaginal discharge. Take a focussed history and explain your recommended investigation plan to the patient.	Candidates are required to take a focussed history, and develop an investigation plan, incorporating immediate clinical and preventive health issues.	Health Advocacy Medical Expertise Communication
6. Patient management	Clinical synthesis from notes and documentation	You are on duty in your Short Stay Ward and you have just taken over the care of a 68 year old man with a suspected TIA that has now resolved and is being discharged. You are provided with the notes which contain a full history and examination and patient charts. You are to formulate a management plan for this patient. Write a brief letter to the GP, including the formulated management plan	Candidates are required to write a brief letter to the GP, including relevant findings and an appropriate management plan.	Prioritisation and Decision Making Communication Medical Expertise
7. Risk assessment from history	Assessment of suicide risk	A 19 year old woman has been admitted following an overdose of benzodiazepines. She has been resuscitated and is now threatening to take her own discharge although she has not yet been assessed by the psychiatrist. You are asked to assess her safety risk and explain to her your decision around her request to go home.	Candidates are required to perform a suicide risk assessment and communicate the outcome to the patient.	Communication Health Advocacy Prioritisation and Decision Making

Station	Skill	Scenario	Expected response	Curriculum Reference
8. Communication / confidentiality	Confidentiality and legalities of consent	You have been asked to talk to a distressed father who has just arrived in the emergency department demanding information about his son. His 17 year old son was admitted earlier after being involved in a road traffic accident in which he is suspected to have been driving a stolen vehicle. He has had a full drug and alcohol screen and the police are present. The son has indicated that he does not wish to talk to his father. You will speak to the father in a private relatives' room.	Candidates are required to talk to the father, bearing in mind the right of the patient to confidentiality and the patient's wishes. Candidates are expected to talk calmly and appropriately to the father in a manner that helps to resolve the situation.	Professionalism Communication
9. Consent	Communication and gaining consent X-ray interpretation	A 22 year old soccer player has been brought in with a wrist fracture which requires reduction. You are provided with the X-ray. You are asked to explain the X-ray findings to the patient and gain their consent for a manipulation in the emergency department by the method of your choice.	Candidates are required to correctly interpret the X-ray Candidates are expected to give a thorough explanation to the patient regarding their injury and recommended management. Candidates are also expected to obtain informed consent from the patient.	Medical Expertise Communication

Station	Skill	Scenario	Expected response	Curriculum Reference
10. Resuscitation 11. (Double Station)	Paediatric resuscitation	An 18 month old child is brought in by ambulance because he is unwell. His mother is accompanying him and on arrival he starts fitting. You are the Team Leader in a tertiary hospital, and have 2 nurses and a junior doctor to assist you. The history obtainable is that he has had high fevers for a day, vomiting, lethargic and developed a rapidly spreading rash.	Candidates are required to lead the simulated team resuscitation response while obtaining a focussed history from the mother. Candidates are expected to consider: - IV access including IO if required. - Vital signs - Seizure control - IV antibiotics - IV fluids - Airway management - Appropriate consultation - Explanation to the mother of events	Medical Expertise Teamwork and Collaboration Leadership and Management Communication
12. Teaching	Lumbar puncture	A 25 year old woman requires a lumbar puncture to exclude subarachnoid haemorrhage. The junior doctor has seen a lumbar puncture done recently but has never performed one. An examiner will be playing the role of the junior doctor. A training mannequin is available. You are asked to explain to a junior doctor how to do the lumbar puncture.	Candidates are required to explain to a junior doctor how to perform a lumbar puncture, including positioning, anatomical landmarks, indications, contra-indications, sterile technique, procedural technique and investigation.	Medical Expertise Scholarship and Teaching

Station	Skill	Scenario	Expected response	Curriculum Reference
13. Communication	Breaking bad news	A 67 year old woman has collapsed at home. She has a GCS of 5, is spontaneously breathing and maintaining her blood pressure. Her CT scan shows a massive inoperable intracerebral bleed. You have been asked to speak to her husband who was present when she collapsed.	Candidates are required to discuss expected short-term management - ICU, IPPV etc., as well as explaining the grim prognosis.	Health Advocacy Communication
14. History taking	Diarrhoea	You are required to take a history from a 25 year old man who presents with a 2 week history of diarrhoea. After 6 minutes you will be asked by the examiner to summarise your findings and provide a differential diagnosis.	Candidates are required to take a focussed history, and formulate the information into a concise structured summary and differential diagnosis. *See Specific detailed OSCE sample – History Taking*	Medical expertise Communication
15. Communication	Back pain	You are asked by your Resident Medical Officer for assistance with a 35 year old patient who presented with back pain. A thorough history and examination has been performed and is unremarkable. The findings are consistent with mechanical back pain. You have examined the patient yourself and are confident there are no red flags. The patient is not happy with the assessment of the treating doctor and feels that the RMO is not listening to his concerns. You, as the Senior Doctor, are asked to speak to the patient about his concerns.	Candidates are required to address the patient's concerns in a professional manner, support the junior doctor, and manage the patient's expectations.	Communication Health Advocacy Professionalism

Station	Skill	Scenario	Expected response	Curriculum Reference
16. Communication and teaching	Investigation plan	A 27 year old pregnant patient who is 28 weeks gestation presents with chest pain. She is otherwise healthy. Her vital signs are provided and are normal. You are also provided with an ECG and chest X-ray (both normal). You have decided that pulmonary embolism needs to be excluded. Hospital policy requires that investigation of all pregnant patients requires discussion with a radiologist. You are required to speak to the radiologist on the telephone and discuss your imaging choice in a patient with a suspected pulmonary embolism.	Candidates are required to provide a coherent and evidence-based investigation plan that can be justified to a consultant peer.	Medical Expertise Communication Prioritisation and Decision Making
17. Management	Asthma	A 26 year old man is brought into the emergency department of your rural district hospital having a severe asthma attack. You have immediately available an experienced emergency department nurse and an emergency department registrar. The arterial gas result is provided. You are asked to manage the patient.	Candidates are required to describe the blood gas result to the staff and commence therapy for severe asthma. *See Specific detailed OSCE sample - Asthma*	Leadership and Management Medical Expertise

Station	Skill	Scenario	Expected response	Curriculum Reference
18. Interpretation of investigations	Chest X-ray	A healthy 40 year old man who is a non-smoker presents with sudden onset of chest pain. His chest X-ray shows moderate spontaneous pneumothorax. You are required to describe the X-ray result to the patient and discuss treatment options.	Candidates are required to identify the pneumothorax and discuss treatment options and implications with the patient. The patient will be actively questioning as to the pros and cons of treatment options. The patient feedback will be sought as well as examiner observation of the explanation.	Medical Expertise Communication

APPENDIX 5

MCQ Writing Template			
PE/FE/CPD:	Type PE/FE/CPD	Author(s):	Type your name here
Subject and topic reference:	Type subject and topic	LOA	Type LOA here
Text Reference	Type references to text here		
Question stem:	Type clinically relevant question stem to assess application of knowledge here.		

Options (in alphabetical order)	Correct Answer letter: Type correct answer letter here
A	
B	
C	
D	

Reviewer comments:	Please leave this space for reviewer comments

APPENDIX 6

EMQ Writing Template

PE/FE/CPD:	Type PE/FE/CPD	**Author(s):**	Type your name here
Subject and topic reference:	Type subject and topic.	**LOA**	Type LOA
Theme:	Type your theme here		

Options: (insert your options in alphabetical order in the table below)
Add more rows if required. Do not exceed X. Use fewer if required.

A		**I**	
B		**J**	
C		**K**	
D		**L**	
E		**M**	
F		**N**	
G		**O**	
H		**P**	

Lead-in phrase: Type your lead-in phrase here

Stem (insert your stems below- not all have to be used)	Reference	Answer letter

Reviewer comments: Please leave this space for reviewer comments

APPENDIX 7

SAQ Writing Template			
FE/CPD:	Type PE/FE/CPD	**Author(s):**	Type your name here
Subject and topic reference:	Type subject and topic	**LOA**	Type LOA here
		Mark:	Type mark allocation here
Text Reference	Type references to text here		
Question stem:	Type clear, focused stem here.		

Marking Scheme
Type simple, detailed marking scheme to ensure objective and consistent marking

Model Answer
Type a comprehensive and clear model answer- clearly outline expectations.

Reviewer comments:	Please leave this space for reviewer comments

APPENDIX 8

OSCE Writing Template			
FE/CPD:	Type PE/FE/CPD	**Author(s):**	Type your name here
Subject and topic reference:	Type subject and topic	**LOA**	Type LOA here
		Mark:	Type mark allocation here
Text Reference	Type references to text here		
Question stem:	Type clinical scenario stem here.		

Instructions
Candidate: Type candidate instructions
Examiner: Type examiner instructions
Patient: Type patient instructions

Scoring sheet
Type a list of specific items important to task performance.

Reviewer comments:	Please leave this space for reviewer comments

APPENDIX 9

CONFIDENTIALITY AND INTELLECTUAL PROPERTY STATEMENT

TO: **AUSTRALIAN COLLEGE FOR EMERGENCY MEDICINE**

I, _____
[Full Name]

_____
[Address]

_____

_____

AGREE AND ACKNOWLEDGE THAT:

1. all information from or about the examinations for the Australasian College For Emergency Medicine (ACEM) including but not limited to examination questions, information about the ACEM examination, examination papers or results and information concerning the procedures and process of the ACEM **(Confidential Information)** must not be used, copied, reproduced, distributed or disclosed in any format;

2. the questions on the MCQ examination paper are the only copy of these questions and that there are to be no duplicates, extracts or adaptation of these questions;

3. I will not reveal, disclose, amend, use or reproduce the Confidential Information or provide it to any other person;

4. I will provide all assistance reasonably requested by ACEM in connection with maintaining the confidentiality of the Confidential Information;

5. I have established and maintained security measures to safeguard the Confidential Information from unauthorized access, copying, reproduction, distribution, disclosure, use of tampering; and

6. I agree to immediately notify ACEM upon becoming aware of any suspected or actual breach of confidentiality, unauthorized access, copying, reproduction, distribution, disclosure, use or tampering of the Confidential Information. I further agree to return or destroy to the College upon their direction, all Confidential Information and any materials in which Confidential Information may be contained.

DATED this _____ day of _____ _____ (Year)

SIGNED: _____