

Single best answer question-writing tips for clinicians

J L Walsh,^{1,2} B H L Harris,^{2,3} P E Smith²

¹Vascular Medicine Programme, American University of Beirut Medical Centre, Beirut, Lebanon

²Cardiff University School of Medicine, Heath Park, Cardiff, UK

³St Anne's College, University of Oxford, Oxford, UK

Correspondence to

Professor Phil Smith, The Alan Richens Epilepsy Unit, Department of Neurology, University Hospital of Wales, Heath Park, Cardiff CF14 4XW, UK; SmithPE@cardiff.ac.uk

Received 10 December 2015

Revised 7 April 2016

Accepted 31 May 2016

Published Online First

1 July 2016

ABSTRACT

Assessment is essential for progression in medical careers. Thus, an important aspect of developing as a clinical teacher is the ability to produce high-quality assessments for junior colleagues. The single best answer (SBA) question format is becoming ubiquitous in the assessment of the application of knowledge in clinical medicine; writing this style of examination question can be a challenge. This concise guide highlights key SBA question-writing tips, aiming to help aspiring clinical teachers set high-quality knowledge assessments.

INTRODUCTION

The National Board of Medical Examiners (NBME) item-writing manual is a mainstay of guidance for question writers aiming to produce high-quality questions. It gives advice for developing many multiple-choice question formats, including multiple true/false, single best answer (SBA) and extended-matching questions (EMQs). Since the item-writing manual was commissioned, most medical examinations of knowledge have adopted the SBA format. The reasons SBAs have become so widely adopted are varied, but include the pitfalls of the other formats. Multiple true/false questions are the easiest to construct but may comprise simply a series of sometimes unrelated statements. However, they rarely reflect the real-life practices of medical diagnosis and management, which come in multiple shades of grey. EMQs assess application of knowledge rather than simple recall but are hard to write well. It is exceptionally difficult to construct EMQs with a suitable number of distractors applicable to all scenarios without undue redundancy. In contrast, well-written SBA questions can concisely assess the application of clinical knowledge, sampling widely from potentially hundreds of learning outcomes in a single easily marked assessment. These characteristics have led them to become the most widely used multiple-choice question format in medical schools.^{1–5}

The NBME item-writing manual provides valuable advice on constructing SBA questions.¹ Here, we highlight and elaborate upon its key points and offer additional advice derived from writing experience.

SOME GENERAL POINTS

The assessors' work for multiple-choice question examinations is almost entirely in their preparation, rather than in their marking. SBA papers can be prepared many months in advance and avoid the postexam frenzy of essay marking. However, it can be very challenging to produce high-quality, discriminating SBA questions that ensure fair, valid and reliable assessments.

The most clinically applicable SBA questions present a scenario, with relevant and plausible options (at least to the mind of a borderline candidate); the 'best' answer might be judged as 80% correct and the distractors perhaps 20–30% correct. While students in clinical practice obviously do not have the prompt of possible options, SBAs do encourage students to work with conditional probabilities that compare to real clinical practice.

The language and scientific content of questions must show clarity, precision and economy to ensure that they test only knowledge and application of knowledge rather than the interpretation of language and syntax. Furthermore, every question must pass the best practice cover test, which involves covering the answer options and attempting to answer solely from the stem (the scenario) and lead-in question (eg, what is the most likely diagnosis?).¹ This ensures a truly 'best of' question style: the good candidate knows the best answer already, without needing the prompt of possible options. A question such as 'Which of the following statements is correct?' followed by a series of options comprising one true and four false statements inevitably fails the cover test and is therefore not a true SBA question.

We now offer guidance to aspiring question setters looking beyond the cover test to improve further their SBA question writing.

FOLLOW A HOUSE STYLE

Candidates welcome assessments written in a consistent and predictable style (table 1). This allows them to concentrate on the topic of the question rather than being distracted by unexpected language or by findings and investigations presented in an unexpected order. A formal house style document is very helpful,⁶ specifying several minimum requirements:

- ▶ **Question length:** In most SBA exams, candidates answer each question in 60–90 s. Thus, each stem should comprise no more than 5–10 lines (including investigation results) and fewer if there is also a table or picture. More detailed stems do help to make each answer option more plausible (and so make a 'better' question). Longer questions are certainly more difficult but surprisingly are not more reliable.⁷ It is also important to remember that very long questions reduce the total number of questions in a paper for a given testing time, and therefore reduce the reliability of the test as well as restricting the breadth of knowledge tested.
- ▶ **Present tense:** A question written in the present tense often feels more relevant and fresh than one written in the past tense; it also avoids clumsy pluperfects, such as 'he had had', and



CrossMark

To cite: Walsh JL, Harris BHL, Smith PE. *Postgrad Med J* 2017;**93**:76–81.

Table 1 Summary of advice for question writers

To be encouraged	To be avoided
Sticking to a 'house style'	Straying from the standard order of case presentation
Using clear language	Using unnecessarily long stems
Blueprinting questions	Using negative questions
Targeting the appropriate cognitive level	Using data formulations
Using present tense and active sense	Sticking rigidly to 'But it was an actual case!' philosophy
Using standard normal values	Omitting the best answer from the options
Using the traditional cover test and/or the science cover test	Tipping off test-wise candidates
Peer reviewing questions	Using options such as 'none of the above' and 'all of the above'

makes the question slightly shorter. Questions in the present tense still need careful proof reading, for example, changing 'the day before' to 'yesterday', and 'two days previously' to 'two days ago'.

- ▶ **Active sense:** Questions written in active sense are easier to understand, less formal and often shorter than those written in the passive sense.
- ▶ **No negative questions:** Although easy to write, negative questions typically have little intrinsic value when testing clinical medicine. A doctor's knowledge of the 'least likely' diagnosis rarely helps a patient. A negative question also fails the cover test spectacularly as candidates have little hope of identifying a wrong answer (from an infinite number of possible wrong answers) without seeing the options. Negative questions can sometimes be salvaged by using lead-ins such as 'What clinical feature most strongly suggests an alternative diagnosis?' or 'What investigation is most likely to be normal?'
- ▶ **Ordering information:** The order of presented material should reflect an assessment in clinical practice. Thus, age, sex and presenting complaint should precede past, family and social history; the examination and investigations should then appear in a predetermined order.
- ▶ **Lead-ins expressing uncertainty** reflect real clinical practice: Thus, a lead-in should seek the 'most likely' diagnosis, the 'best' investigation or 'most important next step' in management (noting that this differs from the 'next most important step', which actually seeks the second-best answer). Such questions are likely to test application of knowledge rather than just recall.
- ▶ **Consistent writing style:** Style should be consistent and written to an agreed style guide (even where authors' opinions may differ). For example, age may be indicated as a hyphenated adjective (eg, an 86-year-old man) or unhyphenated, and sex as 'man' or 'woman' (not 'gentleman' or 'lady'), or 'boy' and 'girl' if aged below 18 years. Spelling should be consistent, for example, 'fetus' or 'foetus', as should hyphenation versus en dash between disease names of eponymous attribution, and on standardised medication names. Question writers should avoid an excess of italics or underlining owing to their difficulty for people with dyslexia; nevertheless, most authors will still write bacterial and gene names in italic script.
- ▶ **Consistent normal values:** Including normal values in questions reflects the real clinical situation; authors need an agreed, consistent and frequently updated list of these. It is

less important that they exactly reflect data from a specific laboratory or website but more that they are consistent throughout the paper(s).

- ▶ **Blueprinting:** Studying the in-house syllabus and adhering to course learning objectives allows questions to be tailored to a specific course and to avoid unnecessary inclusion of rarities. This focused approach also helps writers in choosing distractors. The act of blueprinting and tagging questions to learning outcomes allows easy indexing and helps to avoid topics that are over-represented or absent from the question bank.⁸

PLACE THE PRESENTING COMPLAINT FIRST

Candidates should expect to find the presenting complaint and its duration (the central topic of the question) in the first line as this best reflects real-life clinical practice. Question writers often overlook this aspect, for example:

A 26-year-old man undergoes major surgery to his thigh after a motorcycle crash. He has asthma and takes a salbutamol inhaler regularly. Three days later, he collapses suddenly on the ward. On examination,

This question becomes clearer and more relevant with the presenting complaint placed first:

A 26-year-old man collapses suddenly on the ward. Three days ago, he underwent major surgery to his thigh after a motorcycle crash. He has asthma and takes a salbutamol inhaler regularly. On examination, ...

PROVIDE RAW DATA RATHER THAN A FORMULATION

Providing actual data or pictures (rather than summary formulations) increases the authenticity of a written assessment of clinical knowledge. For example, 'His temperature is 38.9°C' requires more knowledge than 'He has a pyrexia'. Similarly, 'Her body mass index is 34.2 kg/m², is preferable to 'She is obese'. Using data better reflects clinical practice and requires candidates to draw their own conclusions from the clinical information. Furthermore, formulations such as 'end-inspiratory crackles' or 'low-pitched rumbling diastolic murmur' can cue the answer. Similarly, pathognomonic terms such as 'the chest X-ray shows the sail sign' or 'there are Koplik spots' provide an interpretation beyond that expected from a real clinical encounter. Seeing a chest X-ray with left lower lobe collapse or the palate of a child with measles better tests candidates' understanding.

ENSURE THE BEST ANSWER IS AMONG THE OPTIONS

An option list omitting the best answer not only fails the cover test but also does not reflect best practice. Question writers sometimes omit the truly best answer—perhaps because it seems too easy and the setter seeks the second-best answer. For example:

A 15-year-old boy develops myoclonic jerks and has a generalised tonic-clonic seizure. The doctor discusses the best long-term antiepileptic treatment

What treatment is most likely recommended?

- A. ethosuximide
- B. lamotrigine
- C. levetiracetam
- D. phenytoin
- E. topiramate

Answer C: levetiracetam.

The best (or equally good) answer would be sodium valproate but it is not a presented option; perhaps the question writer was

trying to assess more subtle knowledge of second-line treatments. This is easily solved:

A 15-year-old boy develops myoclonic jerks and has a generalised tonic-clonic seizure. He has previously tried sodium valproate but this caused troublesome weight gain; he wishes to try a different treatment. The doctor discusses the best long-term antiepileptic treatment.

The lead-in and answer options remain the same but now contains the actual best answer.

BE FLEXIBLE WITH QUESTIONS DERIVED FROM REAL CASES

Although many excellent questions are based on real clinical scenarios, question writers should not feel confined by the actual clinical narrative. Many authors protest, 'But it was an actual case!', thinking this somehow justifies designating an unusual answer as the 'most likely'. The question writer's priority is to create a good question, not to tell an interesting anecdote. Trying to be faithful to an actual case is a common fault of novice question writers and a block to producing excellent questions.

MAXIMISE CLARITY, ECONOMY AND PRECISION IN THE LEAD-IN AND OPTIONS

Where the same word or phrase appears in each option, this can move to the lead-in. For example:

What structure is most likely involved?

- A. axillary nerve
- B. median nerve
- C. musculocutaneous nerve
- D. posterior interosseous nerve
- E. ulnar nerve

This is more economical and clear as:

What nerve is most likely to be involved?

- A. axillary
- B. median
- C. musculocutaneous
- D. posterior interosseous
- E. ulnar

'Which of the following ...?' is unnecessarily verbose. It is implicit in the multiple-choice format that the required answer lies among the options given. This phrase may defend against a candidate's complaint that he/she knows the preferred answer but cannot find it among the options. However, if the questioner ensures the best answer is among the options (see above), then 'Which ...?' or 'What ...?' suffice and avoid the more wordy lead-in.

Incidentally, 'Which?' implies the choice lies only among the options presented (certainly true for written examinations but not reflecting clinical practice). If a candidate must first see the options to answer a question, it fails the cover test. By contrast, 'What?' passes the cover test by implying that all options are possible.

REFLECT GOOD AND REALISTIC CLINICAL PRACTICE

A question's clinical scenario should evolve as in real life. For example, for a person attending with a rash after starting a new drug, a lead-in such as 'What drug is most likely caused the rash?' is unrealistic as the doctor would know very well which drug had been given. The scenario becomes more realistic if, for example, the patient presents to the emergency department while on holiday, having started a new drug but unable to remember its name.

Although examples of poor clinical practice are (regrettably) realistic, questions should, where possible, reflect good practice and adhere to national guidelines. If a question writer wishes to include a poor standard of care within a scenario, for example, omission of a heel-prick test as a neonate leading to a childhood presentation of phenylketonuria, then this might involve changing the scenario to, for example, a patient who immigrated in childhood, or someone from a travelling family without access to regular healthcare.

SET ANSWER OPTIONS WITHIN ONE DOMAIN

Distractors work best if drawn from the same domain and appear similar to one another. A common shortcoming is to have distractors from different domains, reducing their potential for discrimination. The most obvious example is including both investigations and treatments among the options, for example, 'laparotomy', 'CT scan of abdomen', 'refer to haematologist' or giving both drug classes and specific drug options, for example, 'benzylpenicillin', 'trimethoprim' and 'aminoglycoside'. Answers in different domains risk allowing two equally good answers from different domains to be equally valid, for example, requesting a CT scan and inserting an intravenous cannula.

PEER REVIEW QUESTIONS IN SUPPORTIVE GROUPS

Peer review undoubtedly improves question quality^{1 9-11} Question-writing groups work best if all members bring questions for discussion and peer review, and all experience having their questions criticised and improved. It is constantly surprising (and appropriately humbling) how often an experienced question writer's 'perfect' question improves further with such scrutiny. These groups must be carefully chaired to permit an equal voice to all participants and to allow all group members' contributions to be equally valued, irrespective of seniority. No member should fear exposing a knowledge blind spot (we all have these); on the contrary, not knowing something is an opportunity for all to learn. Peer review also checks for accurate content, standardises questions, ensures consistent formatting, punctuation, grammar and spelling, and limits superfluous information.

In clinical medicine, doctors must sort large quantities of information, picking out that which is relevant. However, long questions are frustrating to read and reduce the breadth of knowledge that can be tested in a given examination. Peer review helps to strike this balance between making the question realistic while trimming redundant information.

On a similar note, it is important that a paper is proofread by someone independent and experienced before its final acceptance: at this stage, question writers sometimes may be blind to ambiguities or confusing terminology, particularly if the question has gone through several iterations. In many assessment programmes, this role falls to standard setting groups.

PROVIDE FEEDBACK AND REFERENCES FOR QUESTIONS

Formative assessments provide an excellent vehicle for learning. While faculty members may frequently provide formative questions to students, they typically withhold their 'best' questions for summative tests.

Formative questions should ideally include detailed written feedback, justifying the best answer and (in particular) using the evidence base to explain why other options are less correct; this all reinforces the students learning.¹² Including references to the literature also helps, for several reasons:

1. ensuring that questions are accurate and supported by current evidence;

2. facilitating review (including answer key checks after the exam) and updating questions, for example, with new national guidance;
3. checking that the author has not misinterpreted a source or used an outdated source;
4. guiding further reading.

We also encourage students to set up their own question-writing groups to generate formative questions for their peers.^{10–13} Another approach is to use a free online platform such as PeerWise, allowing anonymous authoring and critique of questions,¹⁴ together with rich and well-referenced feedback.

Feedback for summative exams is more difficult to provide in practice as many question writers do not have sufficiently large numbers of questions in their banks to allow dissemination of detailed information about the questions used (they will likely be used again in future assessments). However, mapping questions to specific high-level learning outcomes and transmitting this information back to students together with their performance is a useful compromise.

APPLY THE 'SCIENCE COVER TEST'

Writing basic science questions through the medium of a clinical case is challenging but necessary if students are to be assessed on the basic science relevant to clinical medicine.¹⁵ How does one create a clinical scenario to test knowledge on, say, the endoplasmic reticulum? Simply asking, 'What is the main role of the endoplasmic reticulum?' is an admission that the knowledge needed to answer this question is not sufficiently clinically relevant to form part of a realistic clinical scenario: an uncomfortable message to send to an aspiring clinician. Conversely, paying lip service to the need for clinical scenario might give a question such as:

A 12-year-old boy with muscle weakness undergoes a muscle biopsy. The specimen is examined using electron microscopy and the pathologist notes an abnormality of the endoplasmic reticulum.

What is the main role of the endoplasmic reticulum?

The scenario here provides a clinical setting but is not necessary to the question. It fails the 'science question cover test', where the candidate can answer the question even after covering the stem. The conscientious candidate will read it all and become frustrated by the unnecessary work. The test-wise candidate will quickly see the lead-in and ignore the scenario. Our challenge is to make clinical scenarios that are relevant to basic science questions.

To ensure that the lead-in is not answerable as a freestanding question, it must refer to elements in the clinical scenario to ensure that the scenario is integral to the question. A rewrite of this question that allows the question to pass the science cover test might be:

A 12-year-old boy with muscle weakness undergoes a muscle biopsy. The specimen is examined using electron microscopy and the pathologist notes an abnormality of the endoplasmic reticulum.

What cellular function is most likely to be abnormal?

TIPS FOR USING CLINICAL SCENARIOS IN BASIC SCIENCE QUESTIONS

Basic science questions can often be made clinically relevant and realistic:

- ▶ *Calculation questions* are easily set into a clinical context, for example, through prescribing a paediatric dosing regimen (mg/kg): 'What is the best estimate of the dose given over 6 hours?' Another example, 'The doctor explains that the

estimated glomerular filtration rate has fallen by 80% from its baseline value. What is the best estimate of her current glomerular filtration rate?' Incidentally, the increasing sophistication of smart 'cheating' technology may mean that calculation questions in future examinations will need to be answerable without requiring a calculator.

- ▶ *Anatomy questions* are easily made clinically relevant. For example, 'A 19-year-old man develops a dropped foot after fracturing his fibula. What nerve is most likely affected?'
- ▶ *Histology questions* are also easily made clinically relevant where a patient (with any disorder of any organ) undergoes a biopsy that contains normal tissue, for example, 'What cell type most likely predominates?'
- ▶ *Biochemistry and physiology questions* are generally more difficult as patients are less likely to present to doctors with problems relating to normal function. Key exceptions might be the normal menstrual cycle, normal pregnancy and embryology. A physiology question might be, for example, a 45-year-old man with breathlessness who undergoes an investigation (eg, a treadmill test) and is told the result is normal. What maximum heart rate did he most likely achieve? A biochemistry investigation may relate to normal fluid balance or normal effects of starvation in an otherwise healthy patient undergoing surgery.
- ▶ *Pathology and microbiology questions* fit readily into clinical scenarios as they feature patients with certain diseases.
- ▶ *Social science, ethics and professionalism questions*. These discursive topics do not readily fit into multiple-choice formats and often require longer clinical scenarios to make them plausible. They remain a major challenge for SBA questions and may be better addressed in an EMQ, short answer question or essay format.

BE AWARE OF TIPPING OFF TEST-WISE CANDIDATES

It is sometimes possible to answer poorly written multiple-choice questions correctly without knowledge of their content. Indeed, the NBME item-writing guide alludes to avoiding "technical item flaws that provide special benefit to test-wise examinees".¹ Question setters must be aware of such strategies (10 are listed below) and carefully proof their questions to protect against these. Test-wise candidates might:

1. *Select the middle option in a list of options*, especially with a list of numeric options, since many question writers unthinkingly place a couple of numbers above and a couple below their preferred answer. It is best to put options in alphanumerical ascending order, removing concerns around candidates obtaining the correct answer by guessing, through selecting an over-represented middle option.¹⁶
2. *Use convergence strategy*: If options each have two components (eg, benzylpenicillin and gentamicin), choose the option that contains the two that appear most frequently among other options. Thus, among the options:
a+b
a+c
a+d
c+d
c+e
a and c appear most often and so a+c is likely to be the best option. The question writer has tried to hide a+c by mixing other options with each of these but in doing so has cued the answer.
3. *Choose one of two opposites or one of two similar options* ('best of two'); question writers often try to balance their preferred answer by offering its opposite.

4. *Ignore the answer that offers 'x only' or 'y only' when there is an option of 'x+y'*; the question writer usually wants to check that candidates know that both components are necessary.
5. *Choose the longest, most elaborate, most complex or most detailed answer*, on the grounds that the question setter would put this all in only if it were needed; similarly, choose an eponymous term if offered rather than a descriptive term.
6. *Choose 'no treatment necessary', 'none of the above' or 'all of the above'* if these are offered. Using these phrases as options is not good practice. 'None of the above' severely impacts on the functionality of the question and cannot test if the candidate actually knows the best answer. In addition, a question with 'all of the above' as an option not only fails the cover test, but risks being alphabeticised to the top!
7. *Select any answer that is cued in the question*, for example, 'What is the most appropriate treatment?' might be followed by four management strategies and one treatment.
8. *Note questions on similar themes elsewhere in the same paper* as one may give the answer to the other.
9. *Avoid options containing absolutes such as 'always', 'never', 'must', 'invariably', etc.*, since question writers typically qualify their best answer appropriately; options containing 'may' or 'could' are more attractive.
10. *Choose an answer written in language characteristic of a particular examiner* (eg, an idiosyncratic phraseology used in lectures).

CREATING SBAS FOR UNDERGRADUATE VERSUS POSTGRADUATE ASSESSMENT

When examining at different levels, it is important to pitch the question at the right level for the candidate. Question construction requires an understanding of the cognitive level being tested, whether it be simply remembering a fact or analysing complex data. Bloom's taxonomy is one way to help to organise these thoughts.¹⁷ The best questions test application of knowledge rather than simply recall. A doctor who only recalls knowledge will help fewer patients than one who can apply, adapt and flex that knowledge to an individual situation. A question such as 'What is the most common leukaemia affecting children?' simply tests recall, whereas a question that describes the typical presentation of a child with leukaemia and asks 'What is the most likely diagnosis?' from a list of leukaemias ensures the candidate applies that knowledge to the specific scenario.

A higher level of question requires two or more mental steps to reach the answer. For example, a candidate might first have to determine the diagnosis before selecting an investigation or treatment. In the main, early-level undergraduate SBAs focus on the basic sciences and clinical diagnosis. Such assessments also tend to have more questions assessing simple recall only. Conversely, SBAs in the final years of medical school focus more on management plans after the candidate has synthesised the information within the stem.

Postgraduate SBAs tend to focus on management but the stem may have added complexities. For example, the patient whose history in an undergraduate question suggests type 2 diabetes mellitus, might, in a postgraduate examination have type 2 diabetes mellitus, hypertension, multiple sclerosis and tuberculosis, having just returned from Iraq. Furthermore, postgraduate SBAs tend to draw upon recent research and new changes in clinical guidance (eg, a subtle change in the British Thoracic Societies guideline on bronchiolitis), whereas undergraduate questions

draw more upon traditional treatment options (eg, the best management of hyperkalaemia).

SHARING THE LOAD

Writing large numbers of SBA examination questions can be a daunting task. It is challenging to cover all course learning outcomes with fixed format questions to be completed in a set time. Luckily, the approach to learning medicine is similar nationally and internationally, making collaboration in question generation fruitful and increasingly popular. For instance, in the UK, the Medical Schools Council Assessment Alliance has a collaborative shared bank of questions, from which most UK medical schools draw a proportion of their final examination written questions.¹⁸ Question setters who feel overwhelmed by gaps in their database may find intra-institutional or inter-institutional collaboration very helpful. Another option is to use or modify student-generated questions from free platforms such as PeerWise.¹⁴

CONCLUSION

SBA questions—if written well—can assess higher-order learning and discriminate between candidates of differing ability. Clinicians have an essential role in undergraduate and postgraduate assessments and their contribution towards producing high-quality questions will help to ensure reliable and valid written assessments. Writing questions is challenging, requires a careful and methodical approach honed through experience and an awareness of the approach of test-wise candidates.

Main messages

- ▶ Single best answer questions are widely used in medical assessment; well-written question items can effectively assess application of clinical and scientific knowledge.
- ▶ Question writing requires a systematic approach, with a consistent style, alphanumeric ordering of options, use of the cover tests and checking for question flaws that may hint towards the correct answer.
- ▶ Peer review is a key step for optimising question quality.
- ▶ Mapping each question to a learning objective will ensure questions remain germane to the curriculum and appropriate to the required level of learning.
- ▶ Detailed feedback is essential for enhancing learning from formative written questions.

Acknowledgements The authors thank Professor D A Harris for proofreading the article.

Contributors JLW and BHLH are joint first authors.

Competing interests JLW recently completed his academic foundation training in Wales. He is currently a research fellow in cardiology at the American University of Beirut Medical Centre, Lebanon. He is active in educational research, undergraduate and postgraduate medical teaching, writing and reviewing questions for medical school examinations and sits on the Programmes in Medical Education board at the American University of Beirut. He is co-founder of an online SBA question database for medical student revision. BHLH has recently completed his academic foundation training in Oxford. He teaches preclinical and clinical students in his roles as a lecturer at St Anne's College, Oxford and clinical teaching associate of Green Templeton College, Oxford. He is co-founder of an online SBA question database for medical student revision. PES is a consultant neurologist and sub-dean for Assessments at Cardiff University School of Medicine and is associate medical director for Quality for the Federation of Royal College of Physicians (UK).

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- 1 Case SM, Swanson DB. *Constructing written test questions for the basic and clinical sciences*. 3rd edn. Philadelphia: National Board of Medical Examiners, 2001.
- 2 Hays R. Assessment in medical education: roles for clinical teachers. *Clin Teach* 2008;5:23–7.
- 3 Swanson DB, Holtzman KZ, Allbee K. Measurement characteristics of content-parallel single-best-answer and extended-matching questions in relation to number and source of options. *Acad Med* 2008;83(10 Suppl):S21–4.
- 4 Chandratilake M, Davis M, Ponnampuruma G. Assessment of medical knowledge: the pros and cons of using true/false multiple choice questions. *Natl Med J India* 2011;24:225–8.
- 5 Haladyna TM, Downing SM. A taxonomy of multiple choice item-writing rules. *Appl Meas Educ* 1989;2:37–50.
- 6 Ware J, Vik T. Quality assurance of item writing: during the introduction of multiple choice questions in medicine for high stakes examinations. *Med Teach* 2009;31:238–43.
- 7 Case SM, Swanson DB, Becker DF. Verbosity, window dressing, and red herrings: do they make a better test item? *Acad Med* 1996;71(10 Suppl): S28–30.
- 8 Dacre J, Potts HW, Sales D, *et al*. The development of a new method of knowledge assessment: tailoring a test to a doctor's area of practice. *Acad Med* 2009;84:1003–7.
- 9 Malau-Aduli BS, Zimitat C. Peer review improves the quality of MCQ examinations. *Assess Eval High Educ* 2012;37:919–31.
- 10 Harris BH, Walsh JL, Tayyaba S, *et al*. A novel student-led approach to multiple-choice question generation and online database creation, with targeted clinician input. *Teach Learn Med* 2015;27:182–8.
- 11 Abdulghani HM, Ahmad F, Irshad M, *et al*. Faculty development programs improve the quality of Multiple Choice Questions items' writing. *Sci Rep* 2015;5:9556.
- 12 Butler AC, Roediger HL III. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem Cognit* 2008;36:604–16.
- 13 Walsh J, Harris B, Tayyaba S, *et al*. Student-written single-best answer questions predict performance in finals. *Clin Teach* 2016;13:352–6.
- 14 Walsh JL, Denny P, Smith PE. Encouraging maximal learning with minimal effort using PeerWise. *Med Educ* 2015;49:521–2.
- 15 Smith PE, Mucklow JC. Writing clinical scenarios for clinical science questions. *Clin Med (Lond)* 2016;16:142–5.
- 16 Shantikumar S, Watson M, Handa A. Does alphabetisation ensure randomisation of single best answer questions? *Clin Teach* 2010;7:136–7.
- 17 Bloom BS, Engelhart MD, Furst EJ, *et al*. *Taxonomy of educational objectives: the classification of educational goals. Handbook I: cognitive domain*. New York: David McKay Company, 1956.
- 18 Medical Schools Council website (cited 1 January 2016). <http://www.medschools.ac.uk/MSCAA/Pages/default.aspx>