Twelve tips for introducing very short answer questions (VSAQs) into your medical curriculum

Laksha Bala, Rachel J. Westacott, Celia Brown & Amir H. Sam

MEDICAL TEACHER

Taylor & Francis
Taylor & Francis Group

🔓 OPEN ACCESS Check for updates

Twelve tips for introducing very short answer questions (VSAQs) into your medical curriculum

Laksha Bala[a]* 🆔, Rachel J. Westacott[b]* 🆔, Celia Brown[c] 🆔 and Amir H. Sam[a] 🆔

[a]Imperial College School of Medicine, Imperial College London, London, UK; [b]Birmingham Medical School, University of Birmingham, Birmingham, UK; [c]Warwick Medical School, University of Warwick, Coventry, UK

ABSTRACT

Most undergraduate written examinations use multiple-choice questions, such as single best answer questions (SBAQs) to assess medical knowledge. In recent years, a strong evidence base has emerged for the use of very short answer questions (VSAQs). VSAQs have been shown to be an acceptable, reliable, discriminatory, and cost-effective assessment tool in both formative and summative undergraduate assessments. VSAQs address many of the concerns raised by educators using SBAQs including inauthentic clinical scenarios, cueing and test-taking behaviours by students, as well as the limited feedback SBAQs provide for both students and teachers. The widespread use of VSAQs in medical assessment has yet to be adopted, possibly due to lack of familiarity and experience with this assessment method. The following twelve tips have been constructed using our own practical experience of VSAQs alongside supporting evidence from the literature to help medical educators successfully plan, construct and implement VSAQs within medical curricula.

Introduction

Single best answer questions (SBAQs) are widely used to assess knowledge in written examinations. They have been shown to demonstrate high reliability and can be machine-marked efficiently (Norcini et al. 1984; Coughlin and Featherstone 2017). However, several concerns regarding the use of SBAQs have been raised by educators.

First, there are many situations in which a *single* best answer does not exist in medicine. Clinical uncertainty is inherent in medicine. Decisions around diagnosis and management are often nuanced and indeed even experts do not always agree on a single diagnosis or best course of action. Furthermore, as summarised by Surry et al. (2017), 'patients do not walk into the clinic saying, I have one of these five diagnoses, which do you think is most likely?' (p. 1082). Assessing students using an artificial situation whereby a patient presents with a list of five possible diagnoses is inauthentic and does not reflect the environment in which clinicians practice. Rather, clinicians formulate a list of possible diagnoses following assessment of a patient including taking a history, examining and considering available investigation results. Given the role assessment has in driving learning behaviours (Epstein 2007), it is essential that assessment methods encourage learning that prepares students for the realities of clinical practice.

Second, students have also been shown to perform better on SBAQs compared to very short answer questions (VSAQs) in multiple studies (Sam et al. 2016, 2018; Sam, Fung, et al. 2019; Sam, Peleva, et al. 2019; Sam, Westacott, et al. 2019) where the same knowledge is tested in both formats. This may be attributable to cueing and answer recognition behaviours (Newble et al. 1979; Schuwirth et al. 1996; Veloski et al. 1999; Shaibah and van der Vleuten 2013; Desjardins et al. 2014). This calls into question the validity of SBAQ assessments, which may be measuring students' ability to recognise the correct answer, rather than generate the answer themselves (Elstein 1993), providing a false impression of students' knowledge. In turn, students' learning approach in preparation for SBAQ assessments is likely to focus on superficial recognition rather than a deeper understanding of the subject being tested (Newble and Entwistle 1986; McCoubrie 2004; Willing et al. 2015) and retention of that learned information (Larsen et al. 2008).

Finally, the content tested using SBAQs may also become skewed from the 'optimal' blueprint as the question format can discourage question writing in areas where it is difficult to identify a sufficient number of plausible distractors. This, along with the premise that testing of core knowledge using an SBAQ format may be considered too easy (Elstein 1993; Fenderson et al. 1997), can result in the testing of obscure material.

A strong evidence base has now developed for the use of VSAQs as an alternative assessment tool in both formative and summative undergraduate assessments (Sam, Fung, et al. 2019; Sam, Peleva, et al. 2019; Sam, Westacott, et al. 2019). To the best of our knowledge, there has not been any published data on the use VSAQs in postgraduate or non-medical settings thus far. However, interest has been expressed regarding the introduction of VSAQs to postgraduate Royal College membership examinations for Physicians (Phillips et al. 2020) and Psychiatrists (Scheeres et al. 2022).

CONTACT Amir H. Sam ✉ a.sam@imperial.ac.uk 🖥 Imperial College School of Medicine, Charing Cross Campus, Reynolds Building, St Dunstan's Road, London W6 8RP, UK.
*Joint first author

VSAQs consist of a clinical scenario (which includes the presentation, examination findings, and investigation results, as necessary) and a lead-in question. Instead of the candidate selecting an option from a predetermined list, they must provide their own free-text answer. We recommend questions are constructed so that the answer required is one to five words in length to facilitate marking and reduce response burden for students.

VSAQs are considered a type of constructed-response or open-ended question. According to the constructivist theory of learning, constructed-response questions are thought to test higher-order cognitive processes, such as analysing/critiquing new information, and generating hypotheses/constructing knowledge (Cantillon 2010). Furthermore, assessments requiring students to produce (as opposed to recognise) knowledge are thought to promote better retention of learning (Larsen et al. 2008).

VSAQs have been shown to demonstrate higher reliability, discrimination, and authenticity compared to SBAQs (Sam et al. 2018). Students consistently score more highly in SBAQs compared to VSAQs, despite questions testing exactly the same knowledge base (Sam et al. 2016; Sam, Fung, et al. 2019; Sam, Peleva, et al. 2019; Sam, Westacott, et al. 2019). The difference in scores is attenuated if students sit SBAQs first, prior to answering the same questions in a VSAQ format. This suggests that VSAQs have a higher degree of validity in testing the ability to arrive at a correct answer without cueing or guessing (Sam et al. 2018).

Evidence also suggests that VSAQs encourage more authentic clinical reasoning strategies compared to SBAQs. For example, students are more likely to use analytical reasoning methods (specifically identifying key features) when answering VSAQs, whereas the use of test-taking behaviours (e.g. using answer options or 'buzz words' to help reach an answer) is more common for SBAQs. In addition, students acknowledge uncertainty more frequently when answering VSAQs (Sam et al. 2021).

The following twelve tips have been constructed using our own practical experience of VSAQs, alongside supporting evidence from the literature, to help medical educators successfully plan, construct and implement VSAQs within medical curricula.

VSAQ content and construction

Tip 1

Identify areas of applied knowledge that are best assessed by VSAQs

It is possible to use SBAQs to test applied knowledge in many areas of medical curricula, but not all. We recommend identifying those areas or learning outcomes where existing assessment instruments, such as SBAQs are not fit for purpose, which can sometimes lead to a lack of assessment of such outcomes. Examples include testing areas of core knowledge where it is difficult to find four plausible distractors to create an SBAQ (e.g. arterial blood gas analysis or prescribing as per Example 1) or where providing the correct answer allows for easy recognition of the correct response.

Example 1. A 22-year old man has acute breathlessness. He has a known history of asthma for which he takes regular beclomethasone and theophylline, and salbutamol as required. His temperature is 36.5 °C, pulse rate 95 bpm, BP 110/68 mmHg, respiratory rate 30 breaths per minute, and oxygen saturation 94% breathing air.

He is unable to complete sentences in one breath, and has a loud wheeze bilaterally. His peak flow is 35% of predicted. He is initially treated with supplementary oxygen, salbutamol *via* oxygen-driven nebuliser, and hydrocortisone 100 mg intravenously. A combination of salbutamol and ipratropium is then given, however his symptoms fail to improve significantly. The intensive care unit has been called to review the patient. He weighs 70 kg.

Please prescribe the most appropriate next medication (give drug name, dose, and route).

Accepted VSAQ answers: (Students have access to the British national formulary)

Magnesium sulphate 1.2 − 2 g intravenous over 20 min
Source: (Sam, Fung, et al. 2019)

VSAQs are ideally placed to test those areas in the curriculum which are inherently not amenable to the SBAQ format, such as prescribing. SBAQs test the ability to select a correct prescription out of a choice of five options, whereas VSAQs require students to generate details of the medication themselves (e.g. dose, route, frequency, etc.). The Prescribing Safety Assessment, a national examination taken by medical students in the UK that is being adopted in Canada, Australia, and New Zealand (Maxwell et al. 2015, 2017; Hardisty et al. 2019), does test this skill well, using a variety of computer-marked question formats. However, the examination is often taken in the last few months of the undergraduate medical course. It is therefore not able to identify gaps in prescribing knowledge early enough, nor does it provide the opportunity for longitudinal feedback for medical schools to be able to address deficiencies in prescribing knowledge and adjust course content to strengthen skills in these areas.

A prospective study analysing data from a pilot prescribing assessment showed that the median percentage score for the VSAQ test was significantly lower than the SBAQ test (28 vs. 64%, $p < 0.0001$). Significantly more prescribing errors were detected in the VSAQ format than the SBAQ format across all domains, notably in prescribing insulin (96.4 vs. 50.3%, $p < 0.0001$), fluids (95.6 vs. 55%, $p < 0.0001$), and analgesia (85.7 vs. 51%, $p < 0.0001$). The study demonstrated that prescribing VSAQs are an efficient tool for providing detailed insight into the sources of significant prescribing errors, which were not identified by SBA questions (Sam, Fung, et al. 2019).

Using VSAQs to assess prescribing is a valuable tool, particularly in the age of electronic prescribing, whereby it has become increasingly more difficult for students to practice prescribing in the workplace. This form of assessment can enhance students' skills in safe prescribing and potentially reduce prescribing errors.

Tip 2

Not all SBAQ vignettes work as a VSAQs

The resources required for the introduction of any new assessment need to be considered. One way of easily

generating VSAQs is to transform pre-existing SBAQs simply by removing the answer options. Most SBAQs that pass the cover test (i.e. can be answered solely from the stem of the question, without the need for answer options) can theoretically be turned into VSAQs. However, there are some situations where conversion of an SBAQ does not work. For example, SBAQs which ask students to choose the best of five options mentioned in the clinical vignette (e.g. five risk factors and five drugs) do not work well as VSAQs. Even if the five answer options are removed, the question still remains a best of five questions based on the information provided in the clinical vignette (see example 2).

Example 2. A 78-year-old man attends the Emergency Department following a collapse. He is dizzy on standing. He has osteoarthritis, gout, and benign prostatic hypertrophy. He is taking allopurinol, finasteride, ibuprofen, lansoprazole, and tamsulosin.

His temperature is 37.0 °C, pulse rate 84 bpm, BP 146/86 mmHg (lying), and 118/72 mmHg (standing), respiratory rate 12 breaths per minute and oxygen saturations 96% breathing air.

Investigations:

Haemoglobin	126 g/dL	(130 − 175)

Which medication is most likely to contribute to his current presentation?

A. Allopurinol
B. Finasteride
C. Ibuprofen
D. Lansoprazole
E. Tamsulosin

Correct SBAQ answer

Tamsulosin

Tip 3

Use authentic scenarios

When writing SBAQs, authentic clinical scenarios are often tailored to create four plausible distractors. This can make the clinical case contrived and reduce authenticity. By contrast, VSAQs can be constructed using authentic cases without alteration, since there is no requirement to create four plausible distractors.

VSAQs are also a useful assessment method for testing clinical reasoning under conditions of uncertainty, e.g. around diagnosis, investigation, or management. The constructed response format used in VSAQs allows for the acknowledgement and management of clinical uncertainty, which SBAQs limit through the provision of answer options. However, VSAQs still need to be written with care to ensure the uncertainty they create is related to the clinical problem rather than the question construction itself.

In Example 3, the lack of information in the clinical vignette regarding the location of the patient, may lead to uncertainty regarding management options. For example, in a General Practice clinic, a same day blood test result (i.e. repeat urea and electrolytes) would not necessarily be available. Furthermore, if the setting were an inpatient ward

where an ECG would be readily available, an electrocardiogram (ECG) is not an unreasonable alternative answer.

Example 3. A 47-year old man with hypertension attends for annual review. He takes ramipril (10 mg once daily). His BP is 138/78 mmHg.

Investigations:

Sodium	139 mmol/L	(135–146)
Potassium	6.2 mmol/L	(3.5–5.3)
Urea	5.0 mmol/L	(2.5–7.8)
Creatinine	90 µmol/L	(60–120)

Which is the most appropriate immediate action?

Accepted VSA answers

Repeat urea and electrolytes
Source: (Putt et al. 2022)

Tip 4

VSAQs require specific lead-ins

To further ensure that any uncertainty generated by the VSAQ is authentic and related to the clinical problem, as opposed to the construction of the VSAQ, question writers must ensure the lead-in is specific and tests a particular and discrete area of knowledge. This in turn avoids uncertainty being generated about what the question is asking.

For example, the lead-in 'what is the diagnosis?,' may not be specific enough. In a VSAQ concerning proteinuria, there may be uncertainty as to whether an overarching diagnosis is being asked for (i.e. nephrotic syndrome) or the underlying aetiology specific to the patient (e.g. amyloidosis). A more specific lead-in could be, 'what diagnosis is most likely to be confirmed with renal biopsy?'

This point is also illustrated in Example 4. The lead-in does not specify whether students should offer an appropriate medication or non-pharmacological management. A more specific lead-in could be, 'which class of pharmacological treatment would be most appropriate to manage this patient?'

Example 4. A 75-year-old woman is reviewed 4 days after a fractured neck of femur repair. She has been agitated and upset, particularly at night. She has punched nurses and keeps trying to leave the ward. She has seen strange men in black capes entering the ward and believes that they are controlling the hospital. When she was seen in the memory clinic 6 months ago, she was found to have mild cognitive impairment.

What is the most appropriate management?

Possible VSAQ answers (list not exhaustive)

Move patient to quiet, well-lit side room
Continuity of care from healthcare staff
Encourage visits from family and/or friends
Provide adequate analgesia

Keep a stool chart and treat constipation if needed
Haloperidol
Olanzapine

Tip 5

VSAQs require a defined range of correct answers

As demonstrated in Example 4 regarding the management of delirium, several answers for a given VSAQ may be correct, especially with regards to clinical management. This of course is not an issue for SBAQs, where this can be controlled by the provision of answer options.

The fact there may be no one single best answer to a given question is an authentic representation of clinical medicine. This notion should be embraced and highlighted to students in their preparation for professional practice. However, in order to make the marking process feasible, time-efficient, and semi-automated, it is important to use the clinical vignette and lead-in to guide students towards a defined range of correct answers. Furthermore, a lead-in that allows many different correct answers (as in Example 4) will also likely reduce the question's discrimination, as the area of knowledge being assessed becomes less well defined (e.g. management of delirium vs. indication for haloperidol).

It is important to note that if the assessors are happy to accept a broader range of answers, recognising that marking will be much more time and resource-intensive, then VSAQs do not necessarily need to be directive in this manner. Assessors should carefully consider their objectives, resource availability (i.e. time and assessors available for marking), and what specifically they are trying to assess.

The recommendation of having answers that are five words or less also allows for a more stringent marking scheme and elimination of inter-marker subjectivity, which can be a problem in other forms of free text examinations (Sam et al. 2016). Evidence has shown that for prescribing VSAQs, answers can be further guided by dropdown menus. For example, students can enter the medication name and dose in two separate free text fields, but select route and frequency from two separate dropdown menus (Sam, Fung, et al. 2019), thereby limiting the variability of responses.

Of course, one of the advantages of VSAQs is the flexibility offered in students' answers. If a student provides a viable alternative answer which was not previously considered, this can still be awarded a mark in a VSAQ, whereas this would not be possible in a SBAQ.

VSAQ implementation

Tip 6

Consider your marking strategy

One of the barriers to using free-text answer questions in assessment has been the time required for marking. Online examination management software now allows VSAQ assessments to be conducted on iPad tablets or fixed terminal computers and for marking to be semi-automated. Evidence from an assessment containing 60 VSAQs and involving 299 students showed that the total time taken to review the machine-marked answers by two clinicians was 95 min, 51 s (1 min, 36 s per question) (Sam et al. 2018).

All identical responses to VSAQs can be grouped in blocks by examination software, and then machine marked using an automated matching algorithm. This compares the student's answer against a set of preapproved acceptable answers created by subject experts. The algorithm compares each student response to the correct answers using a measure called Levenshtein distance (Levenshtein 1966). All student answers that are identical to the list of approved answers are automatically marked as correct.

All match failures are highlighted by the software. Processes for reviewing these responses can either involve a panel discussion, or review by two clinicians simultaneously (with or without a third individual acting as an arbitrator) to establish whether any of the non-exact matches should be allowed as correct answers. Marks for responses deemed correct by the examiners are awarded manually, but the examination software also permits answers marked manually as correct to be added to the correct answer database for future use (Sam et al. 2018).

In one study, it was shown that based on the preloaded acceptable answers, the system was able to identify 80.2% of correct answers prior to review. Of answers marked correct by the system, 0.2% was deemed to be incorrect on review, because a spelling error significantly changed the meaning of the answer. In 8.3% of questions at least one student offered an alternative answer to the question that was judged to be correct following the review by examiners (Sam et al. 2018).

Things to consider when deciding on your marking strategy for VSAQs include:

- How close to the correct answer students need to be (e.g. using Levenshtein distance) to account for misspelling, but not allow clinically incorrect answers (e.g. difference between hypo and hyper). This may depend on the type of questions contained within the assessment for example, misspelling 'apendicitis' instead of appendicitis may be acceptable but misspelling a drug name may not.
- Whether half-marks are allowed for partially correct responses. This requires explicit marking schema, will add to marking time and is more likely to require moderation to ensure consistency.
- Accuracy of human marking – how many markers will mark the items that are not machine marked? Should there be an arbitrator or a second review? If VSAQs are used for high stakes assessments, then marking accuracy is paramount.
- Post-hoc review process – a review of question and candidate performance is advisable (particularly for summative assessments) including processes to manage poorly performing questions (are these removed?) or moderation of responses if the review feels that additional correct responses should be allowed.

Example 5 demonstrates a binary marking system. Light grey shading shows answers that are automatically marked as correct based on the preapproved answers. The unshaded answers have been marked as correct based on their similarity to preapproved answers. Answers marked as incorrect are shown in darker shading; however, during the review process this can be overridden (e.g. primary

hypoadrenalism) and all identical answers automatically given the same mark (Sam et al. 2018).

Example 5. A 24-year-old woman reports 2 months of lethargy, dizziness, weight loss, and nausea. She has type 1 diabetes and reports erratic blood sugars and one episode of loss of consciousness. She has hyperpigmentation in her palmar creases and her oral mucosa. Her temperature is 36.8 °C, pulse rate 101 bpm, blood pressure 78/61 mmHg (standing), respiratory rate 16 breaths minute, and oxygen saturation 99% breathing air. Her capillary blood glucose is 3.2 mmol/L.

Investigations:
Sodium: 129 mmol/L (135–146)
Potassium: 5.4 mmol/L (3.4–5.0)
Urea: 7.7 mmol/L (2.5–7.8)
Creatinine: 67 μmol/L (50–95)
What is the most likely diagnosis?

Accepted VSAQ answers (1.00), unacceptable answers (0.00)

Addison's disease	1.00
Primary adrenal insufficiency	1.00
Addisons disease	1.00
SIADH	0.00
Primary hypoadrenalism	1.00

Source: (Sam et al. 2018)

Tip 7

Give standard setters access to acceptable answers

Answers to VSAQs should be provided to standard setters to help them arrive at a valid standard and appropriate paper pass mark. A study conducted by Sam et al. (currently under review) demonstrated that standard setters produced a significantly lower pass mark for a VSAQ paper when they had access to the range of acceptable answers, compared to standard setting without access to any answers. By contrast, providing access to SBAQs did not make a significant difference to the pass mark set (Sam et al. 2022). Previous literature suggests that access to the answers for SBAQs can even result in standard setters underestimating the difficulty of a question (Verheggen et al. 2008).

Providing access to answers may influence standard setters' perception of VSAQ difficulty and provide a clearer indication of the degree of difficulty of the question. Standard setters may develop a greater appreciation of the increased difficulty of VSAQs compared to SBAQs, which is reflected in lower pass marks and in students' poorer performance on VSAQs compared to SBAQs (Sam, Westacott, et al. 2019).

In addition to providing answers for the VSAQs, detailed training should be made available for standard setters to ensure familiarity with the standard required to pass (especially the fact that this will generally be lower than for an equivalent SBAQ paper), the format of VSAQs and with the standard setting method used (e.g. Angoff or Ebel).

Tip 8

Prepare your students for VSAQs

When introducing a novel assessment to students, the process needs to be made as transparent as possible so that there is no additional anxiety born from unfamiliarity with the assessment method. This will help ensure that the assessment is useful and gives students the best chance to demonstrate their knowledge.

Important areas to cover include:

- Understanding what a VSAQ is and the added value of this method of assessment (e.g. in real clinical practice a patient does not present with five options to choose from and there often is not a single best answer).
- Advising students how to approach questions – e.g. analytical reasoning and avoiding test-taking behaviours students may have developed for SBAQs.
- Advising students how to answer questions including the word limit, whether the programme uses dropdown menus (e.g. for prescribing route and frequency), the need to be committed to their answer (no answers such as ?pneumonia) and the degree of specificity required in their answers, e.g. name the drug and give the dose if requested (e.g. for adrenaline).
- Providing students with the opportunity to answer practice questions to familiarise themselves with any software used and get feedback on their answers. This is particularly important prior to use in summative assessments.

Tip 9

Use VSAQ answers to give students feedback

Another significant advantage of VSAQs compared with SBAQs is the rich feedback provided by student responses. Student responses can highlight errors in judgement that are not necessarily identified by SBAQs. Their responses are frequently different to the incorrect distractors provided by a question author for an SBAQ who has a pre-conceived idea of how a question is likely to be answered incorrectly.

For example, in a study looking at prescribing VSAQs, examiners were able to identify that some students prescribed large doses of rapid-acting insulin for a hyperglycaemia scenario, which in clinical practice would be a serious prescribing error. The same questions in SBAQ format would not have yielded this level of detail in its feedback (Sam, Fung, et al. 2019). This information can in turn be used to provide specific feedback to students regarding the rationale for correct prescribing.

Where feedback is provided for students after an assessment, it can be provided either on an individual level (e.g. students can review their individual performance by logging into the exam platform) or by class-wide feedback on common errors.

Tip 10

Use VSAQ answers to give teachers feedback

In addition to providing feedback for students, VSAQ answers also identify cognitive processes and the basis for student errors. This can help to improve our understanding

about how students approach clinical problems and can therefore provide feedback to teachers about students' clinical knowledge, the concepts or reasoning that they find difficult and where they are prone to making errors (Sam et al. 2021). The use of think-aloud methods whereby students talk through their thought processes when answering VSAQs can offer further insight into cognitive processes used by students (e.g. analytical or test-taking behaviours) and provide rich information for both student and teacher feedback (Sam et al. 2021).

This insight into student misconceptions, cognitive approaches, and errors, may highlight areas of the curriculum which require further investment and development, providing an opportunity to shape and improve clinical teaching (Sam et al. 2021). Such teaching should include approaches to analytical reasoning, as well as increasing awareness of the potential mechanism for cognitive error (Sam et al. 2021). Areas identified through VSAQ answers may represent parts of the curriculum (so-called blind spots) which faculty may not be aware require improvement. These insights cannot be gained through the analysis of incorrect SBAQ answers because students must select from a list of pre-defined answer options, none of which may have been the student's initial response to the question (Putt et al. 2022).

Importantly, several VSAQs have been shown to highlight significant cognitive errors, which were not apparent in their SBAQ counterparts, or indeed even considered as possible student responses by the person authoring the question.

Example 6. A 60-year-old man has a swollen, painful right leg for 2 days. He has a history of hypertension and takes ramipril. He is otherwise well.

He has a swollen right leg. The remainder of the examination is normal.

Investigations:

Haemoglobin	140 g/L	(130–175)
White cell count	8.0×10^9/L	(3.8–10.0)
Platelets	340×10^9/L	(150–400)
Creatinine	94 µmol/L	(60-120)
Calcium	2.5 mmol/L	(2.2–2.6)
ALT	30 IU/L	(10–50)
ALP	99 IU/L	(25–115)
APTT	30 s	(22–41)
PT	12 s	(10–12)

Urinalysis: normal
Chest X-ray: normal
Venous duplex ultrasound scan: thrombus in superficial femoral vein
Which is the most appropriate additional investigation?

Accepted VSAQ answers

CT of abdomen and pelvis
Source: (Sam, Fung, et al. 2019; Sam, Peleva, et al. 2019; Sam, Westacott, et al. 2019)

This is demonstrated in Example 6, where despite a venous thromboembolism being confirmed, therefore rendering a D-dimer irrelevant, 27% of students still chose this option in the VSAQ. Of further concern, 34% of students

would have ordered a CT pulmonary angiogram (CTPA) in a patient with no respiratory symptoms or signs, thereby exposing the patient to a significant dose of unnecessary radiation without any likely therapeutic benefit. It is also possible that further investigation to exclude an occult malignancy would not have been instituted (Sam, Westacott, et al. 2019).

Furthermore, a recently published analysis identified four predominant types of error: inability to identify the most important abnormal value (e.g. carboxyhaemoglobin in suspected carbon monoxide poisoning), over or unnecessary investigation (e.g. CTPA in Example 6 above), lack of specificity of radiology requesting (e.g. requesting an ultrasound scan without specifying which anatomical site to ultrasound) and over-reliance on trigger words (e.g. reading 'sudden onset headache' and assuming the patient has a subarachnoid haemorrhage without reading the rest of the vignette which points to another diagnosis) (Putt et al. 2022). Hence, analysing incorrect answers enables identification of common cognitive errors and provides greater insight into students' knowledge and understanding, which could be used to guide future teaching.

Tip 11

Use student VSAQ answers as plausible distractors for SBAQs
In addition to providing feedback for students and educators, students' VSAQ responses can even be useful for the construction of SBAQs. Students' responses to VSAQs provide a unique opportunity to learn about responses which are real distractors for students, which providing they are plausible, can in turn be used as distractors for an SBAQ. Indeed, the ability of an SBAQ to accurately test knowledge is affected by the quality of the incorrect options (distractors), as one or more implausible distractors will likely increase the facility of the question. Identifying four plausible distractors for SBAQs is not always easy but if students can inadvertently provide this from a VSAQ, the quality of SBAQs may also be improved.

Tip 12

Use VSAQs to enhance assessment for learning
Evidence from a study using VSAQs in formative team-based learning (TBL) sessions, supports the use of VSAQs over SBAQs for student learning (Millar et al. 2021). The study demonstrated that most students thought VSAQs were a better representation of how they would be expected to answer questions in clinical practice compared to SBAQs and the sessions helped improve their preparation for clinical practice.

The study also showed that using VSAQs in a formative TBL session helped to emphasise group discussions. Group discussion has been shown to be a strong activator of prior knowledge and helps to establish students' understanding of a topic. Students perceive a valuable part of TBL to be the inter-learner discussions, allowing them to hear peer answer explanations and aiding recognition of mistakes in their own understanding (Ho 2019). The use of VSAQs in TBL may therefore enrich group discussions when there are multiple plausible answers in a clinical scenario that need

to be debated, thereby enhancing the learning potential of the session. Evidence also showed that students had a wider range of attempts to reach the correct answer in VSAQs compared to SBAQs, which may have led to enriched team discussions during TBL, as the focus of discussion was not limited to the five options available in a SBAQ (Millar et al. 2021).

Using VSAQs as assessment for learning in a formative setting also provides an opportunity to identify and address any misconceptions in students' understanding early (i.e. before high-stakes summative examination). The use of VSAQs for student learning in the formative setting is amenable to not only TBL, but all teaching modalities (small group tutorials, lectures, etc.).

Conclusion

In this article, we have presented key considerations required for the successful planning, construction, and implementation of VSAQ assessments. We have discussed the evidence base and educational theory supporting the use of VSAQ as a more authentic test of applied knowledge with higher reliability, discrimination, and authenticity compared to SBAQs. In particular, we have highlighted how VSAQs provide a unique insight into what students actually know, which in turn can be used to provide student and faculty feedback and shape the undergraduate medical curriculum. Ultimately, through careful construction and implementation of VSAQs, we hope to better prepare our students for clinical practice and for delivering high quality, safe patient care.

Acknowledgements

Disclosure statement

Funding

Notes on contributors

Laksha Bala, MRCGP, FHEA, Research Fellow in Medical Education at Imperial College School of Medicine, London, UK.

Rachel J. Westacott, MB ChB, FRCP, Associate Professor, Birmingham Medical School, University of Birmingham, Birmingham, UK.

Celia Brown, PhD, SFHEA, Professor of Medical Education, University of Warwick, Coventry, UK.

Amir H. Sam, PhD, FRCP, SFHEA, Head of Imperial College School of Medicine, London, UK.

ORCID

Laksha Bala (iD) http://orcid.org/0000-0002-8242-379X
Rachel J. Westacott (iD) http://orcid.org/0000-0001-9846-1961
Celia Brown (iD) http://orcid.org/0000-0002-7526-0793
Amir H. Sam (iD) http://orcid.org/0000-0002-9599-9069

References

Cantillon PWD. 2010. ABC of learning and teaching in medicine. 2nd ed. Malden (MA): Blackwell Publishing Ltd. (ABC Series).

Coughlin PA, Featherstone CR. 2017. How to write a high quality multiple choice question (MCQ): a guide for clinicians. Eur J Vasc Endovasc Surg. 54(5):654–658.

Desjardins I, Touchie C, Pugh D, Wood TJ, Humphrey-Murto S. 2014. The impact of cueing on written examinations of clinical decision making: a case study. Med Educ. 48(3):255–261.

Elstein AS. 1993. Beyond multiple-choice questions and essays: the need for a new way to assess clinical competence. Acad Med. 68(4):244–249.

Epstein RM. 2007. Assessment in medical education. N Engl J Med. 356(4):387–396.

Fenderson BA, Damjanov I, Robeson MR, Veloski JJ, Rubin E. 1997. The virtues of extended matching and uncued tests as alternatives to multiple choice questions. Hum Pathol. 28(5):526–532.

Hardisty J, Davison K, Statham L, Fleming G, Bollington L, Maxwell S. 2019. Exploring the utility of the Prescribing Safety Assessment in pharmacy education in England: experiences of pre-registration trainees and undergraduate (MPharm) pharmacy students. Int J Pharm Pract. 27(2):207–213.

Ho CLT. 2019. Team-based learning: a medical student's perspective. Med Teach. 41(9):1087.

Larsen DP, Butler AC, Roediger HL. 3rd. 2008. Test-enhanced learning in medical education. Med Educ. 42(10):959–966.

Levenshtein V. 1966. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady. 10:707–710.

Maxwell SRJ, Cameron IT, Webb DJ. 2015. Prescribing safety: ensuring that new graduates are prepared. Lancet. 385(9968):579–581.

Maxwell SRJ, Coleman JJ, Bollington L, Taylor C, Webb DJ. 2017. Prescribing safety assessment 2016: delivery of a national prescribing assessment to 7343 UK final-year medical students. Br J Clin Pharmacol. 83(10):2249–2258.

McCoubrie P. 2004. Improving the fairness of multiple-choice questions: a literature review. Med Teach. 26(8):709–712.

Millar KR, Reid MD, Rajalingam P, Canning CA, Halse O, Low-Beer N, Sam AH. 2021. Exploring the feasibility of using very short answer questions (VSAQs) in team-based learning (TBL). Clin Teach. 18(4):404–408.

Newble DI, Baxter A, Elmslie RG. 1979. A comparison of multiple-choice tests and free-response tests in examinations of clinical competence. Med Educ. 13(4):263–268.

Newble DI, Entwistle NJ. 1986. Learning styles and approaches: implications for medical education. Med Educ. 20(3):162–175.

Norcini JJ, Swanson DB, Grosso LJ, Shea JA, Webster GD. 1984. A comparison of knowledge, synthesis, and clinical judgment. Multiple-choice questions in the assessment of physician competence. Eval Health Prof. 7(4):485–499.

Phillips G, Jones M, Dagg K. 2020. Restarting training and examinations in the era of COVID-19: a perspective from the Federation of Royal Colleges of Physicians UK. Clin Med. 20(6):e248–e252.

Putt O, Westacott R, Sam AH, Gurnell M, Brown CA. 2022. Using very short answer errors to guide teaching. Clin Teach. 19:100–105.

Sam AH, Field SM, Collares CF, van der Vleuten CPM, Wass VJ, Melville C, Harris J, Meeran K. 2018. Very-short-answer questions: reliability, discrimination and acceptability. Med Educ. 52(4):447–455.

Sam AH, Fung CY, Wilson RK, Peleva E, Kluth DC, Lupton M, Owen DR, Melville CR, Meeran K. 2019. Using prescribing very short answer questions to identify sources of medication errors: a prospective study in two UK medical schools. BMJ Open. 9(7):e028863.

Sam AH, Hameed S, Harris J, Meeran K. 2016. Validity of very short answer versus single best answer questions for undergraduate assessment. BMC Med Educ. 16(1):266.

Sam AH, Peleva E, Fung CY, Cohen N, Benbow EW, Meeran K. 2019. Very short answer questions: a novel approach to summative assessments in pathology. Adv Med Educ Pract. 10:943–948.

Sam AH, Westacott R, Gurnell M, Wilson R, Meeran K, Brown C. 2019. Comparing single-best-answer and very-short-answer questions for the assessment of applied medical knowledge in 20 UK medical schools: cross-sectional study. BMJ Open. 9(9):e032550.

Sam AH, Wilson R, Westacott R, Gurnell M, Melville C, Brown CA. 2021. Thinking differently - Students' cognitive processes when answering two different formats of written question. Med Teach. 43(11): 1278–1285.

Sam AH, Millar KR, Westacott R, Melville CR, Brown CA. 2022. Standard setting very short answer questions (VSAQs) relative to single best answer questions (SBAQs): does having access to the answers make a difference? (under review)

Scheeres K, Agrawal N, Ewen S, Hall I. 2022. Transforming MRCPsych theory examinations: digitisation and very short answer questions (VSAQs). BJPsych Bull. 46(1):52–56.

Schuwirth LW, van der Vleuten CP, Donkers HH. 1996. A closer look at cueing effects in multiple-choice questions. Med Educ. 30(1):44–49.

Schuwirth LW, van der Vleuten CP, Stoffers HE, Peperkamp AG. 1996. Computerized long-menu questions as an alternative to open-ended questions in computerized assessment. Med Educ. 30(1): 50–55.

Shaibah HS, van der Vleuten CP. 2013. The validity of multiple choice practical examinations as an alternative to traditional free response examination formats in gross anatomy. Anat Sci Educ. 6(3):149–156.

Surry LT, Torre D, Durning SJ. 2017. Exploring examinee behaviours as validity evidence for multiple-choice question examinations. Med Educ. 51(10):1075–1085.

Veloski JJ, Rabinowitz HK, Robeson MR, Young PR. 1999. Patients don't present with five choices: an alternative to multiple-choice tests in assessing physicians' competence. Acad Med. 74(5):539–546.

Verheggen MM, Muijtjens AM, Van Os J, Schuwirth LW. 2008. Is an Angoff standard an indication of minimal competence of examinees or of judges? Adv Health Sci Educ Theory Pract. 13(2):203–211.

Willing S, Ostapczuk M, Musch J. 2015. Do sequentially-presented answer options prevent the use of testwiseness cues on continuing medical education tests? Adv in Health Sci Educ. 20(1):247–263.