

Using very short answer errors to guide teaching

Oliver Putt¹ | Rachel Westacott²  | Amir H. Sam³  | Mark Gurnell⁴  |
Celia A. Brown¹ 

¹Division of Health Sciences, University of Warwick, Coventry, UK

²College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

³Imperial College School of Medicine, Imperial College London, London, UK

⁴Wellcome-MRC Institute of Metabolic Science & NIHR Cambridge Biomedical Research Centre, University of Cambridge & Addenbrooke's Hospital, Cambridge, UK

Correspondence

Celia A. Brown, Division of Health Sciences, University of Warwick, Coventry, UK.
Email: celia.brown@warwick.ac.uk

Funding information

National Institute for Health Research (NIHR) Applied Research Collaboration (ARC) West Midlands; National Institute for Health Research Cambridge Biomedical Research Centre; Medical Schools Council Assessment Alliance

Abstract

Background: Student performance in examinations reflects on both teaching and student learning. Very short answer questions require students to provide a self-generated response to a question of between one and five words, which removes the cueing effects of single best answer format examinations while still enabling efficient machine marking. The aim of this study was to pilot a method of analysing student errors in an applied knowledge test consisting of very short answer questions, which would enable identification of common areas that could potentially guide future teaching.

Methods: We analysed the incorrect answers given by 1417 students from 20 UK medical schools in a formative very short answer question assessment delivered online.

Findings: The analysis identified four predominant types of error: inability to identify the most important abnormal value, over or unnecessary investigation, lack of specificity of radiology requesting and over-reliance on trigger words.

Conclusions: We provide evidence that an additional benefit to the very short answer question format examination is that analysis of errors is possible. Further assessment is required to determine if altering teaching based on the error analysis can lead to improvements in student performance.

1 | BACKGROUND

Single best answer (SBA) examinations are frequently used to assess knowledge in medicine. Studies have shown that some cognitive processes used to answer SBAs are different to those used in real-life scenarios.^{1,2} Very short answer (VSA) questions were developed to provide a more thorough test of knowledge as described by Sam et al.³ VSA examinations provide a clinical vignette with an associated question from which a student must generate an answer consisting of between one and five words. This format can be delivered online with predominantly rapid electronic marking using pre-specified answers. With manual marking kept to a minimum, the cost and time incurred by more open-ended questions are reduced, while the benefits of free-response answers are retained.

While feedback on examinations is sought-after by students^{4,5} and it has been shown that feedback can improve student performance,^{6,7} we were only able to identify one PUBMED-indexed paper that used student examination performance to improve faculty teaching.⁸ If clinical teachers want to enhance their teaching, one approach is for them to seek feedback on student performance. Because of the free text answers of VSA questions, we hypothesised that common incorrect answers could highlight areas of misunderstanding, providing an insight into student misconceptions and hence opportunities for improving clinical teaching. Such insights cannot be gained through the analysis of incorrect SBA 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.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *The Clinical Teacher* published by Association for the Study of Medical Education and John Wiley & Sons Ltd.

Our aim was to consider, with examples, if an analysis of common incorrect answers in VSA format examinations was possible and if it could identify common areas of misunderstanding amongst students. Our ultimate hope would be that by informing clinical teachers of the common errors identified, it could help them reflect on and improve their teaching.

2 | METHODS

We undertook a secondary analysis of data from a study comparing VSA questions and SBA questions.⁹ All UK medical schools with final year students ($N = 32$) were invited to participate between September and November of 2018, and 20 schools enrolled. The students were invited to sit a formative examination. The examination consisted of 50 questions based on the core knowledge required of a medical graduate.¹⁰ The examination was taken twice in succession, first in VSA format and second in SBA format. The VSA format required students to write a one- to five-word answer in response to a clinical vignette (see Examples 1 to 4 in Figure 1). The SBA version gave five options, one of which was the best answer. Information regarding the study was given to the students prior to enrolment. Participation in the examination was deemed as informed consent. Ethical approval was provided by the Medical Education Ethics Committee (Reference MEEC-1718-100) at Imperial College London.

Incorrect VSA responses for each question were collated and reviewed by a single researcher who performed a thematic analysis of errors. Incorrect answers for the SBA questions were not analysed in this secondary analysis. Types and range of error varied between questions; thus, categorisation was subjective and question specific. The processed data were subsequently analysed and categorised by four researchers: three practicing doctors of mixed clinical experience and one lay person. Categorisation was achieved through discussion of errors identified and if/how they could be improved through altered teaching.

3 | FINDINGS

The 42,670 incorrect answers to the 50 VSAs given by 1417 final year students from 20 medical schools within the UK were analysed. The number of students participating per medical school ranged from 3 to 256.

A variety of errors were observed. Many were attributed to a lack of knowledge. Lack of knowledge errors were specific to the question, making generalisable lessons difficult; however, errors were also identified that we believe could be preventable by changing teaching practices. Four types of error were identified in multiple questions, which could be used to inform teaching and learning. These errors were as follows:

1. Inability to identify the most important abnormal value.
2. Over or unnecessary investigation.

3. Lack of specificity of radiology requesting.
4. Over-reliance on trigger words.

Four types of error were identified in multiple questions, which could be used to inform teaching and learning. These errors were as follows:

1. *Inability to identify the most important abnormal value.*
2. *Over or unnecessary investigation.*
3. *Lack of specificity of radiology requesting.*
4. *Over-reliance on trigger words.*

Examples of each type of error—related to the questions shown in Figure 1—are further described below. Table 1 summarises the most common incorrect responses to each question.

3.1 | Inability to identify the most important abnormal value (Example 1)

Clinicians must identify key investigations required for diagnosis and management. Many candidates answered with either an incorrect parameter or something that was not a parameter (e.g., arterial blood gas [ABG] and carbon monoxide). While modern ABG analysers routinely provide carboxyhaemoglobin results, it is important that clinicians can understand and explain both what is being measured and its significance.

3.2 | Unnecessary or over-investigation (Example 2)

Judicious use of investigations is important for patient safety and economic reasons. Unprovoked deep vein thrombosis (DVT) in a 60-year-

Example 1: Inability to identify the most important abnormal value

A 36 year old man is rescued from a house fire.

He is alert and talking but has a dull headache. His pulse rate is 98 bpm, BP 139/86 mmHg, respiratory rate 22 breaths per minute and oxygen saturation 100% breathing 15 L/min oxygen via a non-rebreather mask.

Which is the most appropriate parameter to measure?

Example 2: Unnecessary or over-investigation

A 60 year old man has 2 days of a swollen, painful right leg. 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	$8.0 \times 10^9/L$	(3.8–10.0)
Platelets	$340 \times 10^9/L$	(150–400)
Creatinine	94 $\mu\text{mol/L}$	(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 seconds	(22–41)
PT	12 seconds	(10–12)

Urinalysis: normal

Chest X-ray: normal

Venous duplex ultrasound scan: thrombus in superficial femoral vein.

Which is the most appropriate additional investigation?

Example 3: Lack of specificity of radiology requesting

A 28 year old woman presents to her GP with a neck lump that she noticed incidentally when rubbing her neck.

There is a smooth, non-tender 1.5 cm mobile lump within the thyroid gland.

Investigations:

TSH	2.3 mU/L	(0.3–4.2)
Free T4	17 pmol/L	(9–25)

Which is the most appropriate next investigation?

Example 4: Over-reliance on trigger words

A 72 year old woman has 6 months of increasing pain at the base of her right thumb. She is having difficulty opening jars and sewing. She is otherwise well. No other joints are painful. She is taking regular analgesia.

The first carpometacarpal joint is swollen and tender, with reduced opposition of the thumb.

Which is the most likely diagnosis?

FIGURE 1 Example questions from the very short answer paper

old man should prompt a clinician to look for an underlying malignancy. A significant number of candidates requested a CT pulmonary angiogram (CTPA) or D-dimer as described in our previous study.⁹ The

D-dimer is not helpful in a patient with confirmed DVT, and a CTPA is not indicated in the absence of clinical features of pulmonary embolism.

TABLE 1 Question details and findings

Example question	Correct answer ^a	% Students answering incorrectly (from N = 1417)	Three most common errors (frequency as % of incorrect answers from previous column)	Learning point for clinical teachers
Example 1: Inability to identify the most important abnormal value	<i>Carboxyhaemoglobin</i>	89.3%	<ol style="list-style-type: none"> 1. Observations (29.2%) (e.g., heart rate and blood pressure) 2. Carbon monoxide (28.5%) 3. Arterial blood gas (26.5%) 	Teach students how to identify which investigation result is most important to guide diagnosis or treatment when presented with a list of investigation results.
Example 2: Unnecessary or over-investigation	CT of abdomen and pelvis	97.7%	<ol style="list-style-type: none"> 1. CT pulmonary angiogram (35.2%) 2. D-Dimer (27.8%) 3. ECG (7.7%) 	Teach students about the judicious and step-wise use of investigations alongside the avoidance of unnecessary tests.
Example 3: Lack of specificity of radiology requesting	<i>Ultrasonography of neck</i>	53.4%	<ol style="list-style-type: none"> 1. Imaging—location not specified (31.6%) 2. FNA (18.8%) 3. Thyroid scintigraphy (15.1%) 	Teach students how to be specific when requesting radiology investigations, including requesting the modality and anatomical region for all types of imaging.
Example 4: Over-reliance on trigger words	<i>Osteoarthritis</i>	39.5%	<ol style="list-style-type: none"> 1. Gout (28.6%) 2. Rheumatoid arthritis (20.4%) 3. Fracture (12.3%) 	Teach students to avoid using trigger words as the main determinant of pattern recognition and to be cognisant of when they are using pattern recognition to ensure they consider alternative hypotheses prior to 'closure'.

^aNote that other answers would have been accepted as correct had the context of the answer remained the same.

3.3 | Lack of specificity of radiology requesting (Example 3)

Clinicians often request radiological investigations that require asking for the correct test. Marks were frequently lost because candidates wrote vague responses such as 'imaging' or 'ultrasound' without specifying what type of scan or which anatomical region should be investigated.

3.4 | Over-reliance on trigger words (Example 4)

These errors resulted from candidates incorrectly focusing on trigger words and instigating erroneous pattern recognition. The example shows how focusing on 'base of thumb pain' without considering the additional information provided has resulted in an incorrect diagnosis of gout or a scaphoid fracture.

4 | DISCUSSION

There is minimal literature discussing the impact of using student errors to alter our clinical teaching practices. Analysis of SBA

questions is challenging because cueing can mask students' misunderstanding. VSA questions were developed to create a more authentic measure of knowledge while allowing electronic delivery and predominantly machine marking. In this study, we have shown that analysis of errors is possible and affords greater insight into students' knowledge and understanding.

In this study, we have shown that analysis of errors is possible and affords greater insight into students' knowledge and understanding.

Reviewing the incorrect answers given in a large-scale VSA question examination generated some general and repeated errors. If

these errors were made in real life, they could impact patient safety, so addressing them is not merely about teaching 'to the test'. Providing information to clinical teachers about errors could therefore serve as a prompt for reflection. Our reflections on these errors helped us to identify a learning point for each of the four errors, as shown in Table 1. All of these learning points could be implemented by clinical teachers outside of the 'examination preparation' arena.

Providing information to clinical teachers about errors could therefore serve as a prompt for reflection.

Our reflections on these errors helped us to identify a learning point for each of the four errors.

This study is limited by being retrospective in nature. Discussion groups were not undertaken; thus, we were unable to obtain candidates' reflections on why incorrect answers were made or their actual cognitive reasoning processes. The examination was not critical to educational progression; thus, errors could be due to lack of engagement rather than a lack of knowledge or understanding. VSA questions were a new format for most participants, and previous studies have suggested that students may adapt learning and revision strategies if this type of assessment was used summatively.³ The errors we identified may therefore not be reproducible if widespread adoption of this assessment method occurs and may vary locally (we were unable to investigate error patterns made by students at different medical schools, for example). While there is no evidence that errors made in examinations are carried over to clinical practice, there is evidence that similar errors including over-reliance on a clinical finding or investigation, over-investigating and premature closure can all result in errors in clinical practice.¹¹ Finally, the frequency of each type of error in our dataset depends, in part, on the content of the questions used, and therefore, our discussion may not represent the most prevalent errors across the entire domain of applied medical knowledge.

5 | CONCLUSION

Analysis of student errors in answering VSA questions could provide information to guide clinical teachers. There is a dearth of literature looking at how examination results can be used to improve

teaching. We propose that analysis of errors could be beneficial to both the student and teacher. We have identified four types of error that resulted in repeated mistakes and suggest ways in which teaching could be adapted to mitigate them. This was a descriptive retrospective analysis, and thus, further work needs to be undertaken to determine optimal approaches to using the feedback identified by this, or a similar review to enhance teaching, and then to prospectively evaluate the impact of such changes to teaching.

ACKNOWLEDGEMENTS

We would like to thank the students who participated in the main VSAQ/SBAQ study and their medical schools for facilitating recruitment and logistics. Veronica Davids and Gareth Booth at the Medical Schools Council Assessment Alliance (MSCAA) were instrumental in enabling the main study to take place.

The MSCAA funded the study from which the data used in this study were taken.

MG is supported by the National Institute for Health Research Cambridge Biomedical Research Centre.

CB is supported by the National Institute for Health Research (NIHR) Applied Research Collaboration (ARC) West Midlands. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

CONFLICT OF INTEREST

At the time of the study, AHS, RW, MG and CB were members of the Board of the MSCAA. The authors declare they have no other conflicts of interest.

ETHICS STATEMENT

Ethical approval was provided by the Medical Education Ethics Committee (Reference MEEC-1718-100) at Imperial College London.

DATA AVAILABILITY STATEMENT

Data on student performance on each question in the main study are available via *BMJ Open*. 2019;9:e032550.

ORCID

Rachel Westacott  <https://orcid.org/0000-0001-9846-1961>

Amir H. Sam  <https://orcid.org/0000-0002-9599-9069>

Mark Gurnell  <https://orcid.org/0000-0001-5745-6832>

Celia A. Brown  <https://orcid.org/0000-0002-7526-0793>

REFERENCES

1. Surry LT, Torre D, Durning SJ. Exploring examinee behaviours as validity evidence for multiple-choice question examinations. *Med Educ*. 2017;51:1075-85.
2. Schuwirth L, Van der Vleuten C, Donkers H. A closer look at cueing effects in multiple-choice questions. *Med Educ*. 1996;30(1):44-9.
3. Sam AH, Field SM, Collares CF, et al. Very-short-answer questions: reliability, discrimination and acceptability. *Med Educ*. 2018;52:447-55.

4. Amin R, Patel R, Bamanian P. The importance of feedback for medical students' development. *Adv Med Educ Pract*. 2017;8:249–51.
5. Taylor C, Green K. OSCE feedback: a randomized trial of effectiveness, cost-effectiveness and student satisfaction. *Creat Educ*. 2013;4(6A):9–14.
6. Rees P. Do medical students learn from multiple choice examinations? *Med Educ*. 1986;20(2):123–5.
7. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull*. 1996;119(2):254–284.
8. Sethuraman K. The use of objective structured clinical examination (OSCE) for detecting and correcting teaching-learning errors in physical examination. *Med Teach*. 1993;15(4):365–8.
9. Sam AH, Westacott R, Gurnell M, Wilson R, Meeran K, Brown C. 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*. 2019;9:e032550.
10. General Medical Council. *Outcomes for Graduates*. London: General Medical Council; 2018.
11. Scott IA. Errors in clinical reasoning: causes and remedial strategies. *BMJ*. 2009;339:22–5.

How to cite this article: Putt O, Westacott R, Sam AH, Gurnell M, Brown CA. Using very short answer errors to guide teaching. *Clin Teach*. 2022;19(2):100–105. <https://doi.org/10.1111/tct.13458>