

Reliability and validity of short answer questions in Pharmacology**Syam Sreedharan, Dhanya Sasidharan Palappallil*, Elza Joy Munjely**

Department of Pharmacology,
Government MC, Kottayam,
Kerala, India

Received: 27 February 2019

Revised: 19 March 2019

Accepted: 28 March 2019

***Correspondence to:**

Dr. Dhanya Sasidharan

Palappallil,

Email: drspdharma@gmail.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: Short answer question (SAQ) is an excellent tool for assessment. Objective of this study was to understand the practices adopted in answering SAQ, assessment of reliability and validity of SAQ and faculty feedback on valuing SAQs.

Methods: A cross-sectional study was done in the Department of Pharmacology of a Government Medical College in Central Kerala from December 2018-January 2019. The study participants were given 10 minutes to answer the two chosen SAQs after which the response sheets were collected and valued by 2 independent examiners. Feedback on correction of the same questions was sought from Pharmacology faculties of various Medical Colleges through online survey. The data were analysed using SPSS software.

Results: 100% participants attempted to answer the SAQ 'Definition of a drug' but none of them came out with the World Health Organisation definition of drug. For the second SAQ, 'Two uses of atropine' around 25(20.8%) participants had written two uses of the drug correctly. The reliability was 0 for 'definition of a drug' and 0.50 (95% CI 0.082-0.702) for 'Two uses of atropine' and both the questions had no validity. The feedback on valuing definition and two use of atropine was variable.

Conclusions: None of the students wrote definition of a drug correctly and majority gave multiple responses to uses of atropine instead of two valid responses. The short answer questions we tested were found to have moderate to no reliability and no validity. The faculties had variable perceptions in valuing the two short answer questions.

Keywords: Atropine, Definition, Reliability, Short answer questions, Validity

INTRODUCTION

Assessment in Pharmacology has developed over years and the choice of the most appropriate instrument is highly contested. The essay questions, be it structured or modified assess the evolution of a concept in its wholesome and concentrates on the higher orders of cognitive domains rather than mere recall. Even though the Multiple Choice Questions has been well accepted in terms of reliability and resource effectiveness for cracking the entrance into the medical undergraduate and postgraduate courses, it has been perceived to be equivalent or inferior to other test instruments.^{1,2} The Short Answer Questions (SAQ) is a

focal tool that quantifies the cognitive skills and it capitalizes on spontaneous generation of answers bridging learning to real-life situations. It may be open-ended or closed and assess a wider dimension of the Knowledge domain from recall to comprehension, and some application.³ The versatility of SAQ lies in eliciting simple, short responses to demanding expression of a student's clinical reasoning and decision-making skills.¹ Though deemed to help facilitators in offering remediation, the inherent problems with SAQ lies in the resource intensiveness in construction and marking of written responses, subjectivity, and inter-observer variability in scoring candidates.⁴ This study was done

with an aim to understand the practices adopted by students in answering short answer question, to study the reliability and validity of SAQ under study and to obtain faculty feedback about valuing short answer question in Pharmacology.

METHODS

This was a cross-sectional study done in the Department of Pharmacology of a Government Medical College in Central Kerala for a period of two months (December 2018-January 2019). Since there were no ethical issues the clearance from Ethics Committee was not sought. ‘Definition of a drug’ and ‘Two uses of atropine’ were chosen as the short answer questions in this study. After completing sessions related to the selected SAQ, during a Pharmacology Lecture hour, the study participants were informed about the study by the Principal Investigator. All students who were willing to participate in the study were included after getting informed consent. After 10 minutes, the responses were collected. Anonymity was not maintained so as to provide constructive feedbacks to the students. The answers were reviewed and valued by 2 independent examiners and data were entered into an excel sheet by the investigators. Apart from age and gender the

parameters studied were completeness, correctness, reliability and validity of the answers. A Google fillout form was circulated through WhatsApp, as a group message as well as personal invites among the Pharmacology faculties of various Medical Colleges in Kerala to elicit their responses towards correction of the same questions. The data were sorted and entered into Statistical Package for the Social Science (SPSS) software version 16 (SPSS Inc, Chicago, USA). Descriptive data were expressed using frequencies and percentages. Reliability was tested by Intraclass correlation coefficient, two-way mixed effect model using an absolute agreement definition. Validity was checked by Cohen’s Kappa statistics.

RESULTS

Around 120 third semester undergraduate medical students participated in the study. There were 81 (67.5%) female and 39 (32.5%) males of mean age 20±2.08 years. The first short answer question was to define a drug. Even though 100% participants attempted to answer it, none of the participants came out with the World Health Organization definition of drug.

Table 1: Characteristics of answers of participants.

Number of participants	n=120
Response rate to both questions	100%
Correctness of ‘Definition of drug’ (2 marks)	None of the participant gave WHO definition of drug
Marks awarded for ‘Definition of drug’	Examiner 1- 0-120 (100%) Examiner 2-0-106 (88.3%), 0.5-9 (7.5%), 0.75-2 (1.7%), 1-2 (1.7%), 2-1 (0.8%) Intra Class Correlation Coefficient-0.00
Correctness of ‘Two uses of atropine’	25 (20.8%) participants had written correctly ‘only’ uses of the drug (22 wrote 2 correct responses; 2 wrote 3 correct responses; 1 wrote 5 correct responses)
Types of Responses to ‘Two uses of atropine’	25(20.8%)- two uses of the drug 18(15%) - incorrect uses 66(55%) -action of the drug± use 11(9.2%) - incorrect uses± action of the drug
Marks awarded for ‘Two uses of atropine’	Examiner 1---0-13 (10.8%), 1-56 (46.7%), 2-51 (42.5%) Examiner 2----0.5-2 (1.7%), 1-17(14.2), 1.5-19 (15.8%), 2-82 (68.3%) Intra Class Correlation Coefficient-0.50 (0.082-0.702)

Table 2: Number of responses versus number of correct uses written by participants.

Number of responses	Number of Correct Uses						Total
	0	1	2	3	4	5	
1	1	1	0	0	0	0	2
2	9	40	22	0	0	0	71
3	3	12	13	2	0	0	30
4	0	3	8	1	0	0	12
5	0	0	2	1	1	1	5
Total	13(10.8%)	56(46.7%)	45(37.5%)	4(3.3%)	1(0.8%)	1(0.8%)	120

The second short answer question was ‘Write two uses of atropine.’ All the participants attempted to answer the question but there were variable responses among the students as shown in Table 1. Around 25 (20.8%) participants had written two uses of the drug, 18 (15%) wrote incorrect uses, 66 (55%) wrote action of the drug at least once as their response, 11 (9.2%) wrote both incorrect uses as well as actions of the drug. Of the 120 participants, 45 (37.5%) listed 2 uses of atropine correctly, while four participants wrote three and one each wrote four and five uses of the drug correctly. Instead of giving 2 responses for two uses of the drug, the responses ranged from 1 to 5 as shown in Table 2. Of the 45 participants who listed two uses of atropine correctly 22 (48.9%), 13 (28.9%), 8 (17.8%), 2 (4.4%) had given 2, 3, 4 and 5 responses respectively as shown in Table 2.

The reliability was tested using the Intra class correlation coefficient which was 0 for ‘definition of a drug’ and 0.50 (95% CI 0.082-0.702) for ‘Two uses of atropine’. The validity was tested using Inter rater reliability with Cohen’s Kappa. However, examiner 1 had awarded 0 marks for all the participants for definition of drug and examiner 2 had not awarded 0 marks for any participants for ‘Two uses of atropine’ hence no measures of association were computed as at least one variable in each 2-way table upon which measures of association were computed was a constant.

The perceptions of Pharmacology faculties of various institutions in Kerala on valuing the 2 SAQs under study were collected through a Google fillout form. Around 70 faculties responded to the questionnaire of which 53 (75.7%) were females and 17 (24.3%) males. There were 10 Professor, 12 Associate Professors, 30 Assistant Professors and 18 Lecturers/ Junior Residents. For the first question regarding valuing the definition of a drug faculties were divided into equal halves as shown in Figure 1.

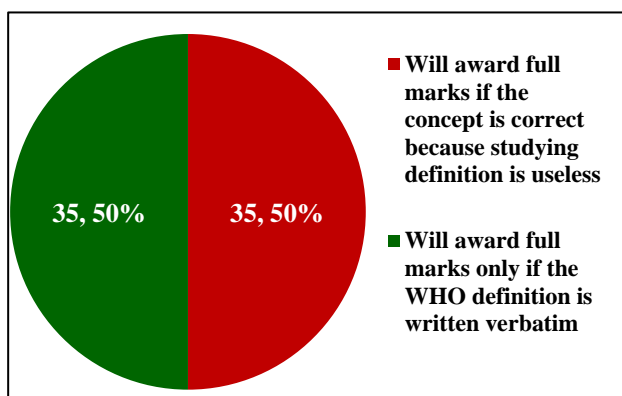


Figure 1: Opinion on valuing ‘Definition of a drug’.

On the question regarding valuing ‘Two uses of atropine’ majority of the faculty 35 (50%) responded that they would give full marks only if specific uses are written with rationale as shown in Figure 2. In a further statement for

the question ‘Write two uses of atropine’ if the student writes anything related to atropine, majority 52 (74.28%) stated that they will not give any marks while the rest said that they will give some marks.

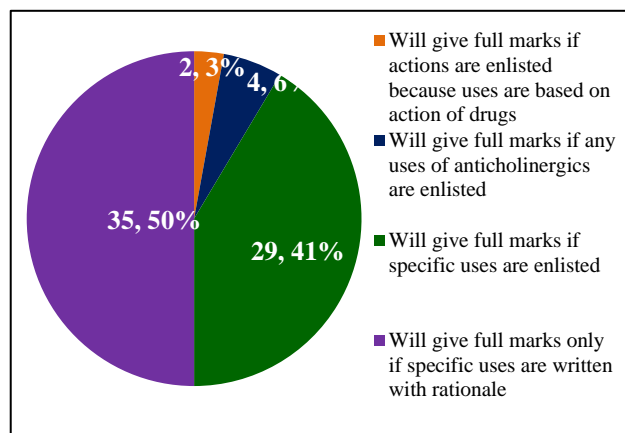
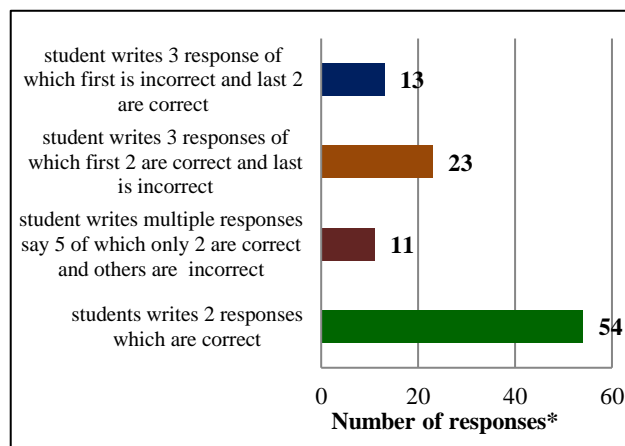


Figure 2: Responses to valuing a question on ‘Two uses of atropine’ with 2 marks.

On enquiring about the attitude of faculty towards the practice of writing ‘Multiple responses instead of two responses’, majority 39 (55.7%) stated that they would give full marks ‘only’ if the student writes 2 responses which are correct. However as shown in Figure 3, there were multiple response from some of the faculties who wanted to give marks for correct answers regardless of the number of incorrect responses written along with it.



*Some faculties gave multiple responses

Figure 3: Perception on correcting ‘Two uses of atropine’ based on the number of responses of students.

DISCUSSION

Reliability is one of the most important elements of test quality and validity is arguably the most important criteria for the quality of a test.⁵ Reliability refers to the extent to which assessments are consistent. We need reliable instruments with internal consistency so as to measure

student achievement. This should not make any difference, whether a student takes the assessment in the morning or evening; today or tomorrow.⁶ There are several methods for computing test reliability including test-retest reliability, parallel forms reliability, decision consistency, internal consistency, and inter-rater reliability.⁵ The values for reliability coefficients range from 0 (no reliability) to 1.0 (perfect reliability). If the reliability of a standardized test is above 0.80, it is said to have very good reliability; if it is below 0.50, it would not be considered a very reliable test.⁶ Validity refers to the accuracy of an assessment whether or not it measures what it is supposed to measure. Even if a test is reliable, it may not provide a valid measure. The ways to estimate the validity of a test include content validity, concurrent validity, and predictive validity.⁵

In this study authors have analyzed the reliability of two short answer questions one was definition of drug which had an Intraclass correlation coefficient of 0 which means that it was not a reliable short answer question. Authors could not assess the validity of the question because examiner 1 had awarded 0 marks for all the participants for definition of drug. The WHO (1966) has given a comprehensive definition. 'Drug is any substance or product that is used or is intended to be used to modify or explore physiological systems or pathological states for the benefit of the recipient'.⁷ This question was deemed to be non-reliable and it may be owed to the faculty feedback that concept is important than correctness of the WHO definition. Another point that may contribute to the variability in the response is that students may not have studied the definition at all.

The second question under study was 'Two uses of atropine' which had an Intraclass correlation coefficient of 0.50 [95% Confidence Interval 0.082-0.702]. This shows that this question had fair reliability. Pham et al, opined that drawbacks of an open-ended exam include less reliability and consumption of resources in construction, delivery, and marking which is reiterated in this study.¹ The low reliability of the question may be attributed to the fact that many students wrote action of the drug instead of writing the use of the drug.

On obtaining the perception of the faculties it was evident that many faculties would mark mechanism of action of the drug as correct when use of the drug is asked. There were also faculties who would give marks for 2 correct answers even if the total number of responses ranged from 3 to 5 with incorrect answers in between. The question 'Two uses of atropine' becomes a valid assessment tool only when it invites answers like used for fundoscopy, refractive error testing, organophosphate poisoning, Muscarine type (Early) mushroom poisoning, sinus bradycardia, as pre-anaesthetic agent and not when we elicit answers like it causes mydriasis, block muscarinic receptors, causes tachycardia, antisecretory action or any use of other anticholinergics.⁷ Schuwirth and van der Vleuten opined that, how a question is asked is more

important than how the question is answered.⁸ However, authors think that how a question is answered also matters. Some faculties opined that they would give full marks for the uses only if the use is properly justified with the rationale for use. Listing use is just recall whereas writing the rationale for the use invokes higher order of thinking like understanding and application. Pham et al, opined that the prerequisite of an SAQ to provide justification for each answer to obtain additional marks capitalizes on intermediate steps in thinking and hence if a single-format approach is used, the appropriateness and degree of "scaling" should be carefully considered.¹

Examiners act as a further source of error as they introduce internal adjustments which create inter-rater variability. As in this study while examiner-1 gave 0 for all definitions, examiner-2 did not give 0 for answer on uses of atropine. The low ICCs which suggest high inter-rater variability between examiners is a matter of apprehension similar to another study conducted.⁴ Expert examiners are prone for subjectivity by deviating from the proposed model/key answer.¹

Single Best Answer questions have become more and more widely accepted owing to testing of a wide range of topics with high reliability with the ease of machine marking with definitive correct answer regardless of the interpretation on the part of the examiner.⁹ However, the short comings include validity, quality of distractors and cueing. Very Short Answer Questions (VSAQ) developed by Sam et al, is machine marked in which students give a very short answer which typically consists of three words or less.

An ideal instrument should foster deep learning rather than cracking the question. SAQ and VSAQ are good tools for formative assessment and may be used as a closed as well as open book tool that require students to produce better memory or construct an answer, than tools that require students to recognize an answer.^{10,11}

The main limitation of the study was that only two short answer questions were under study. The test conducted in students was a surprise test which could have added to the unpreparedness and hence incorrect answers. Test-retest reliability was not assessed.

CONCLUSION

The two short answer questions we tested were found to have moderate to no reliability and no validity. The faculties had variable perceptions in valuing the two short answer questions. Authors should ensure the reliability and validity of these questions with the use of model answers with clear marking guidelines detailing both appropriate and inappropriate student answers. There should be strict guidelines for examiners on what to do, when the student writes multiple responses with correct and incorrect answers.

ACKNOWLEDGEMENTS

Authors are thankful to the students of 3rd semester of this institution who participated in this study. Authors have express their sincere gratitude for Pharmacology faculties of various institutions in Kerala who responded to the request by filling the Google fillout forms.

Funding: No funding sources

Conflict of interest: None declared

Ethical approval: Not required

REFERENCES

1. Pham H, Trigg M, Wu S, O'Connell A, Harry C, Barnard J, et al. Choosing medical assessments: Does the multiple-choice question make the grade?. *Educ Health.* 2018;31:65-71.
2. Hift RJ. Should essays and other "open-ended"-type questions retain a place in written summative assessment in clinical medicine? *BMC Med Educ.* 2014;14:249.
3. Sreedharan S, Hyma V, Palappallil DS. Concept learning through question framing in Pharmacology. *Int J Basic Clin Pharmacol.* 2018;7:319-23.
4. Palmer EJ, Duggan P, Devitt PG, Russell R. The modified essay question: Its exit from the exit examination? *Med Teach.* 2010;32:e300-7.
5. How do you determine if a Test has Validity, Reliability, Fairness, and Legal Defensibility? Last Accessed on 20 February 2019. Available at: http://www.proftesting.com/test_topics/test_quality.php
6. Classroom Assessment. Last accessed on 20 February 2019. Available at: <https://fcit.usf.edu/assessment/basic/basicc.html>
7. Tripathi KD. *Essentials of Medical Pharmacology.* 8th Edition. Jaypee Brothers Medical Publishers; 2019:2.
8. Schuwirth LW, van der Vleuten CP, Donkers HH. A closer look at cueing effects in multiple-choice questions. *Med Educ.* 1996;30:44-9.
9. Sam AH, Hameed S, Harris J, Meeran K. Validity of very short answer versus single best answer questions for undergraduate assessment. *BMC Med Educ.* 2016;16:266.
10. Wood T. Assessment not only drives learning, it may also help learning. *Med Educ.* 2009;43:5-6.
11. Kesavan KP, Palappallil DS. Effectiveness of Formative Assessment in Motivating and Improving the Outcome of Summative Assessment in Pharmacology for Medical Undergraduates. *J Clin Diagn Res.* 2018;12(5):FC08-11.

Cite this article as: Syam S, Palappallil DS, Munjely EJ. Reliability and validity of short answer questions in Pharmacology. *Int J Basic Clin Pharmacol* 2019;8:1019-23.