

Submitted: 30 April 2024
Accepted: 25 September 2024
Published online: 1 April, TAPS 2025, 10(2), 82-85
<https://doi.org/10.29060/TAPS.2025-10-2/SC3345>

Is Self-Mark dependable in Very Short Answer Question formats among pre-clinical medical students?

Sethapong Lertsakulbunlue & Anupong Kantiwong

Department of Pharmacology, Phramongkutklao College of Medicine, Thailand

Abstract

Introduction: Very Short Answer Questions (VSAQs) minimise cueing and simulate actual clinical practice more accurately than Single Best Answer Questions, as multiple-choice options might not be realistic. Phramongkutklao College of Medicine has developed a Self-Marked VSAQ (SM-VSAQ) for formative assessments. This study determines the validity and reliability of the SM-VSAQs.

Methods: Ninety-four third-year pre-clinical students took two occasions of 10-question SM-VSAQ exams regarding cardiovascular drugs. Each question consisted of two steps: (1) clinical vignettes with questions and (2) expected answers with scores, self-marking, and feedback comprehension. Scores ranged from 0.00 to 1.00 in 0.25 increments, though not every increment was applied to all questions. A distribution of the rating agreement between students' and teacher's ratings was presented to determine criterion-related validity and inter-rater reliability.

Results: Criterion-related validity revealed 90.64% and 93.19% of the ratings demonstrated exact agreement between students' and teachers' ratings, with an inter-rater reliability of 0.972 and 0.977 for the first and second occasions, respectively ($p=0.001$). The exact agreement was relatively lower on the first occasion for questions with more diverse expected answers (85.11%, $r=0.867$, $p=0.001$) and drugs requiring their specific full names for a perfect mark (74.47%, $r=0.849$, $p=0.001$). While questions with specific guides do not require complex answers, they received a higher exact agreement.

Conclusion: The SM-VSAQ format effectively combines guided answers with the VSAQ model. The agreement with teacher-rated is excellent. Marking discrepancies rooted in misconceptions underscores the importance of teacher feedback in improving self-grading in formative assessments. Regular self-assessment practice is recommended to enhance grading accuracy.

Keywords: *Very Short Answer Question, Self-assessment, Medical Education, Undergraduate, Pharmacology*

I. INTRODUCTION

Very Short Answer Questions (VSAQs) emerge as a relatively novel assessment format, addressing the constraints of traditional examination methods like Single Best Answer Questions (SBAQs), Constructed Response Questions (CRQs), and Modified Essay Questions (MEQs) (Sam et al., 2018). Although SBAQs are widely adopted in medical education globally, they are prone to cueing effects, leading examinees to depend on contextual clues, promoting a recognition-based learning approach (Sam et al., 2018). Moreover, the absence of multiple-choice options in real-life scenarios diminishes the relevance of SBAQs to medical practice.

Conversely, while CRQs and MEQs better mimic real-life situations, they suffer from rater dependency and significant evaluation time. Whereas VSAQs, free-response questions with 1–5 word answers, lessen rater dependency and evaluation time. Evidence indicates that VSAQs outperform SBAQs in discrimination, validity, and reliability in undergraduate assessments. Their open-ended nature prevents recognition-based learning and cueing. Additionally, VSAQs adeptly pinpoint common errors, often missed by SBAQs, and offer valuable feedback opportunities for educators (van Wijk et al., 2023).

Feedback is crucial for supporting and enhancing learning. Despite its longstanding importance in medical education, effective feedback is frequently deemed insufficient (Kuhlmann Lüdeke & Guillén Olaya, 2020). Self-assessment, enabled by formative exams, allows learners to identify their learning needs (Gedye, 2010). To improve feedback in formative assessments, Phramongkutklao College of Medicine (PCM) developed the Self-marked VSAQ (SM-VSAQ) format, which pairs a VSAQ with possible answers and a marking guide. Students may assess their understanding and pinpoint study areas through SM-VSAQ, enhancing feedback. Although VSAQs offer several benefits, challenges remain in grading the tests, as they may require a longer time. The self-graded format could address this issue in low-stakes examinations. This study assesses whether the SM-VSAQ with partial credit format, utilizing the marking guide, would achieve valid and reliable ratings compared with the teachers.

II. METHODS

Ninety-four third-year pre-clinical students participated in two 10-item SM-VSAQ during a cardiovascular pharmacology course. The exams covered antihypertensive, antiarrhythmic, antianginal, antithrombotic drugs, heart failure drugs, rational drug use, dyslipidaemia treatments, and drugs for atherosclerotic cardiovascular disease (ASCVD). The second SM-VSAQ sessions vary by changing the clinical vignette, the question, or both while maintaining the same underlying blueprint as the first session. Difficulty levels align with the Thai Medical Competency Assessment Criteria. Students had attended lectures on these drug groups before the exams. The VSAQ was content-validated by three professors for relevance, difficulty, feasibility, and simplicity using the Item Objective Congruence method with all over 0.67 of 1.00, indicating acceptable content validity. This approach ensured comparable difficulty.

The formative test was administered through Google Forms under examination conditions within a one-hour timeframe. Ethical approval was obtained from the Institutional Review Board, Royal Thai Army, and the waiver of the requirement for participant consent was deemed unnecessary following national regulations. An information sheet was provided on the first page of the Google Form. This initial test was conducted a day after they completed all lectures. After receiving teacher-led feedback and having time to review, students took a second parallel formative test ten days before the summative exam.

The SM-VSAQs featured four components for each question: clinical vignettes and questions on the first page, answers with scoring guidelines on the next page after they've answered, and a self-scoring option with feedback on answer comprehension. Scores ranged from 0.00 to 1.00 in 0.25 increments, though not every increment was applied to all questions. After the students completed the exam, they provided open-ended feedback on the pros and cons of the format. Examples of the format are shown in supplementary figures 1 and 2.

The self-rated, according to the marking guide, were exported into a Microsoft Excel spreadsheet to facilitate teacher ratings of the VSAQ answers. Using the 'filter' function in Microsoft Excel, the range of answers for each question was examined, and marks were awarded (Sam et al., 2018). Minor misspellings or alternative correct spellings were considered correct. Three pharmacology professors, who assigned scores, reviewed student answers that fell outside the guide. Consensus-determined scores require agreement from at least two of the three professors.

The data analyses were performed using *StataCorp, 2021, Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC*. Consistency reliability was analysed using Cronbach's alpha. Criterion-related validity was demonstrated by the distribution of the rating agreement between student and teacher ratings, presented as frequency and percentages. Inter-rater reliability was calculated using Pearson's correlation.

III. RESULTS

Cronbach's alpha for the SM-VSAQ was 0.741 and 0.721 on the first and second occasions, respectively. The teacher-rated alpha was 0.766 initially and 0.735 on the second. Criterion-related validity was assessed through agreement analysis (Supplementary Tables 1 and 2). Table 1 summarises the results of the agreement analysis. 90.6% and 93.19% of the ratings showed exact agreement between the students' and teachers' ratings, with an inter-rater reliability of 0.972 and 0.977 for the first and second occasions, respectively. The exact agreement is relatively low on the first occasion of Drugs used in heart failure (85.11%) and Anti-angina drugs (74.47%). Conversely, antithrombotics and drugs used in ASCVD received a high exact agreement of 96.81%. Example of questions with high and low agreement is demonstrated in supplementary figures 1 and 2. Additionally, content analysis of student's feedback revealed that they perceived that the format helps identify knowledge gaps, encourages review of missed topics, and aids in recognizing their current knowledge level (Supplementary Table 3).

Item	First Occasion						Second Occasion					
	Exact agreement	0.25 difference	0.50 difference	0.75 difference	1.00 difference	r*	Exact agreement	0.25 difference	0.50 difference	0.75 difference	1.00 difference	r*
	n (%)	n (%)	n (%)	n (%)	n (%)		n (%)	n (%)	n (%)	n (%)	n (%)	
Q1. Antihypertensive drugs	86 (91.49)	0 (0.00)	8 (8.51)	0 (0.00)	0 (0.00)	0.943	90 (95.74)	0 (0.00)	4 (4.26)	0 (0.00)	0 (0.00)	0.969
Q2. Antihypertensive drugs	87 (92.55)	4 (4.26)	3 (3.19)	0 (0.00)	0 (0.00)	0.964	91 (96.81)	0 (0.00)	3 (3.19)	0 (0.00)	0 (0.00)	0.965
Q3. Antihypertensive drugs	91 (96.81)	2 (2.13)	1 (1.06)	0 (0.00)	0 (0.00)	0.981	90 (95.74)	1 (1.06)	1 (1.06)	2 (2.13)	0 (0.00)	0.960
Q4. Antiarrhythmic drugs	90 (95.74)	2 (2.13)	1 (1.06)	0 (0.00)	1 (1.06)	0.961	91 (96.81)	2 (2.13)	0 (0.00)	1 (1.06)	0 (0.00)	0.980
Q5. Drugs used in heart failure	80 (85.11)	7 (7.45)	5 (5.32)	0 (0.00)	2 (2.13)	0.867	88 (93.62)	0 (0.00)	4 (4.26)	0 (0.00)	2 (2.13)	0.922
Q6. Anti-angina drugs	70 (74.47)	9 (9.57)	14 (14.89)	0 (0.00)	1 (1.06)	0.849	79 (84.04)	5 (5.32)	10 (10.64)	0 (0.00)	0 (0.00)	0.918
Q7. Antithrombotic drugs	91 (96.81)	2 (2.13)	1 (1.06)	0 (0.00)	0 (0.00)	0.983	83 (88.30)	6 (6.38)	2 (2.13)	2 (2.13)	1 (1.06)	0.880
Q8. Drugs used in dyslipidemia	84 (89.36)	3 (3.19)	6 (6.38)	0 (0.00)	1 (1.06)	0.915	89 (94.68)	1 (1.06)	2 (2.13)	1 (1.06)	1 (1.06)	0.936
Q9. CVS rational drug used	82 (87.23)	2 (2.13)	10 (10.64)	0 (0.00)	0 (0.00)	0.907	82 (87.23)	3 (3.19)	6 (6.38)	0 (0.00)	3 (3.19)	0.851
Q10. Drugs used in ASCVD	91 (96.81)	2 (2.13)	1 (1.06)	0 (0.00)	0 (0.00)	0.978	93 (98.94)	0 (0.00)	0 (0.00)	0 (0.00)	1 (1.06)	0.973
Total	852 (90.64)	33 (3.51)	50 (5.32)	0 (0.00)	5 (0.53)	0.972	876 (93.19)	18 (1.91)	32 (3.40)	6 (0.64)	8 (0.85)	0.977

* $p=0.001$ for all items, CVS: Cardiovascular system ASCVD: Atherosclerotic cardiovascular disease

Table 1. Comparison of rater agreement between the teacher and the self-rating on the VSAQ assessment

IV. DISCUSSION

VSAQs have demonstrated their discrimination, validity, and reliability among undergraduate assessments and their capacity to identify errors not detectable by SBAQs. However, the marking process poses challenges, potentially requiring more time than SBAQs, even with computerised marking systems (Bala et al., 2023). Delayed marking results in slower feedback delivery to students regarding their examination performance. Therefore, to our knowledge, the study is the first to demonstrate the reliability of using self-guided marking to provide students with immediate feedback after a formative VSAQ examination.

The inter-rater reliability exceeded 0.90 for nearly every question, suggesting the validity of self-grading compared with teacher grading. Moreover, by furnishing students with a partial credit guide, they were encouraged to analyse their answers to each guided answer, fostering a more profound understanding than the singular correct answer required in SBAQs, and encouraging engagement in higher-order thinking. The content analysis of student comments supports this. They found the partial credit guide helpful in identifying key knowledge areas, analyzing expected answers, and

engaging in self-directed learning. Additionally, path analysis showed that the first VSAQ attempt score positively influenced the second VSAQ understanding levels, primarily through the second attempt score, highlighting the benefits of multiple attempts for gaining insights (Supplementary Figure 3).

Discrepancies in ratings with the teacher likely stem from misconceptions. For example, while the correct response involved furosemide acting as a $\text{Na}^+/\text{K}^+/\text{2Cl}^-$ channel inhibitor, some students mistakenly identified it as a " $\text{Na}^+/\text{K}^+/\text{ATPase}$ " and awarded themselves full marks. Some students gave full marks for partially correct and imprecise responses. For instance, concerning the drug interaction between clarithromycin and warfarin, the answer involves enzyme inhibition by clarithromycin, yet some students merely stated, "Drug interaction between drugs." Similarly, in the anti-angina question, the correct answer is "sublingual nitroglycerin or sublingual isosorbide dinitrate." However, those who answered partially correctly still awarded themselves full marks. Additionally, disagreement may also be related to student ability, as those less familiar with the content, which leads to misconceptions, might not rate as well as those who are. To address discrepancies in the ratings, reviewing students' divergent responses could help refine

the marking guide. Furthermore, repeated practice in self-assessment will enhance students' ability to grade their answers accurately.

Conversely, questions with a high level of agreement provided detailed answers consisting solely of the drug name without asking for additional components such as the route of administration or mechanism of action. However, asking for multiple components helped enrich the knowledge and feedback that students could gain.

The present SM-VSAQ format has several strengths. First, it presents a realistic examination, as multiple-choices might not be available in real life. Second, it is simple, feasible, and adaptable, as perceived by the students. Third, it can be administered as an online formative examination, reducing the burden on teachers and providing immediate feedback to students, which has proven reliable and in high agreement with teachers. Nonetheless, this study has certain limitations. It only included a third-year pre-clinical student from a specific educational context, necessitating further research to assess the external validity of the findings.

V. CONCLUSION

SM-VSAQ approach facilitates engagement in higher-order thinking more effectively than the traditional single-best answer method. The format is also simple, adaptable to other subjects, and can be easily reviewed. The agreement between self-graded and teacher-provided ratings is outstanding. Discrepancies between student and teacher evaluations primarily stem from misconceptions in guided answers, highlighting the crucial need for teacher-led feedback to resolve these misunderstandings. This step is essential before implementing self-grading as an alternative in formative evaluations. Regular practice in self-assessment is advised to refine precision in self-grading. The SM-VSAQ format merges the VSAQ model with guided answers and may be further developed to improve feedback timeliness.

Notes on Contributors

SL reviewed the literature, designed the study, collected the data, conducted data analysis and wrote the manuscript. AK reviewed the literature, supervised, designed the study, performed the data analysis.

Ethical Approval

Ethical approval was obtained from the Medical Department Ethics Review Committee for Research in Human Subjects, Institutional Review Board, Royal Thai Army (IRBRTA) (Approval no. S079q/66_Xmp).

The IRBRTA waived the requirement for participant consent, deeming it unnecessary in accordance with national regulations.

Data Availability

Data sets analysed during the current study would be available from the corresponding author upon reasonable request. The Supplementary file for the current study is available from:

<https://doi.org/10.6084/m9.figshare.26507170>

Acknowledgement

This work would not have been possible without the active support of Phramongkutklao College of Medicine faculty members and its academic leaders, who are too numerous to name individually.

Funding

The authors reported no funding associated with the work featured in this article.

Declaration of Interest

The authors declare no competing interests.

References

-
- Bala, L., Westacott, R. J., Brown, C., & Sam, A. H. (2023). Twelve tips for introducing very short answer questions (VSAQs) into your medical curriculum. *Medical Teacher*, 45(4), 360–367. <https://doi.org/10.1080/0142159X.2022.2093706>
- Gedye, S. (2010). Formative assessment and feedback: A review. *Planet*, 23(1), 40–45. <https://doi.org/10.11120/plan.2010.00230040>
- Kuhlmann Lüdeke, A. B. E., & Guillén Olaya, J. F. (2020). Effective feedback, an essential component of all stages in medical education. *Universitas Médica*, 61(3). <https://doi.org/10.11144/Javeriana.umed61-3.feed>
- Sam, A. H., Field, S. M., Collares, C. F., van der Vleuten, C. P. M., Wass, V. J., Melville, C., Harris, J., & Meeran, K. (2018). Very-short-answer questions: Reliability, discrimination and acceptability. *Medical Education*, 52(4), 447–455. <https://doi.org/10.1111/medu.13504>
- van Wijk, E. V., Janse, R. J., Ruijter, B. N., Rohling, J. H. T., van der Kraan, J., Crobach, S., de Jonge, M., de Beaufort, A. J., Dekker, F. W., & Langers, A. M. J. (2023). Use of very short answer questions compared to multiple choice questions in undergraduate medical students: An external validation study. *PLOS ONE*, 18(7), e0288558. <https://doi.org/10.1371/journal.pone.0288558>

*Anupong Kantiwong
Department of Pharmacology,
Phramongkutklao College of Medicine, Bangkok, 10400
Email: anupongpcm31@gmail.com