**ORIGINAL RESEARCH**

# Very Short Answer Questions in Team-Based Learning: Limited Effect on Peer Elaboration and Memory

Kai Yuan Lau[1] · Jessica Yang Huey Ang[1] · Preman Rajalingam[1]

## Abstract

**Introduction** In team-based learning (TBL), single best answer questions (SBAQs) are traditionally used because immediate computer-assisted feedback facilitates team discussions. Recent improvements in digital marking systems and criticisms of non-analytical strategies in SBAQs have prompted the consideration of very short answer questions (VSAQs) as an alternative to SBAQs. We aim to compare the effect of VSAQs and SBAQs on peer elaboration and knowledge retention in TBL.
**Materials and Methods** Twenty-four second-year students from the Lee Kong Chian School of Medicine participated in a mixed-methods study that included a randomised controlled crossover trial with two intervention arms (TBL-VSAQs and TBL-SBAQs). Two TBL sessions were conducted, with one topic covered each. Students were randomly allocated into six teams of four members. Individual and team scores and completion times were measured, and students were surveyed on their TBL experience. A follow-up quiz on the same topics was administered two weeks later to assess knowledge retention.
**Results** Individuals scored lower for VSAQs than SBAQs in the second TBL topic ($7.17 \pm 1.52$ versus $8.25 \pm 1.48$; $p = 0.046$), while findings in other metrics were non-significant. Follow-up quiz scores showed no significant difference in knowledge retention, although effect size and power were low. Students perceived VSAQs as more authentic and challenging, though most preferred the continued use of SBAQs for TBL.
**Discussion** VSAQs have a limited impact on peer elaboration and knowledge retention versus SBAQs in TBL.
**Conclusions** We should be circumspect about implementing VSAQs in TBL especially for medical students at the pre-clinical level.

**Keywords** Assessment · Team-based learning (TBL) · Single best answer · Very short answer · Knowledge retention

## Introduction

Team-based learning (TBL) is a collaborative learning pedagogy that thrives on active student participation. Students undertake pre-class preparation before attempting an individual readiness assurance test (iRAT), a team readiness assurance test (tRAT) and a team application exercise (AE) to apply concepts in a clinical context. Team discussions are a crucial component of TBL where students engage in peer elaboration, fostering a process of knowledge recall, reconstruction and reinforcement (see Fig. 1).

Traditionally, single best answer questions (SBAQs) are used in the iRAT and tRAT because they can assess a broad range of learner knowledge in a short period. Also, the efficient computer-assisted immediate feedback on SBAQs can facilitate peer elaboration during the team discussion [1]. Peer elaboration is a strong activator of prior knowledge and has been proposed to be an integral part of deep learning and knowledge retention [2]. However, studies have shown that SBAQs lead to cueing and the adoption of non-analytical reasoning strategies [3, 4].

Very short answer questions (VSAQs) have emerged as an interesting alternative to SBAQs. While both provide an identical vignette and lead-in question, VSAQs require a free-text answer of one to five words instead of selecting from five choices in SBAQs. In VSAQs, the free-text answer is matched to a predetermined set of correct answers, but the ability of content experts to mark reasonable free-text answers that differ slightly from the predetermined answer key was limited by technology. However, developing a

✉ Kai Yuan Lau
  lauk0021@e.ntu.edu.sg

1  Medical Education Research and Scholarship Unit
   (MERSU), Lee Kong Chian School of Medicine, Nanyang
   Technological University, Singapore, Singapore

**Fig. 1** Cognitive processes in TBL, with a focus on team discussions during tRAT
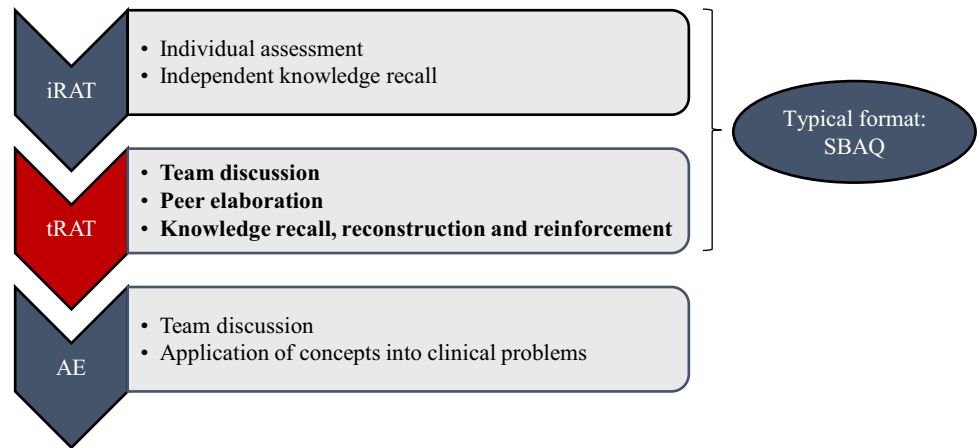


digital marking system has circumvented the challenge of marking free-text answers in real-time, supporting the integration of VSAQs into TBL [5].

VSAQs are a more authentic individual assessment tool than SBAQs because they replicate the open-ended questions that students encounter in clinical postings and real-life clinical practice. VSAQs also confer greater validity in the assessment of independent knowledge recall by eliminating cueing and guessing [3, 6, 7]. Studies on integrating VSAQs into TBL have yielded favourable perspectives from students in terms of enriching group discussions and improving learning approaches [5, 8].

Hence, VSAQs have the potential to enhance not only the individual learning experience through the iRAT, but also the process of peer elaboration and knowledge retention during the tRAT. We define the period of knowledge retention as two weeks. Our study strives to address three main questions. Firstly, are VSAQs more difficult and do they lead to longer team discussions than SBAQs? Secondly, do VSAQs promote better knowledge retention than SBAQs? Thirdly, did students perceive VSAQs to be more effective at facilitating peer elaboration and deep learning than SBAQs? We hypothesise that VSAQs are more difficult than SBAQs, lead to longer discussion times, promote deeper peer elaboration and enhance knowledge retention.

## Materials and Methods

A total of 24 pre-clinical second-year students were recruited from the Lee Kong Chian School of Medicine, where TBL is the main pedagogical method for the first two years of the medical undergraduate curriculum [9]. Ethical approval was obtained from the Nanyang Technological University Institutional Review Board (reference number: IRB-2020–01-029) and informed consent was collected before the commencement of the study. Participants were sourced through mass messaging channels such as email and renumerated with food vouchers for their time spent.

### Pre-TBL

Preparatory materials on two topics were distributed five days before the TBL, covering the mouth and oesophagus and red blood cells, respectively. A TBL session was curated for each topic using an open-source Learning Activity Management Software (LAMS) (www.lamsfoundation.org). Each TBL comprised 10 questions in either an SBAQ or VSAQ format, conducted in the iRAT/tRAT sequence, followed by a clinical question for AE.

### TBL

Students participated in the TBL sessions on 11 August 2021 remotely via an online Zoom meeting platform due to COVID-19 precautions, differing from a traditional in-person classroom session. They had experienced this online TBL format for at least one academic year amidst the pandemic.

The crossover design is outlined in Fig. 2. Students were randomly allocated into six teams of four members each. Teams were randomly allocated to group A ($n = 12$) or group B ($n = 12$). For the first TBL session, group A did SBAQs, whereas group B did VSAQs. During the second TBL session, groups were exposed to a different question format from the first session, respectively, with group A given VSAQs and group B given SBAQs instead.

After completing the iRAT individually, students were placed into the same Zoom breakout rooms with their teams for leader selection and the tRAT. During the tRAT, teams doing SBAQs selected answers from a list of five choices, while teams doing VSAQs submitted a free-text answer of one to five words in length. Both groups received immediate feedback after each attempt and had unlimited attempts until they were correct or chose to move on to the next question. A time limit to complete each section was not strictly enforced, but teams were advised to submit all their answers upon completion.

**Fig. 2** Crossover design with VSAQs and SBAQs on Topic 1 (mouth and oesophagus) and Topic 2 (red blood cells)
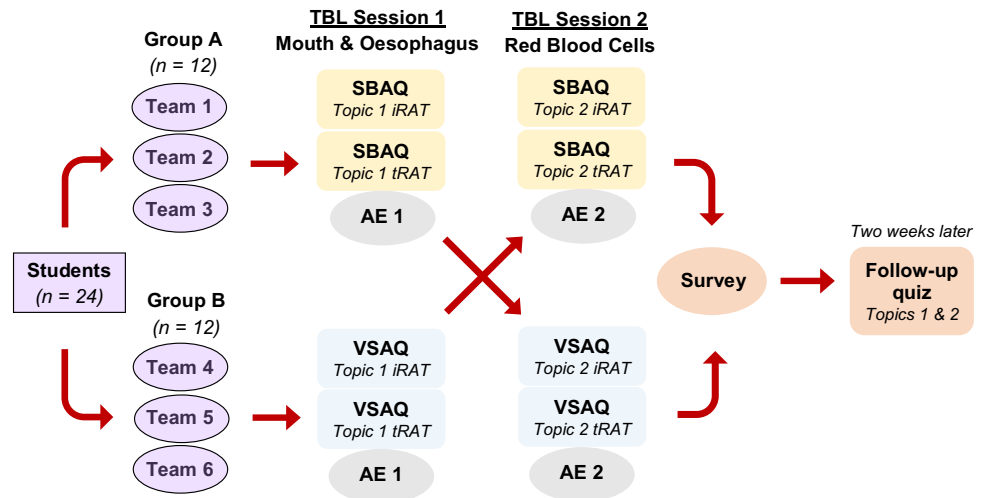


## Post-TBL

Scores and completion times for the iRAT and tRAT of each TBL were analysed using Microsoft Excel with one-tailed unpaired *t* tests performed between groups. A post-TBL survey was conducted to elicit students' perspectives towards VSAQs in TBL, with a focus on their TBL preparation and intra-TBL team discussions. There were eight closed-ended questions using a 5-point Likert scale (strongly agree "5" to strongly disagree "1") and an open-ended question on preferences about the use of VSAQs and SBAQs in TBL.

Students were administered a follow-up quiz two weeks later, comprising 20 questions from both topics in VSAQ format. Calculations were first performed to determine if there were any statistically significant carryover effects between different sets of questions, followed by calculations to determine if there was any significant difference in quiz scores between the two groups to analyse the effect of TBL question format on knowledge retention [10]. Carryover and crossover statistical methods are detailed in Appendix B.

## Results

### Are VSAQs More Difficult and Do They Lead to Longer Discussion Times than SBAQs?

Summarised TBL metrics are displayed in Table 1. Based on iRAT scores, individuals performed significantly better in SBAQs than VSAQs for Topic 2 only ($7.17 \pm 1.52$ versus $8.25 \pm 1.48$; $p = 0.046$, pooled SD = 1.51, Cohen's $d = -0.72$). This suggests that VSAQs may have been more difficult than SBAQs for Topic 2 but not Topic 1. Individuals did not take significantly longer to complete VSAQs than

SBAQs within either of the topics ($p = 0.13$ for Topic 1; $p = 0.15$ for Topic 2).

Based on tRAT scores, there was no significant difference in team performance between SBAQ and VSAQ groups. Hence, VSAQs were not more difficult than SBAQs in a team setting. There was also no significant difference in tRAT completion time between SBAQ and VSAQ groups ($p = 0.31$ for Topic 1, $p = 0.07$ for Topic 2). Unexpectedly, there was a non-significant trend ($p = 0.07$) for teams to complete VSAQs more quickly than the teams doing SBAQs for Topic 2 ($648 \pm 202$ versus $916 \pm 62$ s). This refutes the hypothesis that VSAQs lead to longer team discussions than SBAQs.

**Table 1** Individual and team performance for TBL (one-tailed unpaired *t* tests for scores and times)

| Metrics | SBAQ | | VSAQ | | *p* value |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | |
| *Topic 1* | | | | | |
| | Group A | | Group B | | |
| *iRAT score (out of 10)* | 8.92 | 1.31 | 8.00 | 1.86 | 0.089 |
| *iRAT completion time (s)* | 406 | 342 | 582 | 386 | 0.13 |
| *tRAT score (out of 10)* | 10 | 0 | 10 | 0 | - |
| *tRAT completion time (s)* | 307 | 199 | 470 | 455 | 0.31 |
| *Topic 2* | | | | | |
| | Group B | | Group A | | |
| *iRAT score (out of 10)* | 8.25 | 1.48 | 7.17 | 1.53 | **0.046\*** |
| *iRAT completion time (s)* | 466 | 328 | 600 | 296 | 0.15 |
| *tRAT score (out of 10)* | 9.42 | 1.01 | 8.83 | 0.29 | 0.21 |
| *tRAT completion time (s)* | 916 | 62 | 648 | 202 | 0.07 |

*Bold denotes significance at the 5% level, with pooled SD = 1.51 and Cohen's $d = -0.72$

**Table 2** Individual performance in a two-week follow-up quiz by topic (crossover *t* test for topical scores)

| Quiz score (out of 10) | Group A | | Group B | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| *Topic 1 (mouth and oesophagus)* | 9.25 | 1.14 | 8.92 | 0.10 |
| *Topic 2 (red blood cells)* | 6.17 | 2.21 | 5.83 | 2.59 |
| Crossover statistics | t(22) | | *p* value | |
| *Carryover effects* | 0.627 | | 0.537 | |
| *Difference in treatment effects* | 0 | | 1 | |

## Do VSAQs Promote Better Knowledge Retention than SBAQs?

The mean topical scores for the follow-up quiz are shown in Table 2. As demonstrated in Appendix B, there were negligible within-subject carryover effects when subjects attempted SBAQs on one topic and VSAQs on another topic ($p = 0.539$). There was no significant difference in scores regardless of experience with SBAQs or VSAQs for a particular topic in TBL ($p = 1$). Thus, VSAQs did not confer a better two-week retention rate compared to SBAQs in our study.

## Did Students Perceive VSAQs To Be More Effective at Facilitating Peer Elaboration and Deep Learning than SBAQs?

Survey responses to the Likert scale questionnaires are summarised in Fig. 3, with a detailed breakdown of responses in Appendix C. Student perception towards VSAQs was mixed. Most students agreed that VSAQs were more challenging (95.8%) and authentic in assessing medical knowledge (79.2%) than SBAQs. Exactly half perceived longer discussions for VSAQs than SBAQs. Nearly half agreed that discussions over VSAQs were more exhausting (43.8%), and a similar proportion agreed that more members were involved in the discussion of VSAQs than SBAQs (37.5%). Fewer students felt that VSAQs produced a greater diversity of opinions (20.8%), and an equally small proportion believed that VSAQs gave them more opportunities to share their own answers (20.8%). Overall, a minority found VSAQs more enjoyable than SBAQs for TBL (20.8%).

A large proportion of students (45.8%) preferred SBAQs over VSAQs in TBL, while a fewer number were either favourable towards VSAQs (29.2%) or had no preference (25%). An analysis of the open feedback yielded three main reasons why SBAQs continue to be preferred by the student participants for TBL.

Firstly, some students felt that SBAQs generated more points for discussion than VSAQs (12.5%). They believed that the answer list provided in SBAQs would prompt team members to discuss the concepts behind each option and empower less confident members to take a stand, therefore promoting participation in the discussion. On the other hand, one student opined that it was "easier to stay silent and not contribute" for VSAQs by leaving a blank individual answer, resulting in a muted discussion should few team members offer an answer.

Secondly, one student felt that SBAQs might be more suitable than VSAQs in the foundational learning of "new, unfamiliar topics". It was highlighted that an answer list might facilitate the process of knowledge recall, allowing students to become "familiar and comfortable with certain terms and concepts". Subsequently, VSAQs may play a greater role as a revision tool for students who have already mastered a topic closer to examinations.

Thirdly, one student saw value in certain types of SBAQs which could be as challenging as VSAQs, such as those requiring a choice of the most appropriate statement or the incorrect statement from an answer list. These questions would have otherwise not existed if VSAQs replaced SBAQs in TBL.

A common issue with VSAQs was the need for "exact answers" because free-text inputs with wrong spelling or slight deviations from "specific phrasing" were automatically rejected (25%). One student likened the guessing of phrasing in VSAQs to the guesswork of choosing from an answer list in SBAQs.

Among those favourable towards VSAQs, most still preferred a mixture of SBAQs and VSAQs to a pure VSAQ format for TBL (20.8%). The word "guess" came up most frequently in the survey (20.8%), capturing the essence of how VSAQs prevented students from "guessing and getting the correct answer". A few highlighted how the challenging nature of VSAQs in prompting active recall stretched them to remember important details and understand concepts fully during TBL preparation (16.7%).

## Discussion

### Cueing and Team Discussions

VSAQs did not lead to significantly longer team discussion times than SBAQs. It is postulated that the lack of cueing from an answer list for hard-to-recall questions may lead to more members shying away from the team discussion and relying on others for a consensus answer. While VSAQs are more authentic and promote analytical reasoning in more experienced learners under certain contexts such as summative assessment, students that participate in TBL tend to be in the process of learning new concepts. The cueing provided by SBAQs triggers better retrieval among novice learners, subsequently enabling them to engage in a conversation. In effect, the answer choices in SBAQs act as a form
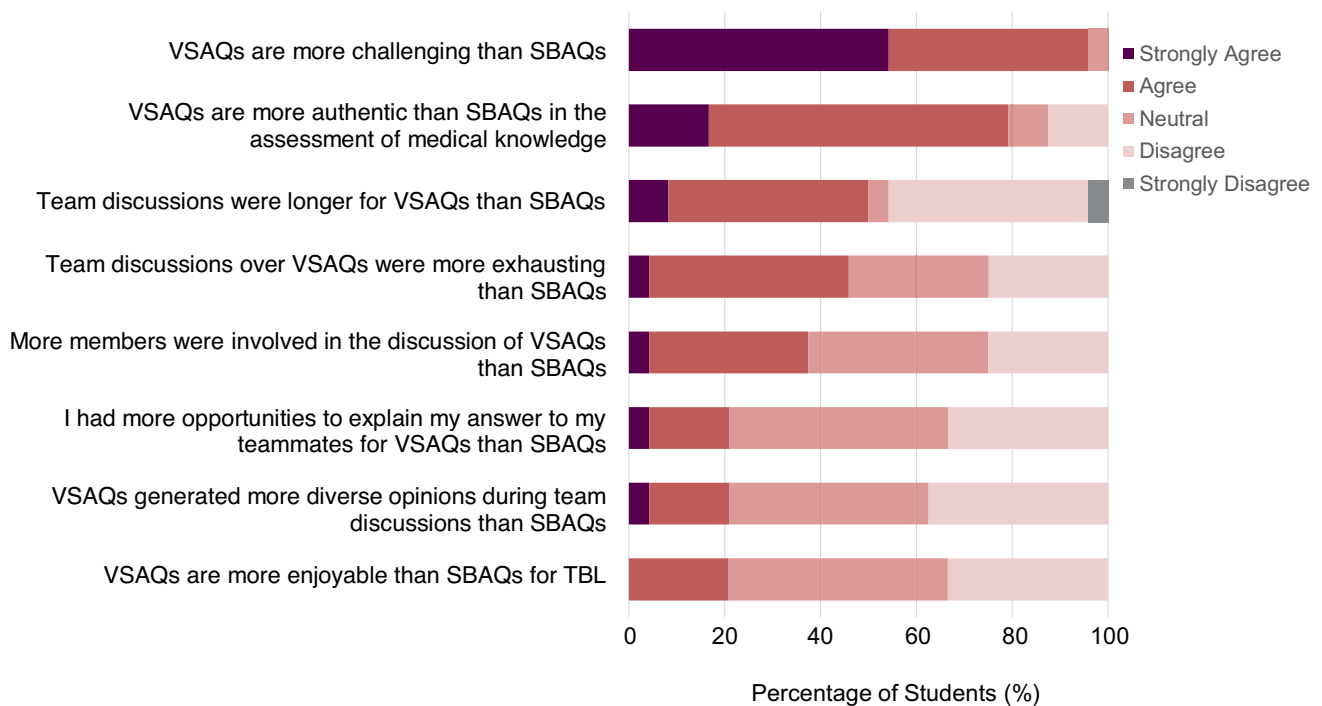
**Fig. 3** Student responses to Likert scale questionnaire in post-TBL survey

of scaffolding for these novice learners [11]. This correlates with findings from the post-TBL survey, where the majority felt that VSAQs did not prolong team discussion times, elicit more diverse views or enable more team members to share their answers, compared to SBAQs.

## Knowledge Retention

VSAQs did not improve knowledge retention compared to SBAQs in our study. This could have been attributed to the low conceptual difficulty of the TBL questions used. The effectiveness of an educational task in producing deep cognitive processes depends on its conceptual level on Bloom's taxonomy [12]. In a study on team discussions in a biology course TBL, questions of a higher Bloom's level, which assessed application and analytical skills, created more instances of peer elaboration such as conceptual explanation, re-evaluation and co-construction compared to lower-order questions that test recall and basic understanding [13]. In our study, questions may have been of a low order to cater to pre-clinical students, focusing on knowledge recall and comprehension skills, respectively. The low conceptual level of the questions could have dampened peer elaboration and shrouded the potential benefit of VSAQs on peer elaboration. This may also explain why students' perceptions of the benefits of VSAQs were mixed. The small sample size also limited our study's effect size and power.

Students generally perceived Topic 2 questions to be more challenging than Topic 1 questions from post-quiz feedback. This may be attributed to the difference in content structure between the two topics. For example, Topic 1 questions on the mouth and oesophagus were largely anatomy-based with a narrow spectrum of content. In contrast, Topic 2 questions on red blood cells focused mainly on mechanisms and diseases from a broader spectrum of content. Thus, students may have found Topic 1 questions easier to study for and attempt.

## Differences in Perceptions from Other Studies

Notably, the student perspectives uncovered in our study differ from the favourable responses seen in a similar study conducted by Millar et al., where most students felt that VSAQs in TBL would improve their preparation for clinical practice [5]. One possible reason why students in our study preferred SBAQs for TBL is that they were just starting their second year of study and had not experienced much exposure to clinical reasoning. Additionally, they may have felt more comfortable with SBAQs as this was a question format used to entrench their learning during their first year of school. On the other hand, students in the study by Millar et al. were third-year students who would have had a longer learning experience and exposure to clinical medicine, such that they perhaps possessed a better appreciation of how

their clinical acumen could be further sharpened with the help of VSAQs. Millar et al. suggest that SBAQs could give students a false impression of their own clinical competence by giving them options that cue their thinking. However, perhaps pre-clinical students who are just beginning to learn the medical sciences find it easier to start discussing if they are cued by the options. This hypothesis is supported by Bird et al. where some students view the lack of a cueing effect as a drawback [14]. More research into student beliefs about the cueing effect and its impact on their learning outcomes is needed.

## Limitations

There were a few limitations in this study. Firstly, it was challenging to implement higher-order VSAQs due to the foundational nature of these TBL sessions. Most of the questions centred around knowledge recall and comprehension skills on pre-clinical biology rather than the interpretation and analysis of clinical-level topics. Lower-order VSAQs were thus utilised to cater to the knowledge level of the pre-clinical medical students. Secondly, there were difficulties in recruiting student participants, resulting in an underpowered study. Thirdly, some students indicated that they were beginning to learn about Topic 1 in their curriculum around the time of the follow-up quiz. This potentially confounding factor should be avoided in future studies.

## Conclusions

Our study showed that VSAQs had no impact on peer elaboration and knowledge retention, with no compelling indication to replace SBAQs with VSAQs at least for TBL in an undergraduate medical course. Further studies with a larger sample size are required to establish the appropriate conceptual level from which VSAQs can improve peer elaboration in TBL. The recording and coding-assisted analysis of conversations could more accurately assess how VSAQs affect team discussions. Given the apparent limitations of VSAQs in foundational learning scenarios such as at the pre-clinical level, expanding the study population to the clinical-level undergraduate or graduate students might yield greater insight into the potential benefits of higher-order VSAQs in TBL.

## Declarations

**Disclaimer** Kai Yuan Lau is an undergraduate at the Lee Kong Chian School of Medicine (LKCMedicine), Nanyang Technological University (NTU). Jessica Yang Huey Ang, MSc, is lead for Educational Development at LKCMedicine, NTU. Preman Rajalingam, PhD, is senior lecturer at LKCMedicine and Director Centre for Teaching, Learning and Pedagogy at NTU.

## References

1. Parmelee D, Michaelsen LK, Cook S, Hudes PD. Team-based learning: a practical guide: AMEE guide no. 65. Med Teach. 2012;34:e275-287. https://doi.org/10.3109/0142159X.2012.651179.
2. Schmidt HG, Rotgans JI, Rajalingam P, Low-Beer N. A psychological foundation for team-based learning: knowledge reconsolidation. Acad Med. 2019;94:1878–83. https://doi.org/10.1097/ACM.0000000000002810.
3. 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. https://doi.org/10.1136/bmjopen-2019-032550.
4. 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. https://doi.org/10.1186/s12909-016-0793-z.
5. Millar KR, Reid MD, Rajalingam P, Canning CA, Halse O, Low-Beer N, et al. Exploring the feasibility of using very short answer questions (VSAQs) in team-based learning (TBL). Clin Teach. 2021. https://doi.org/10.1111/tct.13347.
6. Sam AH, Field SM, Collares CF, van der Vleuten CPM, Wass VJ, Melville C, et al. Very-short-answer questions: reliability, discrimination and acceptability. Med Educ. 2018;52:447–55. https://doi.org/10.1111/medu.13504.
7. Sam AH, Wilson R, Westacott R, Gurnell M, Melville C, Brown CA. Thinking differently - students' cognitive processes when answering two different formats of written question. Med Teach 2021: 1–8. https://doi.org/10.1080/0142159X.2021.1935831.

8. Tan SH, Thibault G, Chew AC, Rajalingam P. Enabling open-ended questions in team-based learning using automated marking: impact on student achievement, learning and engagement. J Comput Assist Learn. 2022;38:1347–59. https://doi.org/10.1111/jcal.12680.

9. Rajalingam P, Rotgans JI, Zary N, Ferenczi MA, Gagnon P, Low-Beer N. Implementation of team-based learning on a large scale: three factors to keep in mind. Med Teach. 2018;40:582–8. https://doi.org/10.1080/0142159X.2018.1451630.

10. Wellek S, Blettner M. On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2012;109:276–81. https://doi.org/10.3238/arztebl.2012.0276.

11. Little JL, Bjork EL, Bjork RA, Angello G. Multiple-choice tests exonerated, at least of some charges: fostering test-induced learning and avoiding test-induced forgetting. Psychol Sci. 2012;23:1337–44. https://doi.org/10.1177/0956797612443370.

12. Adams NE. Bloom's taxonomy of cognitive learning objectives. J Med Libr Assoc. 2015;103:152–3. https://doi.org/10.3163/1536-5050.103.3.010.

13. Leupen SM, Kephart KL, Hodges LC. Factors influencing quality of team discussion: discourse analysis in an undergraduate team-based learning biology course. CBE Life Sci Educ. 2020;19:ar7. https://doi.org/10.1187/cbe.19-06-0112.

14. Bird JB, Olvet DM, Willey JM, Brenner J. Patients don't come with multiple choice options: essay-based assessment in UME. Med Educ Online. 2019;24:1649959. https://doi.org/10.1080/10872981.2019.1649959.