

Assessment as Learning: A Mixed-Methods Study on Pre-Clinical Medical Student-Led Generation of Scenario-Based Multiple Choice Questions

Asmaa F Sharif^{1,2}, Zeinab A Kasemy³, Samar HA Alohari⁴

¹Clinical Medical Sciences Department, College of Medicine, Dar Al Uloom University, Riyadh, Saudi Arabia; ²Forensic Medicine and Clinical Toxicology Department, Faculty of Medicine, Tanta University, Tanta, Egypt; ³Department of Public Health and Community Medicine, Faculty of Medicine, Menoufia University, Menoufia, Egypt; ⁴Department of Public Health and Community Medicine, Faculty of Medicine, Taiz University, Taiz, Yemen

Correspondence: Samar HA Alohari, Department of Public Health and Community Medicine, Faculty of Medicine, Taiz University, Taiz, P.O. Box 6803, Yemen, Tel +9674232429, Email dr.aloharisamar@gmail.com

Purpose: Learning basic sciences divorced from clinical practices is suboptimal. Several student-centered learning strategies were introduced to enhance self-regulated learning while adopting learning techniques centered on assessment receives much less attention. This study aimed to explore the perception and efficacy of engaging early-phase preclinical medical students in generating scenario-based Multiple-choice questions (MCQs) targeting high cognitive levels.

Methods: This study is a quasi-experimental mixed-method study conducted among preclinical second-year medical students. An intervention group involved students who participated in item creation, and a control group included those who did not. The two groups were exposed to mock exams and were asked to complete a self-administered questionnaire and participate in focused group discussions on their perceptions.

Results: Out of 135 students, 48.9% created from 1 to 10 items. Item creators had significantly higher formative scores (34.3 ± 10.5) than non-creators (29.6 ± 10.6) ($p = 0.010$), and their GPA was also significantly better (B+ grade: 42.4% vs 21.7%; $p = 0.015$). Despite these academic variations, above 75% of the two groups were satisfied or very satisfied with non-significant variations. Formative assessment positively correlated with GPAs ($r_s = 0.74$) and the number of developed items ($r_s = 0.39$). The adopted strategy imposed a positive effect on the learning process and fostered procedural soft skills and student values. It allowed students to experience positive emotions and feelings of entertainment during learning and changed their perception of scenario-based MCQs. Students prefer generating such items in other subjects.

Conclusion: The student-led approach to generating clinical scenario-based MCQs demonstrates significant promise as an innovative pedagogical approach in medical education. Involving preclinical students in developing high cognitive levels of MCQs is not merely a tool for evaluation but a versatile pedagogical strategy that can be adapted to various disciplines to enhance learning and engagement and help students transition to the clinical years.

Keywords: assessment, learning, multiple choice questions, basic medical sciences, formative assessments

Introduction

Today, medical education faces significant challenges with shifting to reformed integrated curricula.¹ With increasing calls to integration, almost all medical schools shifted toward integrated curricula. Despite the proven advantages of integration, its application is deviated from what it is assumed to be.² Medical education now adheres to a unified path globally, where basic medical sciences are taught earlier followed by clinical sciences.² The aim of teaching basic sciences is to understand the relationship between the disease and its biological basis on clinical diagnosis.³ Oktar advised not to start medical education curricula with basic sciences followed by a series of clinical subjects. Instead, they proposed involving clinicians in teaching basic sciences side by side with clinical subjects.²



Learning basic medical sciences divorced from clinical practices makes them of low value and impairs student's ability to use these sciences in realistic clinical contexts.² On the other hand, clinical teaching is not an easy task, which implies learning focused on patients and their problems.⁴ This task demands that the tutor should be more than a medical expert. Clinical teacher plays multiple roles, including being an information provider, role model, facilitator, assessor, curriculum planner, and resource material creator.⁵

Indeed, the challenges facing medical education are not related to the integration itself but rather to the separation between basic and clinical medical sciences, where basic sciences are taught by the scientists, particularly the molecular disciplines, and clinicians were forced only to practice bedside teaching for senior students.⁶ Once the students begin the clinical phases, they forget the knowledge they received in basic sciences.⁷ Now, it is the time to take bold steps and apply integration in an optimum way, starting with removing the term preclinical from medical education and replacing it with clinical education through the undergraduate learning journey. Additionally, teaching the basic medical sciences should be under the responsibility of clinicians, ensuring more exposure to clinical practices through all phases of the program.²

Traditional theories of adult learning include instrumental learning theories, transformative learning theories, humanistic theories, social theories of learning, and motivational and reflective models. These theories emphasize that adult students learn through engagement in successive stages of dissonance, refinement, organization, and feedback.⁸ Many student-centered learning strategies were invented aiming to bridge the gap, enhance self-regulated learning (SRL) and enrich medical curricula with clinical backgrounds, including problem-based learning (PBL), team-based learning (TBL), and the use of simulation. Adding clinical hue to the preclinical years of the curriculum was the main scope of these innovative strategies.⁷ It is seen that all efforts are paid to the learning strategies. In contrast to teaching, students are much less involved in assessments.⁹ Developing active learning in undergraduate medical education by adopting learning techniques centered on assessment receives much less attention.¹⁰

Indeed, assessments drive learning, and well-designed assessment tools ensure good evaluation of medical competencies, which are the core of integration.¹¹ Assessment strategies should be tailored in tandem with the advancement of medical curricula to Competency-based medical education (CBME). Collaborative efforts of medical educators, curriculum developers, and students would help in designing ideal realistic assessment tools anchoring clinical skills into the basic solid disciplines.¹² Bloxham and West reported that self and peer assessment strategies boost students' engagement and enhance motivations, which are well correlated with developing higher cognitive skills.¹³

Learning-oriented assessment (LOA) is a term proposed by Carless describing the formative and summative assessments and it comprises three main integral components [assessment of learning (AoL), assessment for learning (AfL) and assessment as learning (AaL)].¹⁴ The best practice of assessment recommends going beyond the AoL which is focused on summative assessments to AfL which is articulated on continuous formative assessment and feedback. Indeed, the highest level and the tip of the assessment pyramid is the AaL, where students self-regulate their own learning by using a learner-centered approach, enabling cognitive and skill reinforcement.¹⁵

Multiple choice questions are the main modality used for medical students' assessment, particularly formative assessment. One advantage of MCQs is that they can test a wide range of knowledge while successfully balancing validity and reliability with logistical ease.¹⁰ One advantage of MCQs is that they can test a wide range of knowledge while successfully balancing validity and reliability with logistical ease.¹⁶ Moving from the old-fashioned MCQs to clinical scenario-based MCQs is deemed critical. These case-based MCQs promote critical thinking and clinical relevance.¹⁷ Case-based MCQs are attractive, standardized assessments that can bring consistency to the evaluation process, ensuring an objective, reliable measure of students' knowledge and skills across large groups within a limited timeframe.¹⁸ Little is understood about the way medical students perceive and deal with case-based MCQs,¹⁹ particularly in pre-clinical phases, where dealing with such assessment tools is challenging.

In recent years, the concept of engaging students as active partners in assessment design has gained momentum worldwide, extending beyond regional contexts. Multicenter initiatives in German medical schools have demonstrated the feasibility and benefits of involving students in producing high-quality items assessing higher-order cognitive skills.⁹ Similarly, the University of Manitoba in North America has successfully involved students in question bank creation, reporting high satisfaction and perceived improvements in learning outcomes.²⁰ Nevertheless, implementation of student-generated assessments is not without challenges. Students generated by unexperienced creators might lack validity and

reliability, particularly if adequate training and faculty oversight are lacking.²¹ The process often requires substantial time and resource investment from educators to train students, rigorously review questions, and provide meaningful feedback.²² Moreover, variability in student motivation and readiness may affect the consistency and educational benefits of these activities.⁹

In this study, we hypothesized that placing AaL, AfL, and AoL side by side with students partnering is an innovative assessment practice that has a significant positive impact on the perception and performance of medical students in the early phases. Therefore, this work aimed to explore the perception and efficacy of engaging early-phase medical students in generating scenario-based MCQs targeting high cognitive levels in basic medical sciences.

Materials and Methods

Study Design and Context

This study is a quasi-experimental mixed method study consisting of two phases: [quantitative phase aiming to assess students' perception and performance and qualitative phase in the form of focused group discussions (FGD)]. The participants involved in this study were year two undergraduate medical students enrolled in the cardiovascular system (CVS) module taught as part of the Bachelor of Medicine and Surgery (MBBS) program at the College of Medicine (COM), Dar Al Uloom University (DAU), Riyadh, Kingdom of Saudi Arabia.

The COM, DAU, adopts seven years of an integrated hyper kid curriculum, where students typically enroll in four phases [university preparation phase (1 year), preclinical phase (2 years), clinical phase (3 years), and the internship year]. Students in the preclinical phases are exposed to several student-centered learning strategies, PBL, project-based learning, and Clinical skills and simulation activities. They are evaluated formatively and summatively, using theoretical exams and Objective Structured Practical Examinations (OSPE).

Sampling and Sample Size Calculation

We employed convenience sampling, where all consenting, eligible students who fulfilled the inclusion criteria were included. To ensure the investigated sample achieved the study objectives, a statistical power analysis was performed based on data from the current study ($N = 135$), comparing item creators versus non-creators. The effect size for a formative score in this study was 0.44, considered to be medium using Cohen's (1988) criteria,²³ with an $\alpha = 0.05$ and sample size = 66 in the item creator group and 69 in the Control group, a post hoc power analysis was conducted with this effect size using GPower version 3.1 software. It was approximately $(1-\beta) = 0.82$.²⁴ Thus, our power analysis, based on a sample size of 135 participants, proved adequate for the objectives of this study.

Inclusion and Exclusion Criteria

All students who were regularly registered in the CVS module during the study period and consented to participate in this study were considered eligible. Suspended students and those attending less than 70% of the CVS module were excluded. Additionally, we excluded patients with incomplete responses.

Ethical Consideration

This study was commenced after obtaining approval from the Institutional Review Board at COM, DAU. Following the Helsinki Declaration, data were anonymized, and every student gave written informed consent to publish anonymized quotes.

Intervention

Subjects Covered by the Questions

The CVS module includes thirteen pharmacology subjects, which were considered the core of this study. The subjects included alpha receptor blockers, Beta receptor blockers, pharmacology of antihypertensive drugs I and II, pharmacology of drugs used in the management of heart failure I and II, pharmacology of antianginal therapy I and II, psychological therapy of dyslipidemia I and II, pharmacological therapy of dysrhythmias I and II and thrombolytic therapy.

Recruiting and Training Students on Item Creation

After finalizing all pharmacology-related sessions, a tutorial was scheduled where students were briefed on the guidelines to prepare scenario-based MCQs. Simplified blueprints were shown for students, and the key considerations for developing MCQs, including the different levels of cognition according to Bloom's taxonomy were explained by the tutor, who is also a medical education expert. Students were divided into 15 teams, and along with the tutor's assistance, every team created two or three items, and these items were discussed and solved with the remaining students during the session. After ensuring that all teams understood the idea, every student was requested to choose the subjects willing to create items, and a deadline of three days was determined to submit the created items to the tutor.

Formulation of the Questions

Students were given handouts showing the criteria for establishing valid MCQs. All submitted questions should contain functioning clinical scenarios, and the complexity level showed to be variable across the items. Questions should contain a stem, lead-in, three distractors, and one key representing the best answer or response. Every item should be submitted in a simple blueprint form, linking the question to the learning objective.²⁵ Students can submit as few as one item and a maximum of 10 items. Students were requested to work individually, not sharing the created items with their colleagues.

Revising the Formulated Questions

The submitted questions were thoroughly revised for their scientific rigor. The phrasing of the item, homogeneity of distractors, and grammatical and spelling mistakes were fixed. Any duplication, irrelevant information, clues in the choices, or negative terms were also resolved, and the item was labeled as ready.

Feedback

As feedback is an integral component of AfL, a constructive feedback session was conducted for every item creator, highlighting the points of strength and weakness in the submitted item, along with the assessed level of cognition. The suggested modifications were also mentioned with justification. Students were encouraged to reflect on their experiences.

Mock Exam

A predetermined date for the mock exam was announced, and students were requested to be prepared and study the 13 subjects. The interval between the student MCQ generation activity and the subsequent mock exam was approximately one week. This timeframe was chosen to allow students sufficient opportunity to engage with the material and reflect on their created questions before assessment. The mock exam consisted of 25 items, accounting for a total of 50 marks, and students had to complete the exam individually within one hour and 15 minutes, considering that 1.5 minutes were allocated for every single item. The mock exam was completed, and students could see their grades once they submitted their answers, with feedback on every solved item.

Grouping

Students were categorized into two groups: one group involved students who participated in item creation, regardless of the number of items created (Intervention group), and the other group included students who did not create any items, although they were involved in the mock exam (Control group).

Data Collection Tool

Self-Administered Questionnaire and Student Performance

After completing the mock exam, students were referred to fill out a self-administered questionnaire sent through the same link of the exam. In this questionnaire, students report their sex (male, female) and the number of items they created. Students involved in the Mock exam but not participating in item creation were also eligible to fill out the survey showing zero created items. All students responded to the item "To which extent you are satisfied with this way in dealing with scenario-based MCQs" on a five-point Likert scale where the scores varied between very satisfactory, satisfactory, neutral, unsatisfactory, and very unsatisfactory. In addition to students' performance in the mock exam, students reported their cumulative Grade Point Average (GPA).

Focused Group Discussion (FGD)

Following the criteria established by Krueger,²⁶ we conducted an FGD to assess students' feedback on the adopted assessment strategy. The FGD consisted of the tutor, a medical education expert, a scribe, and five students (three females and two males). A semi-structured guide was used to guide the conversation, which lasted for 45 minutes. Three open-ended questions were asked to allow students to reflect on their own experiences, including the following:

- i. How do you feel that you get benefits from student-led approach to generate clinical scenario-based MCQs?
- ii. What are the advantages of student-generated clinical scenario-based MCQs compared to tutor-generated MCQs?
- iii. What are other suggested actions you feel would improve your performance in the scenario-based MCQs?

Rigor

Quantitative Data

The survey validity was determined by two experts in medical education, who estimated a time of 3 minutes to fill out the survey. Moreover, we calculate a content validity index and ratio of 0.91 and 1.00, respectively, denoting that all items were relevant.^{27,28} A Cronbach's Alpha of 0.79 was calculated showing very good reliability and the internal consistency reliability was computed using Spearman–Brown Prophecy formula ($r1 = 2(r)/1+r$) where r estimated correlation coefficient computed on the split halves and $r1$ estimated reliability of the entire test and it was 0.878 which represented adequate internal consistency.²⁹

Qualitative Data

Two strategies were applied to ensure the reflexivity of the data analysts, including reflective writing and collaborative reflection.³⁰ Additionally, continuous data analysis and triangulation were deployed until no new themes emerged from the data that was analyzed by two investigators. An inter-rater reliability and the consistency of analysis between the two investigators revealed an overall reliability of 0.89.³¹

Data Analysis

The quantitative data were analyzed using IBM Statistical Package for the Social Sciences Program (SPSS) Statistics for Windows, version 25 (IBM Corp., Armonk, NY, USA). The chi-square test was applied to analyze the relationship between qualitative variables. Comparisons of quantitative data between the two groups were conducted using the Mann–Whitney test (non-parametric). The Spearman-rho was used to assess the association between satisfactory levels, formative and summative exam scores and no. of questions developed by the students. A p -value ≤ 0.05 was considered significant. The significance level was set to P value < 0.05 and a 95% confidence interval (CI). Cohen's d was calculated to determine the effect size for differences in formative assessment scores and GPA between item creators and non-creators, providing a standardized measure of the magnitude of these differences. Additionally, Cramér's V was computed for Chi-square tests to quantify the strength of association between categorical variables, offering an interpretable effect size alongside significance values.

Qualitative data were analyzed using QSR NVivo version 12. Thematic analysis was conducted following the approach described by Braun and Clarke.³² We adopted the inductive approach, where the initial codes were generated through careful reading of the scribed text. From the generated codes, we identified categories and reviewed the potential themes, which were then phrased.

Results

Quantitative Analysis

The study included 135 students, with males comprising 54.1% ($n=73$) and females 45.9% ($n=62$). Approximately 48.9% ($n=66$) of the students created questions, while 51.1% ($n=69$) did not. Among the 66 item creators, the number of questions ranged from 1 to 10, with a mean of 3.6 and SD of ± 2.6 . Figure 1 shows the distribution and percentage of the questions created by students, where 27.3% created one item and 7.5% created 10 items.

The mean formative score was 31.9 ± 10.8 , where the majority were in the B+ (31.9%) and B (26.7%) categories. Regarding satisfaction, 51.9% were very satisfied, and only 3% were unsatisfied, reflecting a generally positive learning

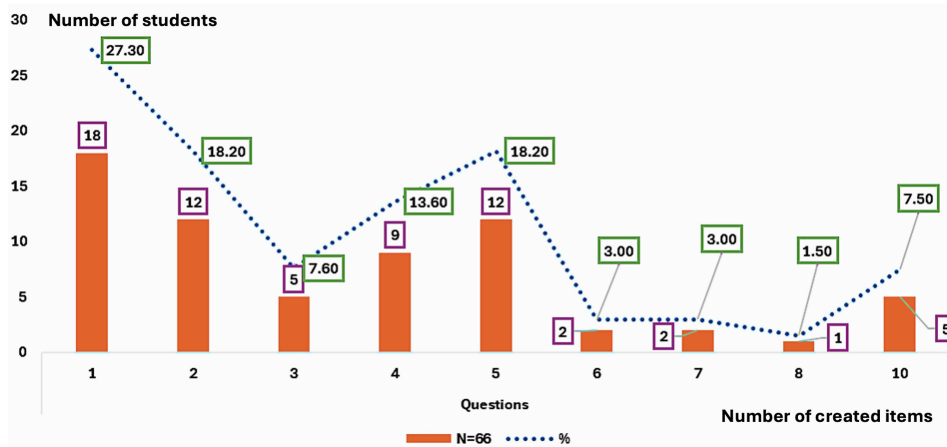


Figure 1 Distribution and proportion of the studied students regarding the number of created items.

experience. Students who created questions (n=66) had significantly higher formative scores (34.3 ± 10.5) compared to non-creators (29.6 ± 10.6), with a p-value of 0.010. Their GPA was also significantly better (eg, B+ grade: 42.4% vs 21.7%; $p=0.015$). Despite these academic differences, satisfaction levels were similar between both groups, where above 75% of participants were satisfied or very satisfied ($p=0.993$) (Table 1).

The current study conveys a strong positive correlation between the scores in formative assessment and the GPAs ($r_s = 0.74, p < 0.001$), and a moderate correlation between the number of questions developed and scores of the formative assessment ($r_s = 0.39, p = 0.001$) (Figure 2). The majority of students exhibited high satisfaction levels, regardless of their GPAs, indicated by the non-significant correlation between satisfaction level and GPA ($p > 0.05$) as shown in Table 2.

As Table 3 depicts, there were no statistically significant differences between male and female students in creating questions ($p=0.651$), number of developed questions (mean 3.9 ± 2.9 for males vs 3.1 ± 2.1 for females, $p=0.250$), or

Table 1 Comparison Between the Students Who Developed and Those Who Did Not Develop Questions Regarding Their Response

	Item Creator (n=66)		Non Creator (n=69)		Total		Effect Size [95% CI]	Test of Sig	P Value
	No	%	No	%	No	%			
Formative score Mean± SD	34.3±10.5		29.6±10.6		31.9±10.8		0.45[0.11–0.79]	2.63	0.010*
Cumulative GPA							0.32[0.24–0.49]	14.02	0.015*
• C	5	7.6	6	8.7	11	8.1			
• C+	0	0.0	7	10.1	7	5.2			
• B	13	19.7	23	33.3	36	26.7			
• B+	28	42.4	15	21.7	43	31.9			
• A	12	18.2	12	17.4	24	17.8			
• A+	8	12.1	6	8.7	14	10.4			
Students' Satisfaction							0.06[0.04–0.27]	0.09	0.993
• Unsatisfactory	2	3.0	2	2.9	4	3.0			
• Neutral	14	21.2	15	21.7	29	21.5			
• Satisfactory	15	22.7	17	24.6	32	23.7			
• Very satisfactory	35	53.0	35	50.7	70	51.9			

Note: *Significance at $p < 0.05$.

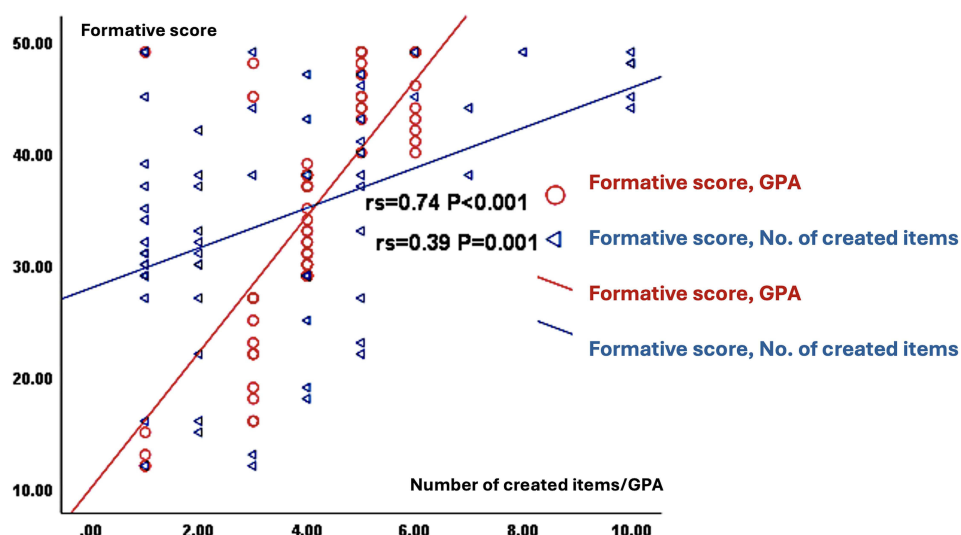


Figure 2 Correlations between satisfactory levels, formative scores and cumulative GPA and no of questions developed by the students.

formative scores (31.8 ± 10.9 vs 31.9 ± 10.7 , $p=0.929$). The GPA and satisfaction levels also did not differ significantly between sexes ($p=0.760$ and $p=0.094$, respectively). These results suggest comparable performance and perceptions between genders.

Table 2 Correlation Between the Satisfactory Levels, Formative Scores, and GPA and the Number of Questions Developed by the Students

	Satisfactory Level		Cumulative GPA		No of Questions	
	r_s [95% CI]	P Value	r_s [95% CI]	P Value	r_s [95 CI %]	P Value
Formative score	0.02 [-0.14–0.20]	0.792	0.74[0.65–0.81]	<0.001*	0.39[0.16–0.58]	0.001*
No of questions	-0.04[-0.28–0.21]	0.764	0.16[-0.09–0.39]	0.193	–	–
Cumulative GPA	0.03[-0.14–0.21]	0.664	–	–	–	–

Note: *Significance at $p < 0.05$.

Table 3 Comparison Between the Male and Female Students Regarding Their Response

	Males (n=73)		Females (n=62)		Effect Size [95% CI]	Test of Sig	P Value
	No	%	No	%			
Students					0.4[0.004–0.221]	0.20	0.651
• Item creator	37	50.7	29	46.8			
• Non creator	36	49.3	33	53.2			
Number of developed questions	3.9±2.9		3.1±2.1		0.33[-0.15–0.83]	1.15	0.250
Formative score	31.8±10.9		31.9±10.7		-0.02[-0.35–0.32]	0.08	0.929
Cumulative GPA					0.14[0.10–0.36]	2.60	0.760
• C	6	8.2	5	8.1			
• C+	4	5.5	3	4.8			
• B	23	31.5	13	21.0			
• B+	21	28.8	22	35.5			
• A	11	15.1	13	21.0			
• A+	8	11.0	6	9.7			

(Continued)

Table 3 (Continued).

	Males (n=73)		Females (n=62)		Effect Size [95% CI]	Test of Sig	P Value
	No	%	No	%			
Satisfaction					0.22[0.10–0.39]	6.38	0.094
• Unsatisfactory	1	1.4	3	4.8%			
• Neutral	11	15.1	18	29.			
• Satisfactory	21	28.8	11	17.7			
• Very satisfactory	40	54.8	30	48.4			

Qualitative Analysis

Analyses of narrative responses in the FGD yielded four interrelated but distinct themes, as follows:

Theme (1) Student-Led Approach to Generate Clinical Scenario-Based MCQs Has a Positive Steering Effect on the Learning Process in the Early Phases of Medical Curricula

Students in the FGD reported enhanced recall and understanding while involved in solving an exam they have participated in its creation. Additionally, they described good critical appraisal and clinical reasoning in creating an exam for their colleagues. Many students expressed that they learned significantly from the mock exam and found it beneficial.

“This experience helped to improve my educational skills to come up with high quality flawless items”.

“This task was amazing! It really helped us improve our skills and made tough questions that challenged us much easier”.

“The mock exam motivated us to review previous lectures and maintain a consistent study routine”.

Theme (2) Student-Led Approach to Generate Clinical Scenario-Based MCQs Fosters Procedural Soft Skills and Values of Pre-Clinical Medical Students

Students conveyed that they felt more confident and not worried during the exam. They reported that participation in exam preparation improved their professional identity and allowed them to feel like real doctors.

“Participating in the exam preparation made me truly felt like I was stepping into the role of a physician”.

Theme (3) Student-Led Approach to Generate Clinical Scenario-Based MCQs Allows Positive Emotions and Feelings of Entertainment During Learning

This theme also included text fragments that showed that students valued the feeling of joy and entertainment, the motivation and positive emotions they felt when seeing their items in the exam. The collaborative nature of the assessments made the scenarios more engaging and relatable, fostering a deeper interest in the material.

“This was the best experience I have ever had in the college. Thank you for making it enjoyable!”.

“The positive emotions I felt while preparing for the exam motivated me to engage more deeply with the material”.

“I was no longer worried during the exam; participation in exam preparation really eased my anxiety”.

Theme (4) Student-Led Approach to Generate Clinical Scenario-Based MCQs Changed the Perception of Students About Scenario-Based MCQs

Students highlight how the implementation of student-led approach to generate clinical scenario-based MCQs transformed their views on scenario-based MCQs from challenging or intimidating to the feeling of relevance and applicability of these scenarios to real-world medical practices.

“I found scenario-based MCQs much more engaging than before”.

“I used to dread MCQs, but now I see them as a valuable learning tool”.

Theme (5) Student-Led Approach to Generate Clinical Scenario-Based MCQs Could Be Successfully Adopted in Other Subjects

Students value practicing generating scenario-based MCQs in further subjects, and they thought it would make complex topics much easier to understand.

“I believe this approach could enhance learning in subjects like anatomy and physiology, just like it did in pharmacology”.

“I can imagine how effective generating scenario-based MCQs would be in my biochemistry class—applying concepts through real-life scenarios would be amazing!”.

“Seeing how well generating scenario-based MCQs worked in pharmacology makes me excited to see it implemented in other areas of our curriculum”.

Discussion

The obtained findings emphasized the significant role of involving students in their assessment as an integral component of the learning process. The intervention used in this work was designed to mirror phases of the regulated learning model proposed by Panadero et al, including preparation, implementation, and appraisal.³³ This study focuses on the appraisal of this form of SRL from the student’s perspective. The adopted strategy was significantly associated with higher academic performance, as well as a high level of satisfaction with this pedagogical approach.

This study highlights the importance of collaboration between medical educators and students, which is consistent with a holistic approach to LOA.³⁴ While in AfL, students receive feedback after formative assessments, in AaL, students take control by dominating the learning process’s discourse and producing a self-regulatory and self-productive identity.¹⁵ In this study, students showed a very good satisfaction with the adopted learning approach. Similarly, Lakhtakia et al reported similar findings about senior final-year medical students.²² Likewise, about 91% of medical students reported satisfaction with their involvement in generating a question bank in an earlier study.²⁰

Student-generated formative assessments are powerful tools for SRL, with an overall high level of satisfaction and positive impacts on the academic performance of medical students.²² Consistently, Khatskevich et al reported that students involved in creating anatomy mock exams for their colleagues significantly outperformed their peers on the official practical examination.³⁵ These outcomes suggest that student-generated item creation positively influences the learning process by enhancing recall, understanding, critical appraisal, and clinical reasoning skills. From a constructivist perspective, the act of creating assessment items encourages deeper engagement with the subject matter, moving beyond surface-level memorization to higher-order thinking skills, including analysis, synthesis, and evaluation.³⁶ The act of creating assessment items necessitates a deeper engagement with the subject matter, moving beyond surface-level memorization to higher-order thinking skills such as analysis, synthesis, and evaluation. This active learning approach, where students are charged with generating assessments, reinforces student-centered in-depth learning and critical thinking.²² A previously conducted study in a dental college revealed that students could generate MCQs assessing high-order cognitive skills more effectively than tutors.³⁷

Students recognized the intervention used as a learning opportunity that enhanced their understanding, critical appraisal, and cognition.³⁸ A unique example of involving students in generating MCQs that target high order thinking skills is demonstrated in a multicenter pilot study conducted in German medical schools. Seventeen medical schools participated in this study, where a selected team, including 17 students, was trained in creating MCQs. The generated items were of high quality, assessing higher order thinking skills and showing high test reliability. Item creators perceived this experience as a feedback opportunity rather than a learning opportunity.⁹ The development of higher cognitive skills is best accomplished through inquiry and investigation, as well as applying core knowledge to solve problems and find solutions to new situations.³⁹ The current findings emphasize how engaging students in creating scenario-based MCQs

motivates them to study the core content of the course, with less anxiety and stress before exams, which is consistent with a previous study.²² Indeed, the process of designing questions requires students to identify key concepts, anticipate common misconceptions, and formulate clear and unambiguous questions, all of which contribute to a more robust understanding of the material fostering SRL.²²

According to the current study, one reported advantage of teaching basic sciences in clinical contexts for early medical students is the improvement of professional identity. Duban et al reported that students involved in the experimental curriculum, where clinical skills were taught earlier showed an increase in the perceived value of basic sciences in their future careers and higher performance in future clinical science subjects.⁴⁰ Involving medical students in such engaging activities may help bridge the known gaps in supporting professional identity within medical curricula. Most medical schools consider the professionalism course the primary source for advocating a professional identity among medical students.⁴¹

It is noteworthy to mention that our students were eager to practice generating case-based MCQs in other basic disciplines. For example, second-year biomedical sciences students who engaged in generating MCQs in a biochemistry course expressed a desire to develop a repository of high-quality items.⁴² Cantor et al stated that “development of the brain is an experience-dependent process”.⁴³ As a function of experiences, the human brain and capacities grow across an interactive developmental spectrum (physical, cognitive, affective). What happens in one domain influences other domains, emphasizing the effect of emotion on both developing and suppressing cognition.³⁹

Another unexpected benefit expressed by students was the enjoyment they derived from item creation. This shift in perception transformed MCQs from a challenging task into a valuable learning opportunity. The positive emotions associated with this process further underscore the effectiveness of involving students in generating high-level MCQs, creating a motivating learning environment. Students appreciated the joy and motivation they felt when seeing their items included in exams, and the collaborative nature of the assessments made scenarios more engaging and relatable. Literature suggests that assessments that stimulate autonomous motivation, particularly those perceived as enjoyable, can lead to increased effort and a stronger connection to the learning material.⁴⁴ In an earlier study conducted by Lakhtakia et al among medical students, item creators described the sense of gaming while creating MCQs. Students described that tossing around distractors became a second habit to them during time of relaxation.²² Furthermore, Gray considered the creativity the main source of critical thinking. Engagement in enjoyable learning has an ever-lasting impact, sustaining self-learning and building peer-learning habits.⁴⁵

However, it is crucial to acknowledge limitations, particularly concerning selection bias. The self-selection of students into the item creation process may mean that more motivated and capable students participated, which could confound our results. This aspect should be considered when interpreting the outcomes, as it may limit the generalizability of our findings. Though students in the orientation session worked in teams according to their preference to formulate preliminary items, they were requested to work individually while creating the final items. Despite this approach introduced self-selection bias, it minimized the effect of team dynamics on students' perceptions. These effects are contradictory in the literature. It is not overtly apparent whether doing the current exercise as part of a team is well perceived or not by students. While some previous works have considered teamwork an advantageous approach in creating MCQs,^{20,22} others were dissatisfied with team performances.³⁷ Nonetheless, peer interactions can enhance learning by allowing students to clarify doubts and collectively refine their understanding, ultimately leading to improved academic performance. Research supports that student-led initiatives and peer-assisted learning can enhance confidence and educational outcomes.³⁵

Strength and Limitation

The main strength of this study is its use of a mixed-methods approach, which helps to clarify the elements of SRL that can be consistently replicated over time. This approach also highlights opportunities to tackle perceived challenges and weaknesses in an adaptive manner. The nature of the intervention tool is another unique strength of this study. While in all previous studies, students were requested to create MCQs, in the current study, the created MCQs should be of the scenario-based type, which is known to be challenging, particularly for medical students in the early phases of the program.

One significant limitation is the limited psychometric analysis of item difficulty and discrimination that was not visited in depth. Future research should incorporate comprehensive item-level analyses to augment the interpretability and robustness of assessment outcomes. Additionally, the duration of the intervention was limited. It was insufficient to determine if this intervention leads to long-term knowledge retention or improved performance in subsequent clinical years. Longitudinal studies are needed to investigate the long-term impact of the adopted strategy. Furthermore, conducting this study enrolling only second year medical students is another limitation. It would be valuable for future research to investigate whether different stages of learning influence students' understanding and perceptions of creating MCQs and assessing high order thinking skills by gathering feedback from students at various points in their educational journey. Moreover, future studies could incorporate a variety of assessment formats to give students better insights into their learning strategies.

Another limitation to be considered when interpreting the results is the quasi-experimental design that permitted self-selection into intervention and control groups, potentially introducing selection bias. Students volunteering to write questions may differ in motivation, prior preparation, or study habits compared to non-creators, complicating causal attribution of performance differences. Lastly, since this research focused on a single course with one cohort, the findings may have limited generalizability. Expanding the study to include multiple cohorts could help address this issue. A follow-up exercise utilizing the same framework in subsequent cohorts will strengthen the tool's reproducibility as well as the consistency and generalizability of the findings.

Conclusion

The student-led approach to generating clinical scenario-based MCQs shows significant promise as an innovative pedagogical method in medical education. By actively involving students in the assessment creation process, educators can foster a deeper understanding of course material, enhance critical thinking and clinical reasoning skills, and promote self-regulated learning (SRL). Engaging students in the assessment process represents a viable strategy for improving the medical curriculum. The adopted approach encourages active learning among medical students and fosters a sense of ownership over their educational assessments. Involving students in generating high-level cognitive MCQs serves not merely as an evaluation tool but as a versatile pedagogical strategy that can be adapted across various disciplines to enhance learning and engagement. While this study demonstrates an association between student-led item creation and positive outcomes, the obtained findings must be interpreted cautiously, given the inherent limitations of the study design, including the self-selection bias. Further research may be needed to establish causative effects. Implementing the adopted strategies has the potential to create a more integrated and effective curriculum, assisting students in their transition to the clinical years and preparing future medical professionals with not only knowledge but also essential soft skills and a positive attitude toward continuous learning and assessment.

Institutional Review Board Statement

This study was commenced after obtaining approval from the Institutional Review Board (IRB) at COM, DAU (approval number HP-01-R-134-DAU-COM-25-01). Following the Helsinki Declaration, data were anonymized, and every student gave written informed consent to publish anonymized quotes.

Abbreviations

AaL, assessment as learning; AfL, assessment for learning; AoL, assessment of learning; CBME, competency-based medical education; COM, College of Medicine; CVS, cardiovascular system; DAU, Dar Al Uloom University; FGD, focused group discussions; GPA, Grade Point Average; LOA, learning-oriented assessment; MCQs, Multiple choice questions; OSPE, Objective Structured Practical Examinations; PBL, problem-based learning; SRL, self-regulated learning; TBL, team-based learning.

Data Sharing Statement

The data analyzed in the current study are available upon reasonable request from the corresponding author.

Informed Consent Statement

Written informed consent was obtained from every participant prior to their inclusion in the current study.

Acknowledgments

The authors extend their appreciation to year 2 medical students at the College of Medicine, Dar Al-Uloom University.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

The authors did not receive any support from funding or nonprofit agencies for this study.

Disclosure

The authors report no competing interests (financial or non-financial) related to this work.

References

- Frenk J, Chen L, Bhutta ZA, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet*. 2010;376(9756):1923–1958. doi:10.1016/S0140-6736(10)61854-5
- Oktar S. Basic medical sciences should be mainly taught by clinicians for a tight integration of basic and clinical sciences in medical education. *Acta Medica Alanya*. 2022;6(3):320–325. doi:10.30565/medalanya.1159175
- Brass EP. Basic biomedical sciences and the future of medical education: implications for internal medicine. *J Gen Intern Med*. 2009;24(11):1251–1254. doi:10.1007/s11606-009-0998-5
- Spencer J. Learning and teaching in the clinical environment. Cantillon P, Wood D, editors. *ABC of Learning and Teaching in Medicine*. 2nd. Wiley-Blackwell: 2003. Available from: <https://www.researchgate.net/publication/10936581>. Accessed October 07, 2025.
- Harden RM, Crosby J. *The Good Teacher Is More Than a Lecturer: The Twelve Roles of the Teacher*. Vol. 22. AMEE; 2000.
- Cooke M, Irby DM, Sullivan W, Ludmerer KM. American medical education 100 years after the Flexner report. *N Engl J Med*. 2006;355(13):1339–1344. doi:10.1056/NEJMra055445
- Prober CG, Khan S. Medical education reimaged: a call to action. *Acad Med*. 2013;88(10):1407–1410. doi:10.1097/ACM.0b013e3182a368bd
- Taylor DCM, Hamdy H. Adult learning theories: implications for learning and teaching in medical education: AMEE Guide No. 83. *Med Teach*. 2013;35(11):e1561–e1572. doi:10.3109/0142159X.2013.828153
- Wagener S, Möltner A, Timbil S, et al. Development of a competency-based formative progress test with student-generated MCQs: results from a multi-centre pilot study. *GMS Z Med Ausbild*. 2015;32(4):Doc46. doi:10.3205/zma000988
- Harris BHL, Walsh JL, Tayyaba S, Harris DA, Wilson DJ, Smith PE. A novel student-led approach to multiple-choice question generation and online database creation, with targeted clinician input. *Teach Learn Med*. 2015;27(2):348–352. doi:10.1080/10401334.2015.1011651
- Sohrman M, Berendonk C, Nendaz M, Bonvin R; Swiss Working Group for PROFILES Implementation T. Nationwide introduction of a new competency framework for undergraduate medical curricula: a collaborative approach. *Swiss Med Wkly*. 2020;150(1516):w20201–w20201. doi:10.57187/smw.2020.20201
- Ferris H, O'Flynn D. Assessment in medical education; what are we trying to achieve? *Int J Higher Educ*. 2015;4(2):139–144. doi:10.5430/ijhe.v4n2p139
- Bloxham* S, West A. Understanding the rules of the game: marking peer assessment as a medium for developing students' conceptions of assessment. *Assess Eval High Educ*. 2004;29(6):721–733. doi:10.1080/0260293042000227254
- Carless D. Learning-oriented assessment: conceptual bases and practical implications. *Innovations Educ Teach Int*. 2007;44(1):57–66. doi:10.1080/14703290601081332
- Dann R. Assessment as learning: blurring the boundaries of assessment and learning for theory, policy and practice. *Assess Educ*. 2014;21(2):149–166. doi:10.1080/0969594X.2014.898128
- Hays R. Assessment in medical education: roles for clinical teachers. *Clin Teacher*. 2008;5(1):23–27. doi:10.1111/j.1743-498X.2007.00165.x
- Al Shuriaqi S, Aal Abdulsalam A, Masters K. Generation of medical case-based multiple-choice questions. *Int Med Educ*. 2023;3(1):12–22. doi:10.3390/ime3010002
- Kurdi G, Leo J, Parsia B, Sattler U, Al-Emari S. A systematic review of automatic question generation for educational purposes. *Int J Artif Intell Educ*. 2020;30(1):121–204. doi:10.1007/s40593-019-00186-y
- Chéron M, Ademi M, Kraft F, Löffler-Stastka H. Case-based learning and multiple choice questioning methods favored by students. *BMC Med Educ*. 2016;16(1):1–7. doi:10.1186/s12909-016-0564-x
- Gooi ACC, Sommerfeld CS. Medical school 2.0: how we developed a student-generated question bank using small group learning. *Med Teach*. 2015;37(10):892–896. doi:10.3109/0142159X.2014.970624
- Arekkuzhiyil S, Professor A. Issues and concerns in classroom assessment practices. *Arekkuzhiyi Santhosh*. 2021;20(8):20–23.

22. Lakhtakia R, Otaki F, Alsuwaidi L, Zary N. Assessment as learning in medical education: feasibility and perceived impact of student-generated formative assessments. *JMIR Med Educ.* 2022;8(3):e35820. doi:10.2196/35820
23. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Law-Rence Erlbaum Associates.; 1988.
24. Faul F, Erdfelder E, Lang AG, Buchner A. GPower. 2007.
25. Chauhan A, Mahajan A, Begum J, Kapoor A, Badyal D, Kulkarni S. Designing effective scenario-based multiple-choice questions for health professionals. *Nat Med J India.* 2024;37(6):335–339. doi:10.25259/NMJI_1004_2023
26. Krueger RA. *Focus Groups: A Practical Guide for Applied Research*. Sage publications; 2014.
27. Yusoff MSB. ABC of content validation and content validity index calculation. *Educ Med J.* 2019;11(2):49–54. doi:10.21315/eimj2019.11.2.6
28. Genareo V. Using microsoft excel to calculate content validity index (CVI) and content validity ratio (CVR): a practical approach. *Quality Assurance Symposium AAQEP.* 2023.
29. Dawson-Saunders B. *Basic & Clinical Biostatistics. ALANGE Medical Book.* 1994:42–161.
30. Olmos-Vega FM, Stalmeijer RE, Varpio L, Kahlke R. A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. *Med Teach.* 2023;45(3):241–251. doi:10.1080/0142159X.2022.2057287
31. Miles MB, Huberman AM. *Qualitative Data Analysis: An Expanded Sourcebook*. sage; 1994.
32. Braun V, Clarke V. Research designs: quantitative, qualitative, neuropsychological, and biological. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf DSK, editors. *Handbook of Research Methods in Psychology*. Ameri Psychol Assoc; 2012:701.
33. Panadero E. A review of self-regulated learning: six models and four directions for research. *Front Psychol.* 2017;8:422. doi:10.3389/fpsyg.2017.00422
34. Zeng W, Huang F, Yu L, Chen S. Towards a learning-oriented assessment to improve students' learning—a critical review of literature. *Educ Assess Eval Account.* 2018;30(3):211–250. doi:10.1007/s11092-018-9281-9
35. Khatskevich K, Hewitt ST, Jang CW, et al. Benefits of implementing student-led review and mock examinations in the medical undergraduate gross anatomy curriculum. *Med Teach.* 2022;44(9):973–976. doi:10.1080/0142159X.2022.2049733
36. Cobern WW. Constructivism. *J Educ Psychol Consultation.* 1993;4(1):105–112. doi:10.1207/s1532768xjepc0401_8
37. Gonzalez-Cabezas C, Anderson OS, Wright MC, Fontana M. Association between dental student-developed exam questions and learning at higher cognitive levels. *J Dent Educ.* 2015;79(11):1295–1304. doi:10.1002/j.0022-0337.2015.79.11.tb06025.x
38. Pangaro L, Ten Cate O. Frameworks for learner assessment in medicine: AMEE Guide No. 78. *Med Teach.* 2013;35(6):e1197–e1210. doi:10.3109/0142159X.2013.788789
39. Darling-Hammond L, Flook L, Cook-Harvey C, Barron B, Osher D. Implications for educational practice of the science of learning and development. *Appl Dev Sci.* 2020;24(2):97–140. doi:10.1080/10888691.2018.1537791
40. Duban S, Mennin S, Waterman R, et al. Teaching clinical skills to pre-clinical medical students: integration with basic science learning. *Med Educ.* 1982;16(4):183–187. doi:10.1111/j.1365-2923.1982.tb01245.x
41. Sarraf-Yazdi S, Teo YN, How AEH, et al. A scoping review of professional identity formation in undergraduate medical education. *J Gen Intern Med.* 2021;36(11):3511–3521. doi:10.1007/s11606-021-07024-9
42. Bottomley S, Denny P. A participatory learning approach to biochemistry using student authored and evaluated multiple-choice questions. *Biochem Mol Biol Educ.* 2011;39(5):352–361. doi:10.1002/bmb.20526
43. Cantor P, Osher D, Berg J, Steyer L, Rose T. Malleability, plasticity, and individuality: how children learn and develop in context 1. In: *The Science of Learning and Development*. Routledge; 2021:3–54.
44. Kusrurkar RA, Orsini C, Somra S, et al. The effect of assessments on student motivation for learning and its outcomes in health professions education: a review and realist synthesis. *Acad Med.* 2023;98(9):1083–1092. doi:10.1097/ACM.0000000000005263
45. Gray P. *Free to Learn: Why Unleashing the Instinct to Play Will Make Our Children Happier, More Self-Reliant, and Better Students for Life*. Hachette UK; 2013.

Advances in Medical Education and Practice

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>

Dovepress
Taylor & Francis Group