

The Generation and Use of Medical MCQs: A Narrative Review

Sinclair Steele¹, Niranjan Nayak², Yehia Mohamed^{1,3}, Debadatta Panigrahi¹

¹Department of Pathological Sciences, Ajman University College of Medicine, Ajman, United Arab Emirates; ²Department of Microbiology, Manipal College of Medical Sciences, Pokhara, Nepal; ³Department of Microbiology and Immunology, Faculty of Pharmacy (Boys) Al-Azhar University, Cairo, Egypt

Correspondence: Sinclair Steele, Ajman University College of Medicine, Ajman, United Arab Emirates, Tel +971 06 748 2222, Email s.steele@ajman.ac.ae

Abstract: The use of multiple-choice questions in medical assessments is ubiquitous and extends in breadth from undergraduate formative and summative assessments to postgraduate certifications and reaccreditations. Despite the clear importance of multiple-choice questions in medicine, in the last thirty years there has been no authoritative peer reviewed publication that reviews the overall generation and use of multiple-choice questions in medicine. Producing effective and useful medical multiple-choice questions is as much of an art form as it is a science - unfortunately, every new writer of such questions tends to rely on their own experiences of such questions and then travel along their individual and tortuous pathway to produce “acceptable” questions. Our review article provides practical guidance and information that will assist writers of medical multiple-choice questions. As well as extensively discussing common pitfalls, we provide a checklist that represents optimum practice in question generation. We directly compare examples of good and poor multiple-choice questions. We describe the metrics for measuring effective questions; specifically, the discrimination and difficulty indices. We discuss the optimum approach to training faculty to produce *flawless* multiple-choice questions. We describe the potential role and benefits of students in writing multiple-choice questions. We address the current and future effectiveness of the major artificial intelligence software brands in developing objectively functional multiple-choice questions. We continue to support the pragmatic use of MCQs in undergraduate and postgraduate medical training, with the *caveat* that ongoing training is made available to the creators in order to facilitate the production of the highest quality questions.

Keywords: multiple-choice questions, difficulty index, discrimination index, problem-based question, undergraduate education, postgraduate education, summative assessment

Introduction

Formative and summative assessments are key components in both driving and quantifying educational progress in the teaching and learning environment. Worldwide, the most popular item tool that is used in medical educational summative and formative assessments is the *multiple-choice question* (MCQ).¹ This versatile question type can be used to assess recall, understanding, application, analytical skills and evaluation. Accordingly, it is also an excellent tool for evaluating medical competencies.² The ubiquity of MCQs at all levels of medical education indicates that essentially all educators are required to be able to write reliable MCQs. So, this particular skill is a *sine qua non* for a medical educator/faculty member. Regrettably, the quality of MCQs produced by medical educators untrained in the preparation of MCQs can be variable and is often very poor.³⁻⁶ For example, Bhat and Prasad’s item analysis of an MB BS undergraduate ophthalmology question bank showed that 50% of the MCQs were flawed, the majority being due to bad stems or distractors.⁶ DiSantis et al’s analysis of Continuing Medical Education (CME) MCQs demonstrated that 43% of questions in major radiology journals were flawed.⁵ Furthermore, in 2006 Stagnaro-Green and Downing showed that 40 of 40 evaluated CME questions in the *New England Journal of Medicine* were flawed - all with poor stems.⁷ Several studies have shown a quantifiable improvement in MCQ writing quality, with fewer stem flaws and non-functioning distractors, after the inexperienced MCQ creators received training sessions.^{3,4,8} For example, Gupta et al’s 2020 study showed a statistically significant reduction in item writing flaws from 27.7% to 15% after just one training session, with

most of the improvement being in the writing of answer options.⁴ Dellinges and Curtis's 2017 study demonstrated a similar improvement in MCQ quality after a training session lasting for just one hour.⁸ This clearly indicates the need for readily available MCQ writing guidance for inexperienced question creators. Consequently, a major purpose of this narrative review is to be a primary introductory source of guidance to inexperienced medical MCQ writers. A large part of the motivation to create this narrative review was the difficulty we noticed that less experienced faculty members had in generating multiple-choice questions that could overcome the vetting procedures in place to guarantee high quality questions that were fit for purpose. This local observation was concordant with the variable standard of MCQ writing by inexperienced creators indicated in the international literature.³⁻⁶ Accordingly, a major objective of our review is to assist our less experienced colleagues; to facilitate question writing by providing structural guidance on good question writing, to indicate areas that require caution because of common errors and to indicate objective established criteria for evaluating the appropriateness of a question. We also understand that inexperienced faculty/question writers should be directed to better strategies for writing questions that make use of formal training resources⁹ and their colleagues' experience. Alerting these inexperienced writers to the latest student and artificial intelligence approaches to generating multiple-choice questions can further assist these writers in developing a personal algorithm that will more reliably and consistently produce high quality MCQs.

Medical training is uniquely intense because of the ultimate professional responsibilities of qualified doctors. Accordingly, before any medical educational institution can certify that the student is appropriately accomplished, a high level of certainty needs to be reached regarding the student's knowledge and clinical skills. In this regard, multiple-choice questions are particularly valuable as an integral part of formative and summative assessments¹⁰⁻¹² in medical training because they can be equally applied to undergraduate and postgraduate education and can be extended to the more than 60 medical specialties. To support inexperienced medical MCQ writers and as an introduction to the field of MCQ writing, in this narrative literature review we describe:

- MCQ basic structure and terminology
- Written examples of flawed and acceptable MCQs
- The basic assessment of the quality and utility of MCQs; difficulty index and discrimination index
- A checklist of pragmatic considerations for writing good MCQs
- The importance of faculty training in writing MCQs
- The educational benefits of students framing MCQs
- Future perspectives - the role of artificial intelligence in generating MCQs

Materials and Methods

This review provides a narrative overview of the literature on the generation and use of MCQs in medical education and training. The research was carried out by searching electronic databases such as PubMed/Medline and Google Scholar for articles published between 1990 and 2023. Key terms used in the search for each database included medical MCQs, MCQs in medicine, medical MCQ review and creating medical MCQs. Furthermore, we screened the reference lists from pertinent articles to ensure a thorough search. This review excluded those works not written in English and published before or after the period specified in the inclusion criteria.

MCQs (Multiple-Choice Questions): Why We Use Them in Medicine

Assessing a large number of medical students by essay type questions is time consuming and is prone to subjective variation in the consistency and transparency of the evaluations. In contrast, evaluating a medical student by using multiple-choice questions (MCQs) is more objective and is more time efficient.^{13,14} Furthermore, good essay writing has a component that reflects the individual's writing ability, rather than their medical knowledge or clinical skills.

In addition, as the Bloom's classification indicates, multiple-choice tests can assess different levels of learning outcomes; beginning with basic recall to the clinical applications of basic sciences and their analyses and interpretations.^{13,15} This is possible because the student must choose the best answer for the given scenario from a set of potential answers which can vary in number from four to six. Additionally, the MCQ method of testing confers

reliability. It is well established that MCQs provide much less scope for guess work, making them a fairly accurate means of assessment.^{13,15} However, there are obvious limitations to this kind of assessment. One such limitation is *versatility* measurement which refers to the inability of the MCQ method to measure the standard to which the learner is able to organize his/her thinking processes, generate and creatively articulate explanations.¹⁶ While it is obviously important to have a clear understanding of the strengths and weaknesses of MCQs such as those we have described above, it is particularly important to always remain aware of the *validity* criterion in MCQ testing. Validity refers to the degree to which a test measures the learning outcomes it claims to assess.

Components of MCQs: Conventions in Question Organization

By convention, a multiple-choice question begins with a problem/statement, which is known as the *stem*. The stem has a list of suggested solutions which are called *options/alternatives*. The alternatives include one correct option which is the most suitable or *single best answer*. The rest of the alternatives are partially correct or incorrect, with respect to the statement provided in the stem and thus, are called *distractors*. The serial number or serial order of the correct answers is provided in the answer *key*. The *key*, whether hard copy or electronic, is understandably separately presented at the end of the set of MCQs. The stem is the most important part of understanding any MCQ that is to be answered. The stem may be in the form of a sentence or a number of sentences often comprising a clinical problem, in which case the MCQ may be designated as a PBQ (problem-based question¹⁷). Stems may also describe a requirement for simple factual recall which the question setter deems important. When the stem is composed of only a sentence, the sentence may be complete or incomplete.

Table 1 exemplifies these MCQ models:

Of course, question setters are always at their liberty to frame the stem in either way (i) or way (ii). However, a stem with an incomplete or partial sentence (although carrying the same meaning) should be less favoured. Because this requires the student to hold the partial sentence in working memory, sequentially completing the sentence with each option rather than just concentrating on the question's answer options - in this case, by recalling the laboratory diagnostic test for typhoid fever. In spite of this consideration, some question setters are fond of constructing the stem with an incomplete sentence. A variation on the incomplete sentence stem model is a stem that starts with a blank space that must be filled. For example:

“ _____ is the most reliable test for the diagnosis of typhoid fever during the first week of illness.”

Such a question, with a blank space at the very beginning of the stem, unnecessarily increases the cognitive load on the student and so, its use should be avoided. It could also be argued that the additional linguistic load placed on the student is inappropriate in a supposed test of scientific/medical knowledge - especially true in the worldwide domain of medical training where so many students use English as a second language.

At this point, it is noteworthy to examine the options given in questions (i) and (ii) of Table 1. Here the correct response has been highlighted in a bold font. This may be preferred by some examiners taking a manual approach to MCQ assessment generation. However, in the modern era, the software platform used to create or format the final examination may have special requirements, such as marking the correct answer with an asterisk.

Table 1 MCQ Examples with Complete and Incomplete Stem Sentences

(i) Stem comprising a complete sentence	(ii) Stem comprising an incomplete sentence
(i) Which of the following is the most sensitive investigation for the diagnosis of typhoid fever during the first week of illness? A. Stool culture B. Urine culture *C. Blood culture D. Widal test	(ii) The most sensitive laboratory investigation for the diagnosis of typhoid fever during the first week of illness is: A. Stool culture B. Urine culture *C. Blood culture D. Widal test

Notes: The asterisk (*) indicates the correct answer.

Attempts should be made to keep all the alternatives homogenous in substance, style and length, because heterogeneous content may, at times provide clues to the students who are relying on guesswork or may even mislead the student. This issue is made more evident through consideration of the following multiple-choice question:

Q1. The factors that are most important in determining the prognosis of a malignancy are:

- A. Stage and grade
- B. Differentiation and grade
- C. Stage and lymphatic invasion
- D. Differentiation and lymphatic invasion
- E. Differentiation and stage

In the above question, all the options offered have two components and are based on common properties of neoplasms. Hence, there is uniformity or homogeneity amongst the possible choices. Therefore, it will be necessary for the student to choose the right answer from the five apparently homogenous alternatives by applying his/her own good judgement. In this case, stage and grade represent the histopathological features that are generally assessed in order to determine the prognosis of a cancer - the correct answer is option A.

The above considerations hold true not only for a question having a stem with a complete or incomplete sentence, but also holds true for an MCQ with a stem based on a clinical problem. Question 2 is an example of such a PBQ that has a stem with a brief clinical history:

Q2. A 45-year-old farmer was an emergency admission to hospital with shortness of breath, a pounding sensation in his head and a tightness sensation in his chest. The farmer had been stung by bees on two previous occasions.

Which of the following statements does not hold true for his current condition?

- A. The clinical condition he is experiencing is anaphylaxis.
- B. Adrenalin and antihistamines should immediately be administered by injection.
- C. Leukotrienes and prostaglandins act as pharmacological mediators to give rise to such a clinical presentation.
- D. Degranulation of neutrophils leads to the secretion of the above-mentioned pharmacologic mediators.

The distractors given in the above question seem good and are appropriate. However, the situation would be less sanguine if an answer option like “degranulation of mast cells leads to the secretion of the above-mentioned pharmacologic mediators”, were added as shown:

Q3. A 45-year-old farmer was an emergency admission to hospital with shortness of breath, a pounding sensation in the head and tightness in the chest. He had a past history of being stung by bees on two earlier occasions. Which of the following statements does not hold true for his current illness?

- A. The clinical condition he is experiencing is anaphylaxis.
- B. Adrenalin and antihistamines should immediately be administered by injection.
- C. Degranulation of mast cells leads to the secretion of the above-mentioned pharmacologic mediators.
- D. Degranulation of neutrophils leads to the secretion of the above-mentioned pharmacologic mediators

This would not be categorized as a good PBQ for the following reasons:

1. Since (C) and (D) are two similar options, differing only in the type of the cell type being degranulated, there will be a natural tendency for the candidate to presume that the correct response is probably one of these two options.
2. Secondly, the proximity and similarity of these two options might easily trigger the student to recall that *mast cell* degranulation would be the right answer.

The next MCQ (PBQ) example shows another classic pitfall in the description of answer alternatives:

Q4. A 38-year-old man with end stage renal disease due to chronic glomerulonephritis was given a cadaveric kidney transplant. On the 3rd postoperative day, his urea and creatinine levels decreased appreciably. On the 7th postoperative day, the graft site became tender and swollen, with signs of inflammation. Which of the following should be adopted as part of the ideal management of this clinical case?

- A. Administration of intravenous third generation cephalosporins
- B. Estimation of serum IgG levels
- C. Renal cortical biopsy and intravenous administration of corticosteroids
- D. Oral corticosteroid administration alone

Option D, stating “oral corticosteroid administration alone”, does not appear to be a good answer alternative because the word “alone” creates an impression that this statement is not likely to be the correct answer. Thus, writing “oral corticosteroid administration” would be a better distractor than writing “oral corticosteroid administration alone.” The lesson here is that adding distractor options that have modifiers (“alone”) that seem absolute can decrease the effectiveness of the distractor.

The above considerations are intended to make the MCQs appropriate for the pertinent student population. Specifically, intending to strike a balance between clarity of writing and fair standard versus inherent complexity and assumed knowledge. Accordingly, a well-prepared student should be able to accurately judge and answer such polished MCQs. At the same time, it is very important for the question setter to be aware that the MCQs should not be so student-friendly that the framed question would be too easy for an average/below average student. Question 5 exemplifies the issue:

Q5. All of the following are antigen presenting cells except:

- A. B cells
- B. Macrophages
- C. Dendritic cells
- D. T cells

For an average student this question is too easy, because such a student should know that antigens are presented to T cells by other cells, but T cells themselves do not present antigens. The lacuna in the above question is that the distractors are not strong enough to challenge a typical student to critically judge and evaluate the answer options. Moreover, this MCQ is a negative question.

Including too many negative questions in a particular examination paper should be avoided; the necessarily increased linguistic complexity and consequent time consumption, decrease their value in actually testing knowledge or understanding of the medical material. Instead of using this imperfect Q5 MCQ, this question should be modified and increased in difficulty, in this case by considering factors related to the recognition of antigen by the effector cells of the immune system.

Q6. The killing of a virally infected cell is best achieved by which of the following effector cells of the immune system?

- A. T lymphocytes
- B. B lymphocytes
- C. Macrophages
- D. NK cells

Compared to Q5, Q6 seems is more challenging because the students must recall their understanding of the mechanism of antibody dependent cell mediated cytotoxicity (ADCC) and the Fc receptors present on Natural Killer (NK) cells, in order to identify the correct answer. Fc receptors are present on the macrophages too but macrophages are seldom involved in ADCC, as granzymes and other mediators for neutralizing a virally infected cell are not liberated by an activated macrophage, rather are elaborated by an activated NK cell. Hence, option C in Q6 is an example of a good distractor. Hence, to write such a good quality question, the question setter must have a clear and accurate understanding of the topic being assessed - or seek more experienced support. Whatever the final form and phrasing of the MCQs, it is essential to validate the questions in terms of their difficulty and discrimination levels. This evaluation is performed after an MCQ test has been conducted and the students’ responses have been obtained. These evaluations include the determination and assessment of the *difficulty index* and the *discrimination index*.

Difficulty Index

Before calculating this index, it is necessary to determine the scoring groups; which are estimated as follows:

Upper scoring group, lower scoring group

- Imagine arranging the total number of answer papers in order from the highest scores to the lowest scores.
- For example, if there are 75 answer sheets in one set, select 25 answers with high scores (upper group), 25 answers with low scores (lower group), and the remaining 25 as the middle group.

Now, the difficulty index is calculated using the formula $P=(H+L)\times 100$ divided by T; where H denotes the number of correct responses in the upper score group, L denotes the number of correct responses in the lower score group, and T, the total number of responses (both correct and incorrect) in all the groups. The value of P is usually expressed as percentage. The lower the percentage value, the more difficult is the question. Currently, the best quality examination software will automatically perform these described calculations.

Interpretation - A question with a difficulty index of 30–40% is difficult; below 30% being too difficult. An MCQ with a difficulty index of 50–60% is acceptable and one with 70% is too easy.^{18,19}

Discrimination Index

The calculation of the discrimination index is performed using the following formula:

$$d=2\times(H-L) \text{ divided by } T$$

Here d can be expressed as a decimal or fraction.

Interpretation - An index of 0.35 is excellent, ie this question can obviously discriminate well between a good and an average student. A discrimination index of 0.25 is good, 0.20 acceptable and less than 0.15 is designated as unacceptable, which should be revised.²⁰

In the above context, it is undoubtedly true that the construction of flawless MCQs²¹ is an art form in itself, at which some teachers inherently excel as part of their explicit evaluation of their students. Accordingly, at this point in the review, it worth evaluating the pros and cons of “faculty written MCQs” vs the recent trend towards “student framing and designing of MCQs.”

Faculty Written MCQs

Studies conducted recently highlighted that faculty members in a substantial number of medical institutions globally, were not adequately conversant with framing flawless MCQs for MBBS level examinations.^{17,18} Reports from India and other parts of the world^{22,23} stated that a few hours of conventional faculty training (as imparted by the Medical Council of India) was not adequate for the generation of the desired improvement in item writing flaws (IWFs). In contrast, Gupta et al¹¹ documented a statistically significant reduction in the frequency of occurrence of IWFs as a result of training the faculties in the writing of MCQs. According to their observations, only 15% of the MCQs were flawless prior to training, whereas flawlessness increased to a maximum of 27.7% after the workshop; the workshop being three hours in duration. The authors demonstrated that after the workshops, there was a statistically significant percentage increase in flawlessness, with a remarkable improvement in the quality of incorporated options/correct answers and distractors, which also reflected a better question homogeneity. In comparison with the 27.1% heterogeneity among the options prior to the workshop, a significant drop in heterogeneity to 5.6% was seen after the workshop training. It was therefore concluded that faculty training involving single/short discussion/training session(s) was not sufficient to correct all the flaws in writing MCQs. Rather, long-term training with a number of interactive workshop sessions should be a requirement for faculty MCQ training curricula.^{24,25}

Issues Related to the Student Framing of MCQs

Students becoming involved in the process of designing MCQs is a relatively new facet of the students’ teaching and learning activities. In a recent study conducted by Grainger et al²⁶ involving 4th year medical students learning anatomic pathology in the University of Otago, it became apparent that medical students were quite capable of creating cognitively challenging MCQs if supported with able guidance and direction. The majority (74%) of the MCQs thus framed, required the respondent to analyze and apply pathological knowledge in answering the questions. Thus, indicating that MCQ writing by senior medical students under proper guidance was quite a reasonable task, which could engage students in self-learning, self-evaluation and synthesis of information from a range of sources based on their theoretical knowledge. Other authors too, reported high quality student engagement in learning, when students were actively involved in the creative process of constructing MCQs as compared to passively answering MCQs.^{27,28} Accordingly, these studies indicated that a student generated MCQ approach should be introduced in medical institutions as part of active learning activities rather than as just online practice tools. Such a student centered MCQ generation process would actively

engage students, which would not only reinforce concepts learned in the seminar/lecture but could also develop a sense of responsibility for greater in-depth learning, including knowledge analysis and evaluation.

Overview and Practical Considerations in Developing MCQs

Undergraduate and graduate medical educators have a tendency to start *de novo* and accordingly, reinvent the wheel when first creating MCQs - this generally arises from a lack of realization of the number of helpful supportive resources²⁹⁻³¹ that already exist and are publicly available. The checklist proffered by Case and Swan³¹ is a particularly useful summary that is adapted and presented in Figure 1.

We have found the checklist(s) to be generally clear and straightforward in application. In relation to these guides, we would like to draw attention to a common error made by inexperienced writers - the issue of convergence that is exemplified in Table 2:

The large number of repeated occurrences of *ileum* and *stomach* in the answer options for the poorly prepared MCQ are difficult to ignore and would suggest to the alert student that the option containing both of these words would most likely be correct. A better formed version of this question would diminish or remove convergence entirely - or even

The Whole Question

- Has a **single best answer** format been used? (Whether single question, multi-question sets or extended matching questions)
- Does the question test the **application** of knowledge or the **analytical** use of knowledge?
- Does the question satisfy the informal rule that the correct **answer can be formulated** without seeing the answer options?
- Does the question have most of the text in the stem with answer options that are short?

Question stem and lead-in checklist

- Is the question structured as a **scenario**, describing a specific situation?
- Is the lead-in focused so that a **clear question** is being asked of the reader?
- Does the phrasing of the lead-in avoid making the described scenario irrelevant?
- Is the lead-in structured as a complete question with a **question mark**?
- Is the lead-in phrased **positively** rather than negatively? Avoiding "not" "except" "least"

MCQ answer options checklist

- Can the options be ranked on a single dimension from most correct to least correct?
- Are the answer options homogeneous in content and phrasing? So, that the right and wrong answers are not obvious by inspection.
- Are the answer options similar in length and parallel in structure? So, that the right and wrong answers are not obvious by inspection.
- Does each option follow grammatically and logically from the lead-in?
- Does the correct answer avoid repeating words used in the stem?
- Are the distractors phrased to avoid repetition that inadvertently clues the correct answer option (*convergence*)?
- Do the options avoid the use of absolute terms such as "always" and "never"?
- Do the options avoid the use of vague terms such as "rarely" and "usually"?
- Do the options exclude answers such as "all of the above" and "none of the above"?

Figure 1 MCQ Preparation Checklist.

Table 2 Poor and Good Convergence in MCQs

Poor question because of convergence in the answer options. Note the number of times “ileum” and “stomach” are included in the options. An uncertain student might then choose option D, the correct answer.	Better question; note the greater variation in answers that avoids multiple use of words that might trigger an uncertain student to the correct answer.
<p>Q7. Which two organs are most directly involved in the pathogenesis of pernicious anaemia?</p> <p>A. Duodenum and ileum</p> <p>B. Stomach and colon</p> <p>C. Pancreas and stomach</p> <p>*D. Ileum and stomach</p> <p>E. Pancreas and ileum</p>	<p>Q7. Which two organs are most directly involved in the pathogenesis of pernicious anaemia?</p> <p>A. Bone marrow and liver</p> <p>B. Duodenum and pancreas</p> <p>*C. Stomach and ileum</p> <p>D. Liver and pancreas</p> <p>E. Colon and caecum</p>

Notes: i) “Stomach” in red to emphasize its use in 3 of the 5 answer options.ii) “Ileum” in blue text to emphasize its use in 3 of 5 answer options.iii) “Liver” in red to emphasize its use in only 2 answer options. iv) “Pancreas” in blue to show its use in only 2 answer options. v) The asterisk (*) indicates the correct answer.

better, use convergence for misdirection to the wrong answer option (Table 2). As there is a rules-based approach to writing MCQs, it is inevitable that these rules would be either implicitly or explicitly incorporated into attempts to generate computer-based preparation of assessments; artificial intelligence (AI) based MCQ production. Although many of the studies^{32–34} were related to the use of AI software to answer medical MCQs, there have been a small number of attempts to use AI to create credible medical MCQs. The most pertinent being Agarwal’s study³⁵ which compared the effectiveness of ChatGPT, Google Bard and Microsoft Bing AIs to generate medical MCQs. All three were tasked to generate 110 multiple-choice questions in physiology, with appropriate competencies, for the MBBS (Bachelor of Medicine, Bachelor of Surgery) undergraduate exams. Although ChatGPT and Bard were able to generate 110 MCQs for the required competencies, Bing could only generate 100 MCQs for the necessary competencies. Furthermore, Bard and Bing generated significantly less difficult questions than ChatGPT. The researchers concluded that there were still limitations to the functionality of each AI at this early stage in their development. We believe that the rapid progression of AIs in learning how to produce MCQs will continue unabated and in the future will lead to good question generation. However, we suspect that human curation of MCQs³⁶ by local subject experts will still be necessary for the preparation of the final formative and summative assessments.

Conclusion

Besides the above-mentioned sections on the general theory and practice of MCQ writing, it is also important to maximize the fairness and reasonableness of MCQs, which is an increasingly important component of summative written assessments in undergraduate and postgraduate training. Poorly produced MCQs may be prone to validity issues; In order to improve validity, all MCQ assessments should be aligned with theoretical learning goals and with practical methods. The following practice points are always to be borne in mind for the purpose of accepting an MCQ as fair and equitable:

- (a) MCQs should be closely aligned with the target syllabus.
- (b) MCQs should be able to assess practical competencies, in addition to assessing theoretical knowledge and theoretical competencies.
- (c) MCQs should be free from constructive errors.
- (d) MCQs should be scrutinized to prevent question re-use or significant question repetition.

Our article is the most comprehensive peer-reviewed overview of the practical generation and use of MCQs in medicine, published in the last three decades. We have indicated and exemplified the common pitfalls in MCQ writing in medicine. We have identified tools and approaches that should be used in the generation of high quality multiple-choice questions. We have indicated objective criteria for assessing the quality of MCQs. We have critically discussed the role of AI in MCQ generation. We support the pragmatic use of MCQs in undergraduate and postgraduate medical training, with the *caveat* that training and support are made available to the creators in order to generate the highest possible quality questions.

Abbreviations

MCQ, Multiple-choice Question; PBQ, Problem-Based Question; MBBS, Bachelor of Medicine Bachelor of Surgery; NFD, Nonfunctioning Distractors; AI, Artificial Intelligence; NK, Natural Killer cell; ADCC, Antibody Dependent Cell Mediated Cytotoxicity; IWFs, Item Writing Flaws.

Data Sharing Statement

All data and materials are available upon request to the corresponding author.

Acknowledgments

The authors thank Ajman University for its support.

Author Contributions

All authors made a significant contribution to the conception, study design, execution, acquisition of articles, analysis and/or interpretation; participated in drafting, revising or critical reviewing of the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and have agreed to be accountable for all aspects of the work.

Funding

The authors did not receive additional funding for this study.

Disclosure

The authors declare that they have no competing interests in this work.

References

1. Ali K, Zahra D. Ten tips for effective use and quality assurance of multiple-choice questions in knowledge-based assessments. *Eur J Dent Educ.* 2024;28(2):655–662. doi:10.1111/eje.12992
2. Owolabi LF, Adamu B, Taura MG, et al. Impact of a longitudinal faculty development program on the quality of multiple-choice question item writing in medical education. *Ann Afr Med.* 2021;20(1):46–51. doi:10.4103/aam.aam_14_20
3. Abdulghani HM, Ahmad F, Irshad M, et al. Faculty development programs improve the quality of Multiple-choice questions items' writing. *Sci Rep.* 2015;5(1):9556. doi:10.1038/srep09556
4. Gupta P, Meena P, Khan AM, Malhotra RK, Singh T. Effect of faculty training on quality of multiple-choice questions. *Int J Appl Basic Med Res.* 2020;10(3):210–214. doi:10.4103/ijabmr.IJABMR_30_20
5. DiSantis DJ, Ayoob AR, Williams LE. Journal club: prevalence of flawed multiple-choice questions in continuing medical education activities of major radiology journals. *AJR.* 2015;204(4):698–702. doi:10.2214/AJR.13.11963
6. Bhat SK, Prasad KHL. Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology: a cross-sectional study. *Indian J Ophthalmol.* 2021;69(2):343–346. doi:10.4103/ijo.IJO_1610_20
7. Stagnaro-Green AS, Downing SM. Use of flawed multiple-choice items by the New England. *J Med Medical Educ.* 2006;28(6):566–568. doi:10.1080/01421590600711153
8. Dellenges MA, Curtis DA. Will a short training session improve multiple-choice item-writing quality by dental school faculty? a pilot study. *J Dent Educ.* 2017;81(8):948–955. doi:10.21815/JDE.017.047
9. Coughlin P, Featherstone C. How to write a high quality Multiple-Choice Question (MCQ): a guide for clinicians. *Eur J Vasc Endovasc Surg.* 2017;54(5):654–658. doi:10.1016/j.ejvs.2017.07.012
10. Pugh D, De Champlain A, Touchie C. Plus Ça Change, plus C'Est pareil: making a continued case for the use of MCQs in medical education. *Med Teach.* 2018;41(5):569–577. doi:10.1080/0142159X.2018.1505035
11. Gupta V, Williams E, Wadhwa R. Multiple-choice tests: a–Z in best writing practices. *Psychiatr Clin North Am.* 2021;44(2):249–261. doi:10.1016/j.psc.2021.03.008
12. Goyal M, Agarwal M, Goel A. Interactive learning: online audience response system and multiple-choice questions improve student participation in lectures. *Cureus.* 2023;15(7):e42527. doi:10.7759/cureus.42527
13. Cox KR, Ewan CE. *The Medical Teacher.* London: Churchill Livingstone; 1982. Chapter 37. Writing multiple-choice questions. Bandarnaike RC, Cox KR. Chapter 38. How to construct a fair multiple-choice question paper? Cox KR. Chapter 39. What can teachers and students get out of multiple-choice questions. Cox KR.
14. Mee J, Pandian R, Wolczynski J, et al. An experimental comparison of multiple-choice and short-answer questions on a high-stakes test for medical students. *Adv Health Sci Educ.* 2023;29(3):783–801. doi:10.1007/s10459-023-10266-3
15. Cox KR. How did you guess? *Med J.* 1976;1:884–886.
16. Picketting G. Against multiple-choice questions. *Med Teach.* 1979;1(2):84–86. PMID:24479932. doi:10.3109/01421597909019397
17. Joorabchi B. How to: construct problem-solving MCQs. *Med Teach.* 1981;3(1):9–13. doi:10.3109/01421598109081736

18. Ebel RL. *Essentials of Educational Management*. Englewood Cliffs; New Jersey: Prentice-Hall Inc; 1972.
19. Hubbart JP. *Measuring Medical Education*. Philadelphia: Lea and Febiger; 1981.
20. Anderson J. For multiple-choice questions. *Med Teach*. 1979;1(1):37–42. PMID: 24483175. doi:10.3109/01421597909014340
21. Boland RJ, Lester NA, Williams E. Writing multiple-choice questions. *Acad Psychiatry*. 2010;34(4):310–316. doi:10.1176/appi.ap.34.4.310
22. Vyas R, Supe A. Multiple-choice questions: a literature review on the optimal number of options. *Natl Med J India*. 2008;21:130–133.
23. Singh T, de Grave W, Ganjiwale J, Supe A, Burdick WP, van der Vleuten C. Impact of a fellowship program for faculty development on the self-efficacy beliefs of health professions teachers: a longitudinal study. *Med Teach*. 2013;35(5):359–364. doi:10.3109/0142159X.2013.769672
24. Khafagy G, Ahmed M, Saad N. Stepping up of MCQs' quality through a multi-stage reviewing process. *Educ Prim Care*. 2016;27(4):299–303. doi:10.1080/14739879.2016.1194363
25. Tenzin K, Dorji T, Tenzin T. Construction of multiple-choice questions before and after an educational intervention. *JNMA J Nepal Med Assoc*. 2017;56(205):112–116. PMID: 28598446. doi:10.31729/jnma.2976
26. Grainger R, Dai W, Osborne E, Kenwright D. Medical students create multiple-choice questions for learning in pathology education: a pilot study. *BMC Med Educ*. 2018;18(1). doi:10.1186/s12909-018-1312-1
27. 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
28. Singh L. Technology enhanced peer learning with PeerWise: experiences and perceptions from a developing country. *The Caribbean Teaching Scholar*. 2014;4(1):5–22.
29. McCoubrie P. Improving the fairness of multiple-choice questions: a literature review. *Med Teach*. 2004;26(8):709–712. doi:10.1080/01421590400013495
30. Naidoo M. The pearls and pitfalls of setting high-quality multiple-choice questions for clinical medicine. *South Afr Family Pract*. 2004;65(1):e1–e4. PMID: 37265132; PMCID: PMC10244925. doi:10.4102/safp.v65i1.5726
31. Case SM, Swanson DB. *Constructing Written Test Questions for the Basic and Clinical Sciences*. Philadelphia, PA: National Board of Medical Examiners; 2016.
32. Fowler T, Pullen S, Birkett L. Performance of ChatGPT and bard on the official part 1 FRCOphth practice questions. *Br J Ophthalmol*. 2023;324091. PMID: 37932006. doi:10.1136/bjo-2023-324091
33. Meo SA, Al-Khlaiwi T, AbuKhalaf AA, Meo AS, Klonoff DC. The scientific knowledge of bard and ChatGPT in endocrinology, diabetes, and diabetes technology: multiple-choice questions examination-based performance. *J Diabetes Sci Technol*. 2023;19322968231203987. doi:10.1177/19322968231203987
34. Meo SA, Al-Masri AA, Alotaibi M, Meo MZS, Meo MOS. ChatGPT knowledge evaluation in basic and clinical medical sciences: multiple-choice question examination-based performance. *Healthcare*. 2023;11(14):2046. PMID: 37510487; PMCID: PMC10379728. doi:10.3390/healthcare11142046
35. Agarwal M, Sharma P, Goswami A. Analysing the applicability of ChatGPT, Bard, and Bing to generate reasoning-based multiple-choice questions in medical physiology. *Cureus*. 2023;15(6):e40977. PMID: 37519497; PMCID: PMC10372539. doi:10.7759/cureus.40977
36. Gajjar S, Sharma R, Kumar P, Rana M. Item and test analysis to identify quality Multiple-choice Questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. *Indian J Community Med*. 2014;39(1):17–20. PMID: 24696535; PMCID: PMC3968575. doi:10.4103/0970-0218.126347

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