Acta Scientific Dental Sciences

Volume 10 Issue 8 August 2026

Review Article

Best Practices in Designing Multiple Choice Questions – A Review

Chandrasekharan Nair K[1]*, Pradeep C Dathan[2], T Mohan Kumar[3] and D Bheemalingeswara Rao[4]

[1]Professor Emeritus, Department of Prosthodontics, Sri Sankara Dental College, Akathumuri, Thiruvananthapuram, Kerala, India

[2]Professor and Head of the Department of Prosthodontics, Sri Sankara Dental College, Akathumuri, Thiruvananthapuram, Kerala, India

[3]Director, Centre for Temporomandibular Disorders, Kunnukuzhi, Trivandrum, Kerala, India

[4]Professor and Head of the Department of Prosthodontics, Vishnu Dental College, Bhimavaram, AP, India

***Corresponding Author:** Chandrasekharan Nair K, Professor Emeritus, Department of Prosthodontics, Sri Sankara Dental College, Akathumuri, Thiruvananthapuram, Kerala, India.

Scopus Id: https://orcid.org/0000-0003-3114-3015

DOI: 10.31080/ASDS.2026.10.2138

Received: July 03, 2026

Published: August 03, 2026

Abstract

Multiple-choice questions (MCQs) are among the most widely used assessment tools in health professions education because of their objectivity, reliability, ease of scoring, and ability to sample a broad range of knowledge and cognitive skills. This review provides an overview of the historical evolution of MCQs, the different formats currently in use, and the principles underlying effective item construction. Common item-writing flaws and methods of item analysis used to evaluate question quality are discussed. The advantages and limitations of MCQs are examined, highlighting their strengths in providing standardized, objective, and reliable assessment, while acknowledging concerns related to cueing effects, guessing, and overemphasis on factual recall. Recent developments, including computerized item banks, computer-adaptive testing, and artificial intelligence-assisted item generation, are also reviewed. Although these emerging technologies are expected to enhance the efficiency and quality of assessment, careful adherence to established principles of test construction and validation remains essential to maintain the validity, reliability, and educational value of MCQs. With continued technological advances, MCQs are expected to remain a cornerstone of knowledge assessment and clinical reasoning in medical and dental education.

Keywords: Multiple-Choice Questions; Assessment; Medical Education; Dental Education; Item Analysis; Artificial Intelligence

Introduction

Assessment is an integral component of medical and dental education. Its primary purpose is not only to measure students' knowledge and competence but also to promote learning and guide educational improvement. Assessment enables learners to identify their strengths and weaknesses and respond appropriately to their individual learning needs [1].

Assessments are broadly classified as formative and summative. Formative assessment provides continuous feedback on student performance and offers guidance for improvement, thereby promoting learning. In contrast, summative assessment is primarily intended for grading, certification, and determining whether the required competencies have been achieved. An effective assessment system should also discriminate accurately between different levels of student performance. Multiple-choice questions (MCQs) are valuable in both formative and summative assessments. In classroom settings, they are particularly useful for providing periodic formative feedback, whereas in professional examinations they serve as a reliable tool for summative assessment [2].

A multiple-choice question consists of four essential components: the stem, options, the key, and distractors. The stem presents a question, statement, or clinical problem. It is followed by a set of alternative responses (options), from which the examinee selects the most appropriate answer. The correct response is known as the key, while the remaining alternatives are called distractors, which should be plausible enough to challenge students who have not mastered the content. The most commonly used format is the single best answer (single best response) question, in which only one option is correct. When more than one option is correct, the item is referred to as a multiple-response question. Table 1 illustrates the structural components of typical MCQs [3] (Table 1).

Select the single best answer Example - 1 #A single bilateral edentulous area located anterior to the remaining natural teeth is classified as Kennedy-Apple-gate- (Stem) 1. Class I (options) (Distractor) 2. Class II (options) (Distractor) 3. Class III (options) (Distractor) 4. Class IV (options) (Key)
Example - 2 #If the clasp assembly is not being placed on a tooth adjacent to an edentulous area, the minor connector must be positioned (Stem) 1. in the embrasure between two teeth (options) (Key) 2. on the lingual surface of the tooth distal to the edentulous space (options) (Distractor) 3. on the lingual surface of the tooth mesial to the edentulous space (options) (Distractor) 4. on any tooth with a convex lingual surface (options) (Distractor)

Table 1: Multiple Choice Question (MCQ).

History

The origins of objective testing can be traced to the early twentieth century through the pioneering work of Frederick J. Kelly, an educational psychologist at Columbia University. In 1914, while completing his doctoral research, Kelly developed the Kansas Silent Reading Test, one of the earliest standardized objective tests. The examination required students to select the correct word or phrase from several alternatives to complete a sentence, allowing reading comprehension and speed to be assessed objectively without relying on handwritten essays. The test represented a major shift from subjective essay-based evaluation to standardized, objective assessment.

The need for efficient large-scale assessment became even more apparent during the First World War, when the rapid recruitment of military personnel demanded reliable methods for evaluating large numbers of candidates. Adaptations of Kelly's objective testing principles were incorporated into the U.S. Army Alpha and Beta intelligence tests, enabling millions of recruits to be assessed quickly and consistently.

A major technological advancement occurred during the mid-twentieth century with the introduction of optical mark recognition (OMR) technology. Early OMR systems, developed during the 1930s and subsequently refined, enabled answer sheets containing specially designed bubbles to be scored. Modern OMR scanners detect the presence or absence of pencil or ink marks using reflected light, compare responses with a stored answer key, and generate scores rapidly with exceptional accuracy. Automated scoring greatly reduced examiner workload, minimized human error, and made large-scale standardized testing feasible.

Another important milestone was the publication of Bloom's Taxonomy in 1956, which encouraged educators to design assessment items that measured higher-order cognitive skills such as application, analysis, synthesis, and evaluation rather than simple factual recall [4,5]. Consequently, MCQs evolved from testing memory alone to assessing clinical reasoning and problem-solving abilities. The national board of medical examiners' item writing guide is the gold standard for developing MCQs in health sciences (Table 2). It emphasizes creating one-best-answer questions based on clinical vignettes to effectively test diagnostic and problem-solving skills rather than rote memorization.

The Clinical Vignette (Stem) Focus on a problem: The stem should present a clinical problem featuring a patient, setting, and manifestation. Avoid clues: Omit unnecessary specific physical findings or lab values that give away the diagnosis. Positive framing: Frame the question positively (e.g., "What is the most likely diagnosis?"). Avoid negative phrasing like "All of the following are true EXCEPT". Option to answer without looking: The lead-in question should be focused enough that an expert could determine the answer before reading the options.
The Options Four choices: Use exactly four answer options (one correct answer and three distractors) to optimize reliability. Avoid complex combinations: Never use "All of the above," "None of the above," or "Both A and B". Homogeneity: Ensure all options are mutually exclusive, logically ordered (e.g., alphabetical or numerical), and grammatically parallel. Plausible distractors: Distractors should be realistic errors or common misconceptions for the learners.
General Rules One objective: Assess only one core educational objective per item. Clear language: Avoid vague frequency terms like "often," "rarely," or "usually," which can lead to ambiguity. Avoid verbatim text: Do not copy phrases directly from textbooks; questions should require application.

Table 2: The National Board of Medical Examiners' Item Writing Guide.

During the latter half of the twentieth century, advances in psychometrics and educational measurement further strengthened the scientific basis of MCQ assessment. Item analysis techniques, including the calculation of difficulty index, discrimination index, and distractor efficiency, enabled educators to evaluate the quality and performance of individual questions objectively. As medical and dental education expanded worldwide, MCQs became the preferred assessment method because of their objectivity, reliability, and ability to evaluate large cohorts efficiently [6,7].

The beginning of the twenty-first century witnessed the emergence of computer-based testing and computer-adaptive testing (CAT), in which the difficulty of successive questions is adjusted according to the candidate's previous responses. Modern digital assessment platforms now support multimedia content, interactive graphics, radiographic images, videos, and complex clinical scenarios, enabling more authentic assessment of diagnostic reasoning and clinical decision-making. Recent developments in computerized item banking and artificial intelligence-assisted item generation have further enhanced the efficiency of question development, quality assurance, and psychometric evaluation. Consequently, contemporary MCQs have evolved into sophisticated assessment tools capable of measuring not only factual knowledge but also higher-order cognitive and clinical competencies.

Different types of MCQs currently used

MCQs have become a popularly used assessment tool in professional examinations and in health sciences education because of their ability to evaluate a broad range of cognitive capabilities. Also, they are known for objectivity and reliability. Traditional MCQs were of recall type but now it is adapted or expanded to assess higher order thinking, decision making and clinical reasoning. Major types of MCQs are described below.

1. Single best answer (SBA) or Single best option (SBO)

The characteristics of SBA are briefly mentioned in the introduction. More than one option may appear plausible (seemingly reasonable, believable, or likely to be true). Only one option can be considered as the best response. Level of knowledge and application can be evaluated by this type of MCQs. Getting plausible alternatives is not an easy job. 'All of the above' or 'None of the above' if included in the list of options, the candidate need check only two options and if he knows that both are correct or wrong, all of the above and none of the above selection becomes easier. If the key is longer than the distractors due to grammatical requirements, it will serve as a clue to the candidate.

Example

Stem: Which vitamin deficiency causes scurvy?

A. Vitamin A

B. Vitamin C

C. Vitamin B12

D. Vitamin D

Key: B. Vitamin C

2. Multiple completion type (Type K)

Multiple completion (Type K) MCQs feature a question stem followed by four or more options, of which more than one can be correct. Test-takers evaluate each statement individually and selects the matching combination (e.g., I and II are correct). This requires higher levels of cognition. This is an upgraded version of the single best response type of MCQ.

Example

Stem: Which of the clinical features are typically associated with Acute Pseudomembranous Candidiasis (Oral Thrush)?

I. Painful, white, curd-like plaques on the oral mucosa.

II. Plaques that can be easily wiped off, leaving an erythematous base.

III. Firm, white lesions that cannot be scraped off.

IV. A predisposition in immunocompromised patients or those on broad-spectrum antibiotics.

Options

A. I, II, and IV only

B. II and III only

C. I and III only

D. I, II, III, and IV

Key: A (I, II, and IV only)

3. Multiple response or multiple correct answer questions (MCAQs)

These questions contain more than one correct option. Students are required to select all correct responses to obtain full marks. Such items can assess comprehensive understanding and reduce the probability of guessing.

Example

Stem: Which of the following dental materials chemically bond to the structure of the tooth?

A. Glass Ionomer Cement (GIC)

B. Zinc Phosphate

C. Resin-Modified Glass Ionomer (RMGI)

D. Dentin bonding agents (used with composite resin)

E. Zinc Oxide Eugenol (ZOE)

Key: A, C, D

4. Extended matching questions (EMQs)

EMQs have a highly effective format, mainly intended for problem-solving. They consist of a primary theme, a list of 8-15

possible options (which act as the answers), and a series of 2-5 clinical scenarios. Each scenario should be matched with the most accurate option.

Example

Theme: Oral Medicine and Pathology

Options:

A. Ameloblastoma

B. Candidiasis (Pseudomembranous)

C. Erythroplakia

D. Herpes Simplex Virus (Primary)

E. Lichen Planus (Reticular)

F. Leukoplakia

G. Median Rhomboid Glossitis

H. Pemphigus Vulgaris

I. Squamous Cell Carcinoma (SCC)

Scenarios

A 60-year-old male patient presents with an asymptomatic, raised white patch on the lateral border of his tongue. He is a heavy smoker. The lesion cannot be scraped off, and a biopsy confirms dysplasia.

Key: F (Leukoplakia)

A 25-year-old immunocompromised patient presents with white, curd-like plaques on the buccal mucosa and palate. These plaques can be easily wiped away, leaving behind an erythematous and sometimes bleeding base.

Key: B (Candidiasis - Pseudomembranous)

A 40-year-old female presents with bilateral, lacy, white, spider-web-like striations (Wickham's striae) on her buccal mucosa. She has no complaints of pain, but states she noticed them a few months ago.

Key: E (Lichen Planus - Reticular)

A 65-year-old male presents with a non-healing, indurated ulcer with rolled borders on the floor of the mouth. He smokes and consumes alcohol daily.

Key: I (Squamous Cell Carcinoma)

5. Assertion-Reason type question

This is a two-part multiple-choice format designed to test deep conceptual understanding rather than simple memorization. It consists of a declarative statement (Assertion) and an explanatory statement (Reason). The candidate is given four options to choose from.

Example

Assertion (A): Maxillary first premolars frequently present with two root canals, despite often having a single root at the cervical level.

Reason (R): The pulp chamber of a maxillary first premolar is wide mesiodistally and constricted buccolingually, frequently bifurcating into buccal and lingual canals.

a. Both Assertion (A) and Reason (R) are true, and Reason (R) is the correct explanation for Assertion (A).

b. Both Assertion (A) and Reason (R) are true, but Reason (R) is not the correct explanation for Assertion (A).

c. Assertion (A) is true, but Reason (R) is false.

d. Assertion (A) is false, but Reason (R) is true.

Key: A

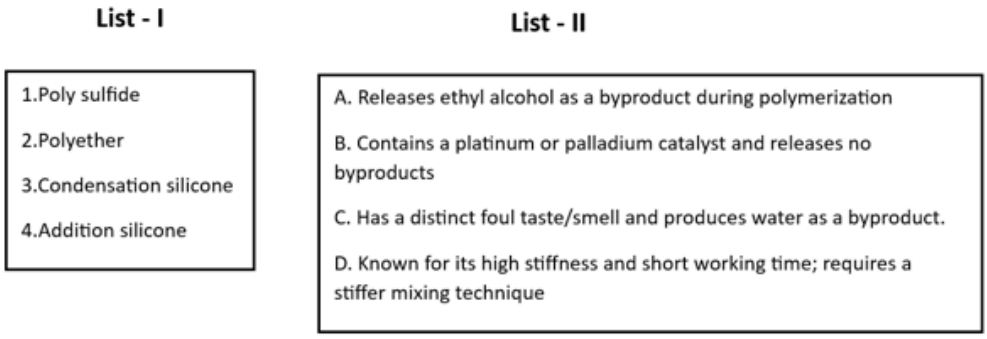
Explanation: The maxillary first premolar is well-known for its anatomical variance; it typically has two root canals (one buccal and one lingual) even when the root itself is not completely divided, largely dictated by its crown and pulp chamber morphology.

6. Matching type MCQs

This type is used to test the ability to associate concepts. Two parallel lists are given – one a column of premises and the other a column of responses. Candidates are expected to pair the related elements by recognizing relationships, patterns or logical correlations.

Example:

Question: Match the elastomeric impression material (List I) with the property or chemical component (List II).



Key: 1-C, 2-D, 3-A, 4-B

It is desirable to have a greater number of choices in the list II so that there will not be a chance to do guess work. In this example if three are matched correctly, the fourth one will become automatically correct.

7. Multiple true-false MCQ

In this particular format, a stem is followed by four or five statements and the candidates are asked to determine whether each statement is 'true' or 'false'. This type of MCQ allows broad content coverage and probably may promote a little guesswork [8].

Example

About tooth anatomy and enamel

A. Enamel is the hardest substance in the human body.

B. Enamel contains living cells and can repair itself after it is fully formed.

C. Dentin is softer than enamel but harder than cementum.

D. The periodontal ligament holds the root of the tooth to the alveolar bone.

Key

A. True: Enamel is highly mineralized (composed mostly of hydroxyapatite) and forms the hardest structure in the body.

B. False: Enamel does not contain living cells (ameloblasts are lost after eruption) and cannot naturally repair itself once damaged.

C. True: Enamel is the hardest, followed by dentin, and then cementum.

D. False: The periodontal ligament attaches the cementum to the alveolar bone [8-10].

8. Scenario-based multiple-choice question (SBMCQ)

Stem is replaced with a clinical vignette. Vignette is a brief snapshot or a short descriptive scene. Instead of telling a whole story with a beginning, middle, and end, it simply captures a specific moment, mood, or feeling. In tests of health science, the vignette describes a clinical presentation (e.g., patient details, symptoms, and radiographic findings), requiring the candidate to synthesize information and derive a diagnostic reasoning or make a clinical decision. The candidate is expected to methodically analyse the patient's symptoms, history, and test results to deduce a diagnosis, identify the root cause, or determine a treatment plan and apply knowledge rather than just use rote memorization [11].

Example: Dentistry

Clinical Vignette (Stem): A 45-year-old female patient presents to the dental clinic with chief complaint of dull, throbbing pain and swelling of the right posterior mandible with a duration of three days. History of poorly controlled type 2 diabetes. Intraoral

examination reveals a severely carious mandibular right first molar (molars) with associated buccal soft tissue swelling which is tender on palpation. A periapical radiograph shows a diffuse, ill-defined radiolucency around the apex of the tooth. Vitality testing yields no response.

Based on the clinical and radiographic findings, which of the following is the most appropriate primary intervention?

options:

A. Perform endodontic access and debridement under local anaesthesia

B. Prescribe systemic antibiotics only and defer treatment for one week

C. Administer local anaesthesia, perform incision and drainage of the fluctuant swelling, and establish endodontic access

D. Extract the tooth immediately without prior drainage

Key (Correct Answer): C. Administer local anesthesia, perform incision and drainage of the fluctuant swelling, and establish endodontic access.

Rationale

For an acute localized fluctuant swelling (abscess) of endodontic origin, the primary treatment principle is to establish drainage (e.g., via incision and drainage and endodontic access) to relieve pain and pressure. Treating the fluctuant swelling surgically takes precedence over root canal instrumentation alone, and antibiotics are secondary to drainage, especially in a medically compromised diabetic patient.

Key benefits of this format

- **Real-World Application:** Tests a clinician's diagnostic and management skills exactly as they happen in practice.
- **Minimizes Guessing:** Requires the student to identify the underlying problem before looking at the choices.
- **Higher-Order Thinking:** Upgrades the assessment from recall to analysis and evaluation [11,12].

Number of items to be included in MCQ examination

An exam will have validity if sufficient number of items are included and it will also ensure breadth of sampling. There is no universal agreement on the number of items to be included. 60-100 items if included, is considered as optimum for a 60 – 90 minutes examination. The scores obtained in an examination of 25-30 items may not truthfully indicate the depth of knowledge of the candidate and hence it cannot be considered as a reliable method of evaluation [10,13].

Time allotted for each type of MCQs

The time allotted for the candidate should be sufficient to attempt all the MCQ items in the question paper. One fact that is usually overlooked is that the purpose of the examination is to test the knowledge of the candidate and not the speed of response. Comprehension of the item may also depend on the language used for framing the item. Complicated structuring may hinder the student's understanding and framing the item in simple language is an important factor which we cannot overlook. Different formats of MCQs require different duration for the candidates to complete the answering and a reasonable duration is given below:

- Single best answer (40 seconds)
- Multiple completion type (50 seconds)
- Multiple Response (50 seconds)
- Extended Matching Questions (60 seconds)
- Assertion-Reason type question (60 seconds)
- Matching type (60 seconds)
- Multiple true-false (60 seconds)
- Scenario-based multiple-choice question (60-90 seconds)

Observational studies have to be conducted periodically to ascertain the duration with different groups of students and based on that the question paper setters can design the number and type of MCQs [10].

Principles of formulating MCQs

Designing an examination that makes use of MCQs is a challenging and time-consuming task. Teachers should formulate a few questions every week especially after taking a lecture on a particular topic. Teacher will be fresh with the subject and the questions can be used in class tests.

Students should be instructed to select the best answer rather than the correct answer. Every distractor may have an element of correctness and if the 'best answer' is asked for, the possibility of arguments can be avoided.

The terminology used in the question should have been used during the course and in the course materials. Unfamiliar terms that appear in the distractor section may be considered as incorrect by the students. If the words and usages that appear in the stem and the key are similar, it may act as a clue and hence to be avoided. Intentionally misleading the student to a distractor which may appear as the key should not be attempted.

MCQs with negative wording like NOT, EXCEPT or FALSE in the stem should be avoided if possible. Emphasize the negative word by making them bold, CAPITALIZING or underlining.

Example:

Which of the following is NOT a primary colour?

A. Red

B. Yellow

C. Green

D. Blue

Key: C

Stem should contain the full problem or question. Student should be able to answer the question without referring to the options. This will serve as a test for checking the clarity of the question. All the relevant content should be included in the stem. Same usage need not be repeated in all the options. That usage should be made part of the stem.

In presently used MCQs of all India entrance examinations for medicine and dentistry have four options like A,B,C,D. Three to five alternatives were tried and it has been observed that three choice items are as effective as four or five choice items. The key should be unarguably correct and the distractors should be incorrect. Make all the options grammatically consistent with the stem. If it is relevant, the choices should be arranged in an order – numerical, chronological or conceptual. The key for each question should be positioned differently with equal as well as random distribution.

Options like 'all of the above' and 'none of the above' should not be used because it provides a chance for guess work. Now there are only two options left. Of the two, if the student identifies that one is incorrect, the 'all of the above' option becomes irrelevant. What is being tested with an option like 'none of the above'? At the end he knows that the distractors are incorrect [14].

Advantages and disadvantages of MCQ

Advantages

- The major quality of an MCQ is its ability to test objectively. The answer to every question is predetermined and it is either completely right or wrong. There is no scope for personal interpretation or examiner bias.
- Candidates are expected to choose the correct answer from the given list and need not formulate a lengthy answer which is time consuming.
- Examiners can include a vast portion of the curriculum in the evaluation and the time required is very short.
- MCQs can be adapted to test different levels of cognition like recall, comprehension or application; in other words, it is flexible.
- It can be used to conduct periodic tests in the form of quiz and it can be repeated any number of times.
- It is a good tool to conduct formative evaluation and gives feed back to both teachers and students.
- Question bank can be generated and each question can be subjected to critical evaluation and modification.

Disadvantages

- Formulating a question with a clear stem and equally plausible distractors is not an easy job and is time consuming. Because of this reason, teachers tend to make MCQs which are simple and that explores recall level information.
- Provides scope for guess work and with limited number of questions, depth of knowledge cannot be assessed.
- There is no provision to give credit to the student who has better knowledge on the subject asked. MCQ provides opportunity to select the correct answer from the list provided.

- May expose students to misinformation that influence subsequent thinking about the content [15,16].

Item analysis

Item analysis is the process of analysing the performance of an MCQ after it has been used in a test. This exercise will determine (i) whether the item is of appropriate level of difficulty (ii) whether the item is capable of discriminating between well informed and ill informed students (iii) to provide a feed back on the functionality of the distractors. By conducting item analysis, well performing questions can be identified, flawed items can be detected and teachers can obtain a feed back on the effectiveness of the instruction.

Difficulty index (Facility value)

Difficulty index indicates the proportion of students who answered an item correctly. It will be expressed in percentage. The results are interpreted as follows:

- Easy item: > 70%
- Moderate difficulty: 30 – 70%
- Difficult item: < 30%

Discrimination index

The students are ranked according to the scores obtained in the test. Then they will be divided into high achiever group (H) and low achiever group (L). If group consists of 50 students, each group will have 25 students each. If the group consists of 100 students or more, the top one third of the students will be included in the H group and the lowest one third will be included in the L group. The index will be calculated by the following formula.

$$d = \frac{H - L}{N} \times 2$$

H – number of students answering the item correctly in the high achiever group

L – number of students answering the item correctly in the low achiever group

N – total number of students in the batch

Discrimination index (di) is a measure of the ability of the item to discriminate between good students and not so good students. While difficulty index is expressed in percentage, discrimination index is indicated as a fraction. The maximum value obtained for discrimination index is 1.0 which indicates that the item has ideal quality.

- Recommended for inclusion: >0.25
- Acceptable with revision: 0.15 – 0.25
- Non inclusion of the item : <0.15

Distractor effectiveness

Any of the distractors in an item that has not attracted the attention of 5% of the candidates, is considered as non functional.

Items with difficulty index between 50 – 60% are ideal but between 30 – 70% are acceptable. Items with discrimination index between 0.25 to 0.35 are good. Lower than this is considered as a poor item. Item analysis is usually applied in the summative evaluation. Item analysis provides a critical insight into the MCQ on its ability to evaluate the candidates objectively [10,17,18].

Flaws in the preparation of MCQs

MCQs make objective assessment of knowledge and hence they have become popular with medical and dental education. If utmost care is not exercised, flaws may creep in the sectors of validity, reliability and objectivity. Commonly observed flaws are as follows.

Flaws related to the construction of stem

- **Ambiguous stems:** Stem should present a clear problem. Unnecessary information, ambiguous wording or vague expressions can confuse the candidates and which will result in misinterpretation. By reading the stem, the question raised should be understandable without reading the options.
- **Negative wording:** Stems containing negative words such as 'Except' or 'Not', very often lead to confusion and increase the possibility of inadvertent errors. In unavoidable situations, when such words are used, they should be highlighted.
- **Testing for only recall of facts:** Preparing MCQs to test higher order cognitive functions is a tough job and many of the items fall into the recall of facts level. Test for clinical reasoning and decision making abilities get slowly avoided from the question papers. This is considered as the limitation of the MCQ design.

Flaws related to options and distractors

- **Implausible distractors:** Distractors should be credible and attract the less knowledgeable students. Obviously incorrect or absurd alternatives decrease the discriminating power of the question and increase the probability of guessing the correct answer.
- **Unequal option lengths:** If the correct answer is framed longer and detailed than the distractors, students may identify the correct answer based on its appearance rather than the knowledge about it.
- **Grammatical and logical cues:** Grammatical inconsistencies between the stem and options may unintentionally reveal the correct answer. Similarly, options that differ in tense, number, or syntax may provide clues independent of content knowledge.
- **Overlapping in the content of options:** Options should be independent and mutually exclusive. Overlapping options or those containing similar numerical ranges may confuse students and compromise the accuracy of responses.
- **Heterogeneous alternatives:** All alternatives should belong to the same category and match with the stem. Mixing diagnoses, treatments, anatomical structures, and laboratory findings within the same item can make the question illogical and may provide with clues.
- **Use of "All of the above" and "None of the above" in the options:** These options create major issues with validity and testing accuracy. They encourage superficial guessing and fail to test actual subject knowledge.

Flaws affecting validity and reliability

- **Tricky or misleading questions:** Items designed intentionally to confuse students do not assess the competence. MCQs should challenge understanding, not the test taking skills of examinees.
- **Irrelevant difficulty:** Difficulty should arise from the content and not from the poor wording or structure.
- **Lack of blue printing:** If learning objectives are not pre-determined, trivial objectives get over emphasised and important competencies get neglected. A test blueprint ensures balanced content coverage and curriculum outcomes.
- **Failure to perform item analysis:** Failure to evaluate item analysis after examinations may allow flawed questions to persist. Regular item analysis helps improve the quality and reliability of MCQ banks [19,20].
- Flawed MCQ items are common in assessments but they do not cause disadvantage to borderline students. Conversely, high-achieving students are more likely to be penalised by flawed items.

Comparison of MCQs created by teachers, students, computerised banks and AI

Multiple-choice questions (MCQs) may be generated by teachers, students, computerized question banks, or artificial intelligence (AI). Each source has distinct strengths and limitations with respect to content validity, cognitive level, reliability, adaptability, and resource requirements (Table 3).

Feature	Teacher-generated MCQs	Student-generated MCQs	Computerized Question Banks	AI-generated MCQs
Content expertise	High; based on subject knowledge and curriculum objectives	Variable; depends on students' understanding	High if developed by experts	Variable; depends on quality of training data and prompts
Alignment with learning objectives	Excellent	Moderate to good	Good	Good with appropriate prompts and review
Cognitive level addressed	Can cover all levels of Bloom's taxonomy	Often lower levels initially; can improve with training	Usually lower to moderate levels, though higher-order items may be included	Can generate items across Bloom's taxonomy with suitable instructions

Quality of distractors	Good if item-writing principles are followed	Often weak or implausible	Generally standardized and validated	Variable; may generate plausible but occasionally incorrect distractors
Consistency and standardization	Variable among teachers	Variable	High	High, though output quality varies
Speed of generation	Time-consuming	Time-consuming	Immediate retrieval	Extremely fast
Need for expert review	Desirable	Essential	Periodic updates required	Essential before use
Potential for errors	Human bias and ambiguity	Conceptual inaccuracies	Outdated items	Factual errors and inaccuracies
Adaptability and customization	Excellent	Very limited	Limited	Excellent
Educational benefit to creator	Enhances teacher reflection	Promotes active learning	Minimal	Minimal unless used interactively
Security and originality	Good	Variable	Risk of repeated exposure	Generates novel questions but resembles training examples
Best used	Formative and summative assessment	Learning exercises and peer teaching	Large scale examinations	Rapid item generation
Pedagogical quality	Excellent	Poor curriculum perspective	No upgradation	Good
Draw back	Highly time intensive	Factual error, poor format	Static standards unless reviewed	Nonfunctional distractors, requires review

Table 3: Comparison of MCQs created by Teachers, Students, Computerized Question Banks and Artificial Intelligence.

Teacher made MCQs remain the gold standard because they are framed within the curricular frame work and with specific objectives. Preparation of a flawless MCQ is a time-consuming process and requires training and experience. Teachers can frame questions that are capable of assessing higher order cognition.

Palmer and Devitt found that well-constructed Multiple-Choice Questions (MCQs) are often superior to Modified Essay Questions (MEQs) for testing higher-order cognitive skills. Students are engaged in the process of making MCQs as part of active learning, critical thinking and self assessment. These MCQs require faculty review followed by item analysis. Computerised question banks provide standardised, validated, items which are accessible very easily and fit for large scale testing adopted for professional course admissions in medicine and dentistry. AI systems can rapidly produce a large number of customised questions, modify question difficulty and generate explanations. AI generated items should be subjected to fact checking for inaccuracies and misleading distractors. ChatGPT performed at or near the passing threshold of 60% accuracy. Being the first to achieve this benchmark, this marks a notable milestone in AI maturation. Impressively, ChatGPT was able to achieve this result without specialized input from human trainers. Furthermore, ChatGPT displayed comprehensible reasoning and valid clinical insights, lending increased confidence to trust and explainability. However, it is safer to suggest that AI functions best as an adjunct rather than a replacement for human expertise [21-23].

Discussion

Until the 1950s, long- and short-answer essay questions dominated assessment in medical and dental education. Examinations with essay-type questions remain valuable for assessing higher-order cognitive abilities, including critical thinking, synthesis of knowledge, problem-solving, and written communication. However, they are associated with several limitations, particularly the time required for evaluation and the

subjectivity inherent in scoring, both of which reduce reliability. The introduction of multiple-choice questions (MCQs) addressed many of these shortcomings by providing an objective, standardized, and reliable method of assessment. The subsequent transition from paper-based examinations to computer-based testing during the 1990s further enhanced the efficiency of MCQs through automated administration and instant scoring [24].

Early MCQs primarily assessed factual recall and simple comprehension. However, advances in educational theory and assessment methodology have led to the development of clinically oriented formats capable of evaluating higher-order cognitive processes, including application of knowledge, clinical reasoning, decision-making, and problem-solving [10,13]. Consequently, modern MCQs have evolved from simple memory tests into versatile instruments capable of assessing multiple levels of cognition.

An assessment blueprint should be prepared before constructing examination items. Blueprinting is a systematic process that maps assessment items to curricular objectives, learning outcomes, and competencies, ensuring comprehensive and balanced coverage of the curriculum. It is an essential component of assessment planning because it enhances content validity and ensures that the examination reflects the intended educational objectives. Blueprinting also prevents inclusion of questions outside the prescribed syllabus and facilitates an appropriate balance between factual knowledge and higher-order cognitive skills [25].

During blueprint development, curricular content is categorised according to educational importance; commonly as must know, desirable to know, and nice to know. In many institutions, approximately 60% of questions assess must-know competencies, 30% assess desirable-to-know areas, and the remaining 10% cover nice-to-know content. Question setters should possess adequate teaching experience and a thorough understanding of both the curriculum and assessment principles to ensure the preparation of high-quality examination items [26].

Medical and dental students generally express positive perceptions toward MCQs because of their objectivity, standardized format, ease of completion, and rapid feedback. Clinical scenario-based MCQs are particularly appreciated because they emphasize reasoning and decision-making rather than rote memorization. Compared with traditional oral examinations, MCQs are also associated with lower levels of examination anxiety. Furthermore, when incorporated into self-assessment programs, MCQs provide immediate feedback, help learners identify deficiencies in knowledge, and promote self-directed learning [27].

Despite their numerous advantages, MCQs have certain inherent limitations. They demonstrate excellent content validity, high reliability, and permit efficient assessment of large numbers of candidates. Consequently, they have become the preferred assessment tool in undergraduate education, postgraduate entrance examinations, and professional licensure tests worldwide. Nevertheless, one of the most frequently cited criticisms is the cueing effect, whereby the presence of answer options facilitates recognition of the correct response rather than independent recall of knowledge. Students may successfully identify the correct diagnosis, antibiotic or dental material from a list of alternatives but may be unable to generate the same answer independently in a real clinical situation. Excessive dependence on MCQs may therefore encourage superficial learning strategies, factual memorization, and test-oriented preparation rather than deep understanding and clinical reasoning. For this reason, MCQs should complement rather than replace other assessment methods that evaluate communication skills, critical thinking, clinical performance, and written expression [28]. Some of the experienced dental college teachers have observed that over dependence on MCQs has blunted the essay writing skills of the students and which may have a handicapping effect on their career while publishing scientific articles.

Recent advances in artificial intelligence (AI) have introduced new possibilities for the use of MCQs. Programmatic assessment is a recent innovation. It is an educational framework where every assessment is treated as a continuous data point rather than an isolated pass/fail event. Within this system, Multiple-Choice Questions (MCQs) are efficient tools for testing foundational knowledge and clinical reasoning, functioning as the backbone for learning and assessment cycles. When used as low-stakes formative tests, MCQs identify gaps in real-time, helping learners self-reflect and adjust their study trajectories. MCQs used as part of programmatic assessment serve as a powerful study aid, reinforcing memory retention and deep conceptual understanding.

AI-assisted question generation offers substantial reduction in the time required to create high-quality question banks and enables

large-scale production of standardized examination items. Studies have demonstrated considerable savings in faculty time when AI is integrated into a supervised workflow. Large language models can generate grammatically sound questions with appropriate formatting and psychometric characteristics comparable to those of human-written items. However, AI-generated questions often assess lower cognitive levels, produce fewer effective distractors, and occasionally contain factual inaccuracies or ambiguous wording. Therefore, expert human review remains indispensable to ensure clinical accuracy, educational relevance, and psychometric quality [29].

Generative AI is increasingly being incorporated into educational assessment to automate the development of question banks and create authentic clinical scenarios. Contemporary assessments are gradually shifting from testing factual recall toward evaluating clinical reasoning by presenting realistic patient problems and requiring learners to justify their decisions. Modern computer-based assessment platforms also support multimedia content, interactive graphics, radiographic images, videos, virtual reality, and real-time simulations, thereby enabling assessment of practical decision-making in addition to knowledge acquisition. Although these innovations offer exciting opportunities, issues related to examination security, data privacy, and academic integrity remain important challenges that require continued attention.

Despite the remarkable evolution of MCQs, no single assessment method is capable of evaluating all domains of professional competence. Written essays, structured short-answer questions, objective structured clinical examinations, workplace-based assessments, and MCQs each assess different aspects of learning. The future of assessment in medical and dental education is therefore likely to depend on a balanced, programmatic approach that combines the strengths of multiple assessment methods. Such a hybrid model would preserve the efficiency and objectivity of MCQs while ensuring that creativity, communication skills, clinical reasoning, and independent thinking are assessed adequately.

Conclusions

Multiple-choice questions remain one of the most widely used and versatile methods for assessing knowledge, comprehension, application, and higher-order cognitive skills in health professions education. Their effectiveness depends on adherence to established principles of item writing, systematic blueprinting, rigorous item analysis, and periodic revision to maintain validity, reliability, and fairness.

Teacher-generated MCQs continue to represent the benchmark because they are closely aligned with curricular objectives and clinical practice. Student-generated MCQs promote active learning, critical thinking, and deeper engagement with course content. Computerized item banks improve standardization and efficiency, while recent advances in artificial intelligence have introduced unprecedented opportunities for rapid, scalable, and customized item generation. Nevertheless, AI-generated questions should always undergo expert review to ensure factual accuracy, educational relevance, and psychometric quality.

The future of MCQ-based assessment lies not in replacing educators with technology but in integrating human expertise with digital innovations and artificial intelligence. Such a balanced approach will facilitate the development of high-quality, learner-centered, and psychometrically robust assessment systems capable of meeting the evolving needs of medical and dental education.

Conflict of Interest

The authors declare that they do not have any conflict of interest.

CRediT Author Contributions

Conceptualization- K. Chandrasekharan Nair, Review of articles- Pradeep Dathan; T. Mohan kumar, Initial draft preparation: Pradeep Dathan, Bheemalingeswara Rao, Review and editing- K. Chandrasekharan Nair, Supervision-K. Chandrasekharan Nair.

All the authors have read and agreed to the published version of the manuscript.

Acknowledgements

The authors acknowledge the use of an artificial intelligence language model developed by OpenAI, for assistance in editing and improving the readability of certain portions of the manuscript.

Bibliography

1. Auewarakul C., *et al*. "Sources of validity evidence for an internal medicine student evaluation system, an evaluative study of assessment methods". *Medical Education* 39 (2005): 276-283.

2. Case S and Swanson D. "Extended matching items: a practical alternative to free response questions". *Teaching and Learning in Medicine* 69 (1994): S4-S6.

3. https://neetmds.com/trial/index.php?subject=Prosthodontics

4. K Chandrasekharan Nair., *et al.* "Role of objectives in Dental Education - A Review". *Acta Scientific Dental Sciences* 6.2 (2022): 88-94.

5. Bloom BS., *et al*. "Taxonomy of Educational Objectives: Handbook I: Cognitive Domain". New York: David McKay; 1956.

6. J Phelan. "Classroom Assessment Tasks and Tests (chapter)". *International Encyclopedia of Education (Third Edition)* (2010).

7. "The National Board of Medical Examiners (NBME) item-writing guide".

8. Tarrant M and Ware J. "Guidelines for the construction of multiple choice questions tests". *Nurse Education Today* 32.6 (2012): e1-e5.

9. O'Sullivan B., *et al*. "Assessment of global health education: The role of multiple-choice questions". *Frontiers in Public Health* 9 (2021): 640204.

10. Ananthakrishnan Sethuraman., *et al*. "Medical education principles and practice". 2nd Ed., 2000, NTTC, JIPMER, Pondicherry, India

11. Chauhan A., *et al*. "Designing effective scenario-based multiple-choice questions for health professionals: Tips and techniques". *National Medical Journal of India* 37 (2024): 335-339.

12. Badyal DK., *et al*. "Triple Cs of scenario-based multiple-choice question: Concept, construction, and corroboration". *National Journal of Pharmacology and Therapeutics* 1 (2023): 8-12.

13. K Chandrasekharan Nair. Chapter 8, Principles of assessment, Dental Teacher, Springer Nature Switzerland (2025).

14. https://uwaterloo.ca/centre-for-teaching-excellence/catalogs/tip-sheets/designing-multiple-choice-questions

15. https://evidencebased.education/resource/multiple-choice-questions-pros-and-cons/

16. https://www.facultyfocus.com/articles/educational-assessment/advantages-and-disadvantages-of-different-types-of-test-questions/

17. Tejinder Singh., *et al*. "Principles of medical education". 4th Ed, 2013, J P Brothers Medical Publishers, New Delhi.

18. Downing SM. "Twelve steps for effective test development". In: Downing SM, Yudkowsky R, editors. Assessment in Health Professions Education. New York: Routledge; (2009): 3-25.

19. Tarrant M and Ware J. "Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments". *Medical Education* 42.2 (2008): 198-206.

20. Haladyna TM. "Developing and validating multiple-choice test items". 3rd ed. Mahwah (NJ): Lawrence Erlbaum Associates; (2004).

21. Palmer EJ and Devitt PG. "Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper". *BMC Medical Education* 7 (2007): 49.

22. Kung TH., *et al*. "Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models". *PLoS Digital Health* 2.2 (2023): e0000198.

23. Sallam M. "ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns". *Healthcare (Basel)* 11.6 (2023): 887.

24. Mohammed O Al-Rukban. "Guidelines for the construction of multiple choice questions tests". *Journal of Family Community Medicine* 13 (2006): 125-133.

25. Patil SY., *et al*. "Blueprinting in assessment: a tool to increase the validity of undergraduate written examinations in pathology". *International Journal of Applied and Basic Medical Research* 5 (2015): S76–79.

26. Ismail MA-A., *et al*. "Seven steps to construct an assessment blueprint: a practical guide". *Education in Medicine Journal* 12.1 (2020): 71-80.

27. Rajeshwar Reddy Kasarla., *et al*. "Perception towards Online MCQ Test for MBBS and BDS Students' as an Assessment Method during the Covid-19 Pandemic". *Journal of Universal College* (2022).

28. https://psmsurat.com/what-is-programmatic-assessment/

29. Law AK., *et al*. "AI versus human-generated multiple-choice questions for medical education: a cohort study in a high-stakes examination". *BMC Medical Education* 25 (2025): 208.