Review article:

Multiple Choice Questions in Medical Education: How to Construct High Quality Questions

Abdus Salam[1]. Rabeya Yousuf[2], Sheikh Muhammad Abu Bakar[3]

Abstract:

Multiple choice questions (MCQ) are the most widely used objective test items. Students often learn what we assess, and not what we teach, although teaching and assessment are the two sides of the same coin. So, assessment in medical education is very important to ensure that qualified competent doctors are being produced.A good test is the test that assesses higher level of thinking skills. Many inhouse MCQs are found faulty which assess lower level of thinking skills. The main problems in constructing good MCQs are that (i) very few faculty members have formal training in questions construction, (ii) most of the questions are prepared in the last minutes where little time exist for vetting to review the quality of questions and (iii) lack of promise on the standard of the question format and underestimation of the use of blueprint in medical schools. Constructing good MCQs, emphasis should be given that, the stem is meaningful and present a definite problem, it contains only relevant material and avoid negativity. It should be ensuring that, all options present as plausible, clear and concise, mutually exclusive, logical in order, free from clues and avoid 'all of the above' and 'none of the above'. The MCQs can tests well any higher level of the cognitive domain, if it is constructed well. Efforts must be made to prepare and use of test blueprint as a guide to construct good MCQs. This paper describes and offers medical teachers a window to a comprehensive understanding of different types and aspects of MCQs and how to construct test blueprint and good MCQs that tests higher order thinking skills in the future medical graduates, thereby ensures competent doctors are being produced.

Keywords: MCQs, construction, high quality, assessment, higher order thinking skills, test-blueprint.

International Journal of Human and Health Sciences Vol. 04 No. 02 April'20 Page : 79-88
DOI: http://dx.doi.org/10.31344/ijhhs.v4i2.180

Introduction:

Assessment in medical education is very important to ensure qualified competent doctors to be produced. However, no single method is appropriate for assessment of skills, knowledge and attitudes needed in medicine, so a combination of assessment techniques is always required[1]. Among the various test methods, multiple choice questions (MCQs) are widely used to assess learning and knowledge. These are the most common type of written test item used in medical education.

Multiple-choice question: Multiple choice questions are most familiar as the widely applicable and valuable type of objective test items[2,3], introduced into medical examinations since 1950[1]. Such a test usually consists of a number of items that pose a question to which students must select an answer from among a number of choices[4]. An appropriately constructed

1. Abdus Salam, Medical Education Unit, Faculty of Medicine, Widad University College, Kuantan, Malaysia.
2. Rabeya Yousuf, Blood Bank Unit, Department of Medical Lab Services, Faculty of Medicine, Hospital University Kebangsaan Malaysia, Cheras, Malaysia.
3. Sheikh Muhammad Abu Bakar, Department of Anatomy, Faculty of Medicine International Medical College & Hospital,Gazipur, Bangladesh.

Correspondence to: Abdus Salam MBBS (Ctg), MPH (Dhaka), DipMedEd (Dundee), MMedEd (Dundee). Head of Medical Education Unit and Manager -Academic Scheduling & Curriculum, Faculty of Medicine, Widad University College, Kuantan, Malaysia, ORCID iD: 0000-0002-2629-4903
Phone: 06 0199061724(H/p), E mail: abdussalam.dr@gmail.com

MCQs result in objective testing that can measure knowledge, comprehension, application, and analysis. The entire test question is called the item, which consists of a stem and several options. The stem is the question, statement, or lead-in to the possible answers. The possible answers are called options, alternatives, or choices. The correct option is called the keyed response. The incorrect options are called distractors or foils[5]. A good distracter needs to be inferior to the correct answer but should be able to be plausible to a non-competent candidate.

There are many advantages as well as limitations of using the MCQs. It has an excellent reliability which is the reproducibility or consistency of a test score and can be efficiently employed to a large number of students for stratifying individual performance[5]. Their scoring is easier to understand than short-answer test scoring because there is no need to resolve partial and misspelled answers. Furthermore, MCQ test assessment is more objective than the assessment including essay questions. The essay test scores can be affected by the examiner's inconsistencies and are not immune to the influence of bluffing and writing ability factors, which can lower their reliability. The usage of MCQ tests is very important for the examiners because they allow easy and quick evaluation. These tests are particularly essential for the examiners who cover multiple courses with large number of enrolments[6]. MCQs can provide a large number of examination items that encompass many content areas, so a broad domain can be covered. They are time efficient as can be administered in a relatively short period and graded by a computer and have a high reliability[7]. It can also assess taxonomically higher-order cognitive processing such as interpretation, application and evaluation rather than the test of recall of isolated facts. Thus, they could test a number of skills in addition to the recall of factual knowledge, and are reliable, reproducible and cost-effective[1]. However, it is difficult to construct the question with little or no experience or training and examiners need to spend considerable time and effort for this purpose[1]. The MCQ allow for guessing without knowing the actual fact. Also, it does not provide a measure of writing ability and the scores can be influenced by reading ability[8]. It does not assess communication skill which is a very important part of medical education. It also does not assess writing skills. Therefore, this is not an effective way to test students' ability to organize their thoughts or express their creative ideas[6]. It allows students to perform excellently well with repeated questions and there are chances of errors in collation if answers (to 100s of questions) are manually marked[9]. Construction of MCQ is most difficult and time consuming even for the experienced and trained personnel. There is evidence that the basic item-writing principles are not followed mostly while constructing MCQs. Flawed MCQs interfere with accurate and meaningful interpretation of test scores and negatively affect student pass rates. Therefore, to develop reliable and valid tests, items must be constructed that are free of such flaws.

Types of MCQ

MCQ can be presented in various format, commonly used format are 'one correct answer', 'single or one best answer (SBA/OBA)', 'true or false' and 'multiple true or false', 'matching' and the 'extended matching questions or items (EMQ/EMI)' and 'case or problem' format or type[4]. The most popularly used one is the SBA/OBA type, also called A-type MCQ which asks the students to choose the best answer from 4-5 options[10,11]and the second more popular is the EMQ/EMI, also known as R-type MCQ[10].

i) One correct answer format consists of two parts: (i) the stem, which identifies the question or problem and (ii) the responses alternatives. Students are asked to select the one alternative that best completes the statement or answer the question. For example: Which one of the following antibodies is involved in Type 1 hypersensitivity? (A) IgA, (B) IgD, *(C) IgE, (D) IgG, (E) IgM.

ii) Single or one best-option MCQs are the most versatile and widely used MCQs. This format refers to a list of options that can all be correct in the sense that each has an advantage, but one of them is the best. The example is shown in Table 5.

iii) True-false questions contain statements that the student marks as being either true or false. The purpose of this question is to test whether students are able to evaluate the correctness of an assumption[12]. All parts of the statement have to be absolutely true or false in order to qualify as true or false. It is difficult to construct flawlessly and in general, true-false questions check knowledge of facts for which they are best avoided[4,12]. An example of true-false question is: Anabolic Steroids provide

an unfair advantage in sports where strength plays a large part. *(A) True (B) False[13].

iv) Multiple true / false questions consist of a stem which can be a statement, question, case history or clinical data followed by five true or false options; candidate has to answer separately to each of the five options as true or false[9,14]. Multiple true false questions reveal the limits of the SBA format in that the student must think and evaluate each possible answer, rather than selecting the one best answer[15,16]. It is easy to change wording delicately, so that a true statement becomes false or vice versa[16]. It can cover breadth of knowledge (widely sample) though may not assess depth of knowledge. In a paper comprising 60 questions, 300 different items of knowledge can be tested. For each true answer student will get 1 mark, for each false answer get -0.5 mark, there will be no mark for options not attended[3]. Time required to answer each question is about 70 seconds. There needs carefulness while formulating, if not it may test only recall. Terminologies such as never, none, always, all, impossible tend to be false while sometimes, often, usually, generally tend to be true, as such both should be avoided. The example is given in Table 6.

v) Matching type: In this format, students need to match the two lists of statements or words, with specific instructions. The two lists should contain different number of items to avoid cueing[9]. The advanced version is extended matching questions or items (EMQ/EMI) type.

vi) Extended matching questions/items (EMQ/EMI) consists of a number of options and a series of questions where the answers are needed to be selected or matched from the options. A typical set of EMQ/EMI begins with an option list of four to 26 options; eight to ten options are usually used, followed by two or more patient-based items requiring the examinee to indicate a clinical decision for each item[1]. But they are not easy to prepare. Generally EMQ/EMI comprises of four parts (a) a **theme**, which include a symptom, investigation, diagnosis or treatment, e.g. back pain, dyspnoea, diabetes, corticosteroids, (b) an "extended" list (A-H) of possible answer **options** (c) the question, also called the **lead-in** statement, and (d) a clinical problem / scenario or vignette also called the **stem.** There needs more than one, at least two clinical problems / scenario or vignette for each theme. The candidate is asked to match one or more options to each item stem[1]. Extended matching items are most commonly used in medicine as they can be used to test diagnostic ability and clinical judgment[1]. The example is shown in Table 8.

vii) Case or problem format can use a real or realistic resource such as documents or images to use to answer the question. This type of questions can effectively test the application of knowledge.

Guidelines for Writing MCQs: It is important to use good grammar, punctuation, and spelling consistently and minimize the time required to read each item. The ideal question will be answered by 60-65% of the tested population. In a well-constructed MCQ, unintended cues should be avoided such as making the correct answer longer in length than the distractors. Instruction for answering the MCQ could be common for a set of questions. For example, for the single best response type, the instruction could be: "Select the most appropriate answer and darken the corresponding circle in the answer sheet provided"[14]. A good distracter needs to be inferior to the correct answer but should be plausible or as nearer to the correct answer as possible and yet maintaining the correctness of the right answer[11]. The number of options in MCQ varies from three to five. In case of three options, the chances of guessing are as high as 33%, in case of five options one or two distractors tend to be non-functional. It is suitable to use four or five alternatives depending upon the availability of good distractors[11].

Writing objectives : In order to prepare an MCQ test, it is important to consider the sampling of objectives and the content areas to be tested. It is necessary to choose an important learning outcome before constructing an MCQ. Therefore, the item in an MCQ test should be addressed to test a significant learning outcome[11]. The MCQs drawn from a representative sample of content areas of predetermined learning outcomes will allow for a high degree of test validity[5]. Learning objectives are statements of intent or desire which we expect to achieve by our learners at the end of any educational program. Learning objectives should be written as much as possible to SMART, an acronym for specific, measurable, attainable,

realistic and time-bound[17,18].

Benjamin Bloom :

Benjamin Bloom, an educational psychologist divided what and how we learn into three separate domains of learning: Cognitive domain–related to thinking/knowledge (K), Affective domain–related to feeling/attitudes (A), psychomotor domain–related to doing/skills or practice (P). In 1956, Bloom published a taxonomy of cognitive learning, which was described as a hierarchy of (i) knowledge, (ii) comprehension, (iii) application, (iv) analysis, (v) synthesis, and (vi) evaluation[19].In 2001, 45 years later, the last two taxonomies i.e., synthesis and evaluation revised as (v) evaluation and (vi) creation[20]. Buck Walters' Taxonomy describes learning at three level as follows: Level I:Recall of information, Level II: Comprehension and application and Level III: Problem solving ability -involving analysis, evaluation and creation[21]. The Table-1 below shows the revised Blooms' cognitive taxonomy that suitable to test different levels of learning by MCQs[20].

Table-1: Suitability of testing revised Blooms' cognitive taxonomic level learning by MCQs.

Revised Blooms' Cognitive Taxonomy			
Creating	Level-VI	Higher order learning	Higher level, not suitable for MCQs
Evaluating	Level-V		Higher order learning that can be tested with MCQs
Analyzing	Level-1V		
Applying	Level-1II		
Understanding	Level-II	Lower order learning	Lower order learning is what most MCQs target
Remembering	Level-1		

MCQs written to test knowledge (lower-level learning) would not be appropriate to test competence for objectives that reflect analysis (higher-level learning). For example, an MCQ that asks the learner to recognize benign dermal calcifications on a mammogram does not test the learners' problem-solving ability. A question that provides specific patient information and imaging data (i.e., a patient vignette) and that asks the learner to choose the most appropriate management is an example of an item that tests problem solving ability. Objectives should define important knowledge or skills and should be supported by the instruction provided through the educational program. Observable, measurable objectives allow for accurate assessment of whether the learner has achieved the objectives. MCQ should test at the same level of learning as the objective it is designed to assess.

There are many threats to the validity of any test which can be categorized into (i) Content under-representation(CU), meaning to under sampling of the content/domain and (ii) Construct Irrelevance variance (CIV) which refers to the use of inappropriate adjustments (flawed items, ill-structured items) which interfere with the ability to meaningfully interpret scores or ratings[10].

The content underrepresentation can be avoided by the use of test blueprint[10].A test blueprint is also calledtest specification table which is a document that reflects the assessment. It contains (i) content /objectives to be tested (ii) the questions that designs to test the content/ objective (iii) learning domain and levels therein testing. Blueprint is a three-dimensional chart which shows the placement of each question in respect of the objective and the content areathat it tests[22,23].The purpose of making blueprint is to provides a solid foundation for all test development activity thatoffersan evidence for the content related validity; it is a strategic plan in order to ensure that the assessment is more meaningful[10,23,24]. The Table-2 shows an example of test blueprint for better understanding.

How towrite the Stem : The basic rule of writing stem is that students should be able to understand the question without reading it several times and without having to read all the options[14]. The stem is the question or statement, or lead-in to the possible answers which is usually composed first and is best written as a complete sentence or question with all the relevant information and avoid repetition in the options. Any maps, diagrams, graphs, or images can also be incorporated in the stem which should be accompanied by a complete statement. Stems should be written in a simple language so that the stem is not a test of the examinees' reading ability. It should not be tricky or misleading, such that they might deceive the examinee into answering incorrectly[5].Lead-in part of the stem should be construct in such a way that it builds on the information in the stem and pose a clear question or a statement[25,26]. The stem should be stated so that only one of the options can be authenticated and that option should be indisputably correct. When more than one option has some element of truth or accuracy but the keyed response is the best, the stem should ask the student to select the best answer rather than the correct answer[5]. As a general guide, students can complete between one

Tables-2: Example of a test blueprint or test specification table

Test Blueprint or Test Specification Table														
Objectives / Contents /Topics to be tested	Level of Blooms' Cognitive Domain												Total question	
	Questions Format measuring recall / comprehension				Questions Format measuring application/ analysis				Questions Format measuring evaluation / creation				n	% test devoted to topic
	SBA	MTF	MEQ	SAQ	SBA	MTF	MEQ	SAQ	SBA	MTF	MEQ	SAQ		
Topic-1	1	1	0	2	2	0	2	2	1	0	0	0	11	18
Topic-2	1	2	1	1	2	2	2	1	1	0	0	2	15	25
Topic-3	2	0	1	2	1	1	1	3	1	1	0	0	13	22
Topic-4	2	1	1	2	2	3	1	3	2	2	1	1	21	35
Total question	6	4	3	7	7	6	6	9	5	3	1	3	60	100
% test devoted toeach level	10	6	5	12	12	10	10	15	8	5	2	5	100	

SBA=Single best answer; MTF=Multiple true false; MEQ=Modified essay question; SAQ=Short answer question

and two MCQs per minute[5].

Central idea should include in stem instead of the choices. Any excessive, unnecessary and irrelevant statements in the stem should be avoided. Language should be clear and straightforward; complex or imprecise wording may become a test of reading comprehension rather than an assessment of the subject matter. Imprecise terms such as seldom, rarely, occasionally, sometimes, few, and many are not uniformly understood and should be avoided[5]. Word the stem positively and asked a direct question (?) or present a definite statement using a colon (:) or a semicolon (;). Avoid negative words such as not or except; as negative questions tend to be less effective and more difficult for the examinee to understand. If negative stem in some instances cannot be avoided, ensure that the word appears capitalized such as NOT or EXCEPT, e.g. Which one of the following is NOT mechanism of VSD?

To develop a good a clinical problem /scenario or vignette, it should include all or part of the following parts[10].

(i) General information about a patient, such as: age, gender, socio economic status, education level. etc. (e.g., A 40 year old male, low socio-economic class).

(ii) Site of care, for example: hospital, health center, department, clinic (e.g., comes to the outpatient clinic).

(iii) Presenting complaint: (e.g., complaining of chest pain).

(iv) Duration of symptoms: (e.g., that has continued for 3 days).

(v) Patient health history, such as: family history, chronic illness, previous attacks, current treatments.

(vi) Physical findings: examination findings, as results of diagnostic studies, as initial treatment, subsequent findings, etc.

The first three items are vital in the vignette and the other depends on the scenario and the question to be asked.

How to write the Options : The options are the possible answers; they are also called as alternatives or choices or foils. Three to four option items are effective while more than five options is burdensome and often leads to faulty options while increasing the reading demands of the student. All the options need to keep homogeneous in content and of similar consistent length[5]. Options should be in logical or numerical order; avoid repetition in the options; length of options about to be equal; only one of options be the right answer and all are plausible. Options should be independent, not be overlapping[10]. Avoid clues to the right option such as always, never, completely, absolutely, may, could, can and resembling words in the stem, as test wise examinees will know that almost anything is possible. Avoid obvious correct as well as irrational options. Avoid all of the above and none of the above. Using 'all of the above'as correct option, students are exposed to a lot of correct information, but answers may be

more obvious /noticeable, hence learning comes from recall processing; the examinee only need to have partial information in order to answer the question. If they know that only two of the options are correct in a four or five choices question, they can determine that all of the above is the correct answer option. Equally, students only need to eliminate one answer option as implausible in order to eliminate 'all of the above' as an answer option. Using 'none-of-the-above' as correct, students are exposed a lot of incorrect information and may not need to retrieve correct information. Both question types can be harmful to accurate assessment and possible benefits to learning are small[27]. Avoid grammatical inconsistencies that cue the test taker. Options should start with capitalizing the first alphabet responding to a direct question and should start with capitalizing or uncapitalizing of first alphabet depending on whether the statement ends with colon or semi colon. First, key answer of the question needs to be identified followed by distractors assuring that they are plausible. Distractors should directly relate to the lead-in question or statement. Applyingappropriate distractor is a challenging task for the examiners and needs lots of efforts in construction.

Some examples of MCQs with different types or formats, context in the stem, style, options, technical flaws and improved versions are shown here:

Table-3: Example of SBA type of MCQ with context-free and context-rich stem.

Table-3.	
Context-free stem	**Context-rich stem /improved version**
What is the most important mechanism of Heart Failure in achild of one-year old[10]? A. Pressure overload B. Ventricular dilatation C. Ventricular dysfunction D. Ventricular weakness E. Volume overload	A 1yearold baby who is a known case of VSD, was admitted to the ER by his mother, complaining of cough, shortness of breath which disturbs his feeding for the last 3 days. On examination the baby was found ill, his temperature was 38.7C, his pulse was 110/min, which is regular, dyspneic with respiratory rate of 70/min and intercostals recessions, and has a liver was 4 cm below the costal margin[10]. What is the most likely mechanism of his symptoms? A. Pressure overload B. Ventricular dilatation C. Ventricular dysfunction D. Ventricular weakness E. Volume overload
Comment: Context-free version asked totally recall of information, while the context-rich one is stimulating to think and it resembles the real-life situation which the students will face in the future.	

Table-4: Example of SBA type of MCQ assessing higher order thinking level-V "Evaluation".

Table-4.
A 78-year-old man presents to the Emergency Department with sudden onset of left sided chest pain radiating to his back. He has history of stable angina and peripheral vascular disease. His blood pressure is 80/50 mm of Hg with a heart rate of 120/minute[28,29]. Which one of the following tests would most likely confirm the diagnosis? A. Chest radiography B. Computerized tomography of the chest C. Echocardiography D. Electrocardiography
Comments: MCQs can assess well, any higher level of the cognitive domain, if it is constructed well. It depends mainly on the level of competency in the construction of the MCQ items.

Table-5: The following tables (5a, 5b, 5c) shows SBA type of MCQs using different style of lead-in such as using direct question (?), using statement ends with colon at the end (:) or using statement ends with semi colon (;)at the end.

Table-5a: Example of SBA type of MCQ asking for appropriate option using lead-in as direct question (?)
A 20-year-old lady presents to a clinic with a history of lower abdominal pain, abnormal vaginal discharge, dyspareunia associated with fever, nausea and vomiting for 5 days. What is the most likely diagnosis? A. Acute appendicitis B. Acute pyelonephritis C. Ovarian torsion D. Pelvic inflammatory disease E. Ruptured ovarian cyst
Comments: The options or choices starts with capitalizing the first alphabet of the options answering to a direct question.
Table-5b: Example of SBA type of MCQ asking for appropriate option using lead-in as a statement ends with colon (:) at the end.
A 20-year-old lady presents to a clinic with a history of lower abdominal pain, abnormal vaginal discharge, dyspareunia associated with fever, nausea and vomiting for 5 days. The most likely diagnosis is: A. acute appendicitis B. acute pyelonephritis C. ovarian torsion D. pelvic inflammatory disease E. ruptured ovarian cyst
Comments: The options or choices starts with small letter alphabet relating to answer to the statement ends with colon (:)
Table-5c: Example of SBA type of MCQ asking for appropriate option using lead-in as a statement ends with semi colon (;) at the end.
A 20-year-old lady presents to a clinic with a history of lower abdominal pain, abnormal vaginal discharge, dyspareunia associated with fever, nausea and vomiting for 5 days. The most likely diagnosis would be; A. Acute appendicitis B. Acute pyelonephritis C. Ovarian torsion D. Pelvic inflammatory disease E. Ruptured ovarian cyst

Comments: The options or choices starts with capitalizing the first alphabet relating to answer to the statement ends with semi colon (;) which similar to responses answering to direct question.

Over all comments: How many options will be accepted whether 4 or 5 need to be determined by the faculty in consensus.
Also, answering to the lead-in statement ends with colon (:) and semicolon (;), in which aspect option will start with capitalizing first alphabet and in which aspect will start without capitalizing the first alphabet should be standardized by the faculty in consensus.

Table-6: Example of MCQ of multiple true-false (MTF) type/format

Table-6. Multiple true-false type of MCQ. Here the examinee requires to answer all options that are appropriate as true or false.

A 3-year-old girl is brought to the ED by her mother who presented with shock.

What are common causes of shock in this age group?

A. Arrhythmias
B. Cardiac tamponade
C. Gastroenteritis
D. Intussusception
E. Peritonitis

Comments:The above question is asked using a lead-in as direct question (?)

The question may be asked using a lead-in as statement ends with colon (:), where the options will start with small letter. For example;

The common causes of shock in this age group are:

A. arrhythmia
B. cardiac tamponade
C. gastroenteritis
D. intussusception
E. peritonitis

The question may also be asked using a lead-in as statement ends with semi colon (;), where the options will start with capitalizing first alphabet, similar to responses answering to direct question. For example;

The common causes of shock in this age group;

A. Arrhythmia
B. Cardiac tamponade
C. Gastroenteritis
D. Intussusception
E. Peritonitis

The acceptability of the style of lead-in in the stem need to be decided by the faculty in consensus.

Table-7: Example of MCQs (Table-7a) SBA type with repetition of word/phrase in the options with improved version, and (Table-7b) SBA type with overlapped options including improved version.

Table-7

Repetition of word /phrase in options	**Repetition avoided / improved version**
Which of the following action would decrease the radiation dose from the chest CT least?[5,28] A. Decreasing kVp from 140 to 120 B. Decreasing mA from 250 to 125 C. Decreasing the pitch from 2 to 1 D. Decreasing the scan time 1 to 0.5	The following action that would decrease the radiation dose from the chest CT least, by decreasing; A. kVp from 140 to 120 B. mA from 250 to 125 C. the pitch from 2 to 1 D. the scan time 1 to 0.5
Comments: In improved version, repetitions of word in all options are avoided and used the stem as statement instead of direct question.	

Table-7b:

Options are not mutually exclusive, while D is correct answer, overlapped with A, B and C[30].	Improved version, while answer A is correct, not overlapped with other options
How many chromosomes are found in a typical human cell? A. 18 B. 22 C. 32 D. 46 E. 54	How many chromosomes are found in a typical human cell? A. 46 B. 48 C. 50 D. 52 E. 54
Comments: A cell that contain 22 chromosomes also contain 18; a cell that contain 32, also contain 22 and 18 etc. and so on. Overlapped options may be considered as "trick" items by examinee and excessive use of which can erode trust and respect for the testing process.	

Table-8: Example of extending matching questions or items (EMQ/EMI)[26].

Table-8.
Theme: Back pain Option list: A. Ankylosing spondylitis. B. Aortic dissection. C. Intervertebral disc infection. D. Lumbar spondylosis. E. Metastatic malignancy. F. Pars interarticularis defect. G. Prolapsed intervertebral disc. H. Vertebral fracture. Lead-in statement: For each patient with back pain, select the most likely diagnosis. Stems: Question 1 A 23-year-old man has a 6-month history of lower back pain. His pain is predominantly at the thoracolumbar junction and in the right buttock. The pain is worse in the morning and he has difficulty in getting out of bed. There is some improvement during the day. Examination shows restriction of lumbar spinal movements, particularly lateral flexion. (Answer: A) Question 2 A 32-year-old lady presents with acute onset of low back pain. The pain is constant and is not significantly affected by posture. All spinal movements are painful and difficult. Three weeks earlier, she had a urinary tract infection, which had been treated with amoxicillin. (Answer: F)

Table-9: The following table shows different examples of MCQ with technical item flaws.

Table-9a:
Which one of the following strategies is the best for prevention of type-2 Diabetes Mellitus[10]? A. High-risk strategy. B. Population, high risk and screening strategy.* C. Population-based strategy D. Screening strategy.
Comments: Merging options and longest option is the correct option. Context-free stem.
Table-9b:
What is the target population for primordial preventionactivities[10]? A. High risk groups B. Patients with complications. C. Patients. D. Total population.*
Comments: Repetition of word in the stem and key response (correct option); while it is acceptable to repeat the word in the stem and distracter[10]. The stem is also a context-free stem which assess recall of facts.
Table-9c:
Mr. X is a patient admitted to the A & E in shock. After 10 minutes, the doctor observes an ECG abnormality and subsequently a ventricular fibrillation (VF). Mr. X is unconscious and notbreathing. Which one of the following is mostreliable sign of cardiac arrest[10]? A. Absence of breathing. B. Absence of pulse.* C. Unconsciousness. D. All of the above
Comments: Unnecessary information; unclear or ambiguous information e.g. ECG abnormality and the use of 'all of the above' in the optionsis a flaw.
Table-9d:
Which one of the following factors is NOT associated with Type-2 diabetes[10]? A. IGT and IFG B. Obesity C. Physical inactivity D. Viral infections. *
Comments: Use of negative (NOT) in the stem is not recommended. Stem also not context-rich.
Table-9e:
You, as a resident doctor assessing a patient with brainstem injury in a tertiary care hospital. After performing the Glasgow Coma Scale, which one of the following would be the next step in your assessment plan[10]? A. Conduct lumbar puncture. B. Perform arterial blood gas measurement C. Check cranial nerve functioning and respiratory rate.* D. Perform a pulmonary wedge pressure measurement.
Comments: Logical cue is that the word "cranial" used in option C is logically homogeneous with the word "brainstem" in stem. The options are not in chronological order.

Assessment of students is a matter of continuing concerns for medical educators[31]. The major problems in constructing MCQs are that (i) very few faculty members have formal training in questions construction, (ii) most of the questions are prepared as context-free in the last minutes, where little time exist for vetting or review the quality of questions before examination and (iii) may be no agreement on the standard of the question format and underestimation of the use of blueprint of assessment by teachers in the medical schools. The core part of the student life is assessment[32] and students always learn what we assessnot what we teach[10], although teaching and assessment are the two sides of the same coin. The assessment strategies direct and influence the way students learn[31]. A good assessment is the test that assesses higher level of thinking skills. Students

need some scenario or information in assessment tools to solve a problem which assess higher level skills and which should not be just recall of the fact[10]. Constructing a good scenario, emphasis need to be given so that, the stem should present a definite problem to be meaningful and contain only relevant material, it should be a question or statement, and avoid negatively stated question as examinee may find incorrect answer without knowing the correct[33]. All options should be plausible, clear and concise, mutually exclusive, homogeneous, should not include 'all of the above' and 'none of the above', should be in logical order to avoid bias for certain position, should free from clues[30]. Although there are criticisms that, MCQs tests only recall of information, it is evidenced that MCQs can tests well any higher level of the cognitive domain, if it is constructed well. The criticism is basically due to poor construction; so, efforts must be made to construct high quality MCQs.

This paper offers medical and health professional teachers a window to a comprehensive understanding of different types and aspects of MCQs and how to construct good MCQs that tests higher order thinking skills in the future medical and health professional graduates, thereby ensures competent doctors and other health professionals are being produced. Constructinghigh quality MCQs, institutions must emphasize on faculty development programs concentrating on question construction and implementation of test blueprint or test specification table. A test specification table can guide the test development process especially item writing efforts and test assembly that provides strong content validity and legal defensibility of the examination process[10].

References:

1. Al-Rukban MO. Guidelines for the construction of multiple choice questions tests. J Family Community Med. 2006; 3(3): 125-133.
2. Salam A, Rahim AFA, Aziz RA, Fakir NMRM, Jaafar R. Assessment of the students: tools used in UniversitiSains Malaysia. Bangladesh Medical Journal 2005;34(1):11-13.
3. Menon PSN, Multiple choice question construction. In: Sood R, Paul VK, Sahni P, Mittal S, Kharbanda OP, Adkoli BV, Verma K, Nayar U (Eds.). Assessment in medical education. Trends and tools. K L Wig Centre for medical education and technology. All India institute of medical sciences. New Delhi-India. SEARO WHO project WR/IND HRH 001/LCS, 1995.
4. Sood R, Singh T. Assessment in medical education: Evolving perspectives and contemporary trends. The National Medical Journal of India 2012; 25(6): 357-364.
5. Collins, J. Writing multiple-choice questions for continuing medical education activities and self-assessment modules. Radio Graphics 2006; 26(2): 543-551.
6. Jovanovska J. Designing effective multiple-choice questions for assessing learning outcomes. Infotheca 2018; 18(1): 25-42. DOI 10.18485/ infotheca.2018.18.1.2.
7. Amin Z, Seng CY, Eng KH. Practical Guide to Medical Student Assessment. World Scientific Publishing Co. Pte. Ltd. Singapore, 2006.
8. Zimmaro DM. Writing Good Multiple-Choice Exams. Faculty Innovation Centre. University of Texas – Austin. 2016. Retrieved on 10 November 2019 from: https://facultyinnovate.utexas.edu/sites/default/files/writing-good-multiple-choice-exams-fic-120116.pdf.
9. Asani M. Assessment methods in undergraduate medical schools. Niger J Basic Clin Sci 2012; 9(2):53-60.
10. Abdalla ME, Gaffar AM, Sulaiman RA. Blueprints in Health Profession Education Series. Constructing A-Type Multiple Choice Questions (MCQs): Step By Step Manual, 2011.

11. Srinivasa DK, Adkoll BV. Multiple Choice Questions: How to Construct and How to Evaluate? Indian J Pediatric 1989; 56: 69-74.
12. Schuwirth LW, van der Vleuten CP. Different written assessment methods: What can be said about their strengths and weaknesses? Med Educ 2004;38:974-979.
13. Sian Millard, BenChavez. Writing Multiple Choice and True/False Exam Questions. A good practice guide. University of Hawai'i at Hilo. January 2012. Retrievedon 10 November 2019 from http://uhhcopfacultyresource.weebly.com/uploads/2/1/9/8/2198211/multiple_choice_and_true_false_exam_question_design_booklet.pdf.
14. Chaudhary N, Bhatia BD, Mahato SK, Agrawal KK. Multiple Choice Questions- Part II (Classification, Item Preparation, Analysis And Banking). Journal of Universal College of Medical Sciences 2014; 2(3): 54-59.
15. Brian A. Couch, Joanna K. Hubbard, Chad E Brassil. Multiple–True–False Questions Reveal the Limits of the Multiple–Choice Format for Detecting Students with Incomplete Understandings. BioScience 2018; 68(6): 455-462.
16. Bender DA. MCQ, EMSQ or multiple true/false questions? Bioscience Education 2003; 2(1):1-3. DOI: 10.3108/beej.2003.02000009.
17. Salam A. Input, process, output: system approach in education to assure the quality and excellence in performance. Bangladesh Journal of Medical Sciences 2015; 14(1): 1-2.
18. Salam A. Issues of objective, content, method and assessment in the development of relevant curriculum in medical school. Malaysian Medical Association (MMA) News; April 2010; 40(4): 22-24.
19. Bloom, BS, Engelhart, MD, & Committee of College and University Examiners. Taxonomy of educational objectives: The classification of educational goals. London: Longmans, 1956.
20. Anderson LW & Krathwohl DR.A taxonomy for learning, teaching, and assessing, Abridged Edition. Boston, MA: Allyn and Bacon, 2001.
21. Jamaludin R, Jaafar R, Kaur S. Training module series: student-centered learning (SCL), Approaches for innovative teaching. Module 3: Learning Taxonomies. Centre for Development of Academic Excellence (CDAE).UniversitiSains Malaysia, 2012.
22. Sunita Y Patil, Nayana K Hashilkar, Bhagyashri R Hungund. Blueprinting in Assessment:How much is imprinted in our practice? Journal of Educational Research & Medical Teacher 2014; 2(1):4-6.
23. Adkoli B. Attributes of a Good Question Paper. In. R. Sood (Ed), Assessment in Medical Education: Trends and Tools New Delhi. KL Wig Center for Medical Education and Technology. AIIMS, 1995.
24. Downing SM, Haladyna TM. Validity and its threats. In: Downing SM, Rachel Yudkowsky (Editors). Assessment in Health Professions Education. Routledge, New York, 2009.
25. Featherstone C&Hurst Y. (2014), adapted from Case, S. M and Swanson, D. B. (2001). Constructing Written Test Questions for the Basic Clinical Sciences, NBME: Philadelphia. SBA writing guidelines. RCSI Joint committee on intercollegiate examination. Policy No. G33, v.1.0; 2016.
26. Case SM, Swanson DB. Extended-matching items: A practical alternative to free-response questions. Teach Learn Med 1993; 5:107-115.
27. Butler A. For better multiple-choice tests, avoid tricky questions, study finds. EurekAlertAAAS 2018. Retrieved on 18/2/2018 from: https://www.eurekalert.org/pub_releases/2018-10/wuis-fbm100218.php.
28. Javaeed A. Assessment of Higher Ordered Thinking in Medical Education: Multiple Choice Questions and Modified Essay Questions. University of Ottawa. MedEd Publish -An innovative approach to publishing in medical and health professions education, 2018. DOI: https://doi.org/10.15694/mep.2018.0000128.1
29. Touchie C. Medical Council of Canada Guidelines for the Development of Multiple-Choice Questions, 2010.
30. Brame C. Writing good multiple-choice test questions, 2013. Centre for teaching. Vanderbilt University, 2013.
31. Retrieved on 27 November 2019.from https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/.
32. Salam A, Yousuf R, Mahmood CB. Assessment in medical education. JCMCTA 2004; 15(1-2): 15-18.
33. Haque M, Yousuf R, Abu Bakar SM, Salam A. Assessment in medical education: Bangladesh perspectives. Bangladesh Journal of Medical Sciences 2014; 12(4): 357-363.
34. Bothell TW. Assessment of student learning. Brigham Young University Faculty Center. Annual University Conference, 2001.