

How to Cite:

Chettati, M., & Dahami, Z. (2022). Evaluation of medical assays by multiple-choice questions: Example of 10-years' experience of a North African Country. *International Journal of Health Sciences*, 6(S7), 1632–1649. <https://doi.org/10.53730/ijhs.v6nS7.11431>

Evaluation of medical assays by multiple-choice questions: Example Of 10-years' experience of A north African country

Mariam Chettati

Faculty of Medicine and Pharmacy of Marrakech, CADI AYYAD university Marrakesh- Morocco, Postal Address: CHU Mohammed VI BP2360 Principal, Av. Ibn Sina, Marrakesh-Morocco, telephone: +212634617979
Corresponding author email: m.chettati@uca.ma

Zakaria Dahami

Faculty of Medicine and Pharmacy of Marrakech, CADI AYYAD university Marrakesh- Morocco.

Abstract---Construct: Authors examined the use of multiple choice question tests in assessment of undergraduate students, and made a docimological evaluation of the quality of the tests. **Background:** Moroccan universities are gradually integrating multiple-choice questions as a method of evaluation. At the Faculty of Medicine of Marrakech-Morocco, this type of test is used to assess medical students' knowledge, in addition to easier large-scale assessments and automatic correction, test reliability and validity are major qualities of multiple-choice questions, their important disadvantage is that they are difficult and time-consuming to construct. The aim this study was to report the authors experience in assessment of undergraduate students and to make a docimological evaluation of the quality of the tests. **Approach:** we conducted an observational study on tests used for the assessment of semiology and pathology tests in urology and nephrology, during 10 years (2007 to 2017), we performed the analysis of the different criteria for writing an assay. **Results:** six-hundred and forty questions were analyzed in the pathology modules, 56,2% of questions addressed level I of Bloom's modified taxonomy, level II (28,8%), level III (15%), multiple keyed responses (48.88%), while Single keyed response represented (7.11%), compared to 195 clinical vignettes (43.33%), the ambiguity in the wording of the stem was found in only two questions, while the ambiguity in the writing of options was limited to a single proposal, in semiology modules, 100% of questions assessed knowledge recall only, multiple keyed responses MCQs represented 86.77%, while single keyed response (13.05%). **Conclusions:** we provide an overview of guidelines for writing multiple-choice questions of good quality, we believe that they are an important

method of assessment in health science education, provided they are well-constructed.

Keywords---Multiple-choice questions, evaluation, medical education, Faculty of medicine of Marrakesh

Introduction

Moroccan universities are gradually integrating multiple-choice questions (MCQs) as a method of evaluation. At the FMPM, this type of test is now regularly used to assess medical students' knowledge.

MCQ is "a question to which the student answers by selecting from among several proposed solutions, each of which is judged (by the test constructor and by a consensus among specialists) to be correct or incorrect, regardless of the student who must answer it" [1].

During an assessment by MCQ, the student is therefore invited to select one of the proposed options, the student's ability to recognize a "correct" answer, not to produce it, is measured. Of course, the cognitive processes used by the student for this recognition operation can be diverse and more or less complex.

The use of MCQs, particularly in large-scale assessments, has become popular in educational institutions particularly in medicine.

Automatic correction and objectivity are two major qualities of MCQs [2]. Also tests of this type allow a large number of questions to be asked and thus allow students to be assessed over a large portion of curriculum, in a short period of time requiring less effort on behalf of the student, although it takes a lot of effort and time spent by the examiner to make high quality MCQs, as compared to descriptive questions. MCQs are relatively easy to grade but they are difficult and time-consuming to construct.

The MCQ format allows for test reliability. If MCQs are drawn from a representative sample of content areas that constitute predetermined learning outcomes, they allow for a high degree of test validity. Critics of MCQs argue that higher-level learning cannot be tested with MCQs. However, this criticism is more often attributed to flaws in the construction of the test items rather than to their inherent weakness. Appropriately constructed MCQs result in objective testing that can measure knowledge, comprehension, application, and analysis [3].

The principles of writing effective MCQs are well documented in educational measurement textbooks, the research literature, and test-item construction manuals designed for medical educators [4]. Yet, a recent study from the National Board of Medical Examiners showed that violations of the most basic item-writing principles are very common in medical education tests [5].

Docimology is a term proposed by PIERON, which designates the study of examination techniques (grading methods, inter-individual variability of examiners, subjective factors) [6].

In 1938 LAUGIER and WEINBERG conducted an experiment by distributing papers to be marked by parallel juries, they found different ways of obtaining 2 to 3 points in mathematics and physics and 4 points in French and philosophy. There was a standard deviation of 13 points out of 20 between two examiners [7].

According to Docimology, the aim is to exclude the influence of variables that are not related to educational objectives (teaching behaviour, poor formulation, grammatical errors) in order to have a valid and reliable evaluation method. The increase in the number of questions evaluated and the widening of the docimological possibilities will theoretically allow a much more effective discrimination between students. The automation of the correction process will significantly reduce the costs of proofreading these tests.

II- Example of the experience of the FMPM-CADI AYYAD university: Results of 10 years' research study at the faculty of medicine:

II-1- Methods: we have conducted a single-center observational retrospective study on the MCQ tests used for the assessment of semiology and pathology tests in urology and nephrology at the FMPM, during the academic years (2007 to 2017) of the first and second sessions. We performed the analysis of the different criteria for writing a MCQs test, for each test:

1-The level of knowledge the examiner was looking for according to the modified bloom's taxonomy: 1-a level I: *Knowledge* (learner must recall memorized information but not explain or apply it), 1-b level II: *combined comprehension and application* (learner must demonstrate an ability to use, not just explain, new information, applying rules, methods, concepts, principles, laws or theories) 1-c level III: *Problem solving* (learner must understand components of a concept, and their relationships to each other, and analyze information)

2-Types of multiple choice questions:

-Single keyed response; multiple keyed responses, free; clinical vignette.

3-The respect of the guidelines for test construction in writing MCQs: Writing Stems; writing Options; item Analysis.

II-2- Results: We included a total of 35 assays (19 pathology's and 16 semiology assays).

For the urology/nephrology pathology tests during the period 2007-2017: In our study, the 19 urology/nephrology pathology assays contained 640 MCQs representing the predominant type of assessment with 90.5% and 34 open questions representing 9.5 %, Starting in 2013, MCQs have become the exclusive evaluation method for urology-nephrology pathology tests, so we focused specially on the period from 2013 to 2017.

Table 1

Example of items and the levels of cognitive learning tested according to modified Blooms taxonomy

Level	Objective	Question
I	knowledge recall only	The causes of male infertility can be: A/ Obstructive or excretory B / Non-obstructive or secretory C / Secondary to sexual disorders D / Secondary to ejaculatory disorders E / No answer is right
II	Understanding of a number of different processes to answer the question correctly.	A young patient consults for urinary burns with fever, the diagnosis to be mentioned is: A / Acute cystitis. B / Acute pyelonephritis C / Acute prostatitis D / A urethritis E / Tuberculosis
III	Interpretation of data and problem solving required	In a healthy man under the age of 70 with stage T2 prostate gland cancer, check the appropriate treatment option(s): A / Radical prostatectomy B / Simple monitoring C / Surgical castration D / Chemical castration E / Endoscopic prostate resection

- Percentage of MCQs addressing different levels of Bloom's modified taxonomy (Examples in table 1)
Level I: 56,2%
Level II: 28,8%
Level III: 15%
- MCQ with a single concept submitted for reflection:
In our sample, the urology/nephrology pathology tests included multiple choice questions with a single concept submitted for reflection with a total number of 201 multiple choice questions (44.66 %).
- Types of MCQs as illustrated in table 2: The number of Multiple keyed responses was 221 with a percentage of 48.88, while the number of Single keyed response was 32 with a percentage of 7.11, compared to 195 clinical vignette with a percentage of 43.33.

Table 2
Types of MCQs

year	Multiple keyed responses		Single keyed response		Clinical vignette	
	number	percentage	number	percentage	number	percentage
2013 S1	26	50	5	10	19	38
2013 S2	22	44	1	2	27	54
2014 S1	19	38	7	14	24	48
2014 S2	17	34	3	6	30	60
2015 S1	18	36	1	2	31	62
2015 S2	32	64	0	0	17	34
2016 S1	28	56	5	10	16	32
2016 S2	30	60	2	4	18	36
2017 S1	29	58	8	16	13	26

- Example of MCQ with a single concept submitted for reflection:

Acute tubular necrosis:

- A / Represents the first cause of acute organic renal failure.
- B / Is manifested by acute renal failure.
- C / Is manifested by arterial hypertension, which is common.
- D / Is always accompanied by urinary syndrome.
- E / Is of poor prognosis.

- Prevalence of ambiguity in stem Writing:
The ambiguity in the wording of the stem was found in only two questions (table 5), so the prevalence was 0.45%.
- Prevalence of ambiguity in options writing:

In our study the ambiguity in the writing of options was limited to a single example: proposal "e" with a prevalence of 0,23%.

For urology-nephrology semiology module: In our study, the total number of MCQs in the 16 urology-nephrology semiology assays was 151 MCQs with a percentage of 85.25 %, while the total number of open questions was 18 representing 14.75%. Beginning the year 2011, MCQs have become the exclusive evaluation method. The presence of the clear instructions in writing the items was noted in 10 out of 16 assays used, representing 62.5%. Urology-Nephrology semiology tests from 2012 to 2015 respected the same number of MCQs (12 MCQs per assay on average) so we specially focused on analyzing this period. During the "2012-2015" period, 100% of MCQs assessed knowledge recall only, we did not find any MCQ that assessed level II or III of cognitive thinking. The number of Multiple keyed responses MCQs was 80 with a percentage of 86.77, while the number of Single keyed response MCQs was 12 with a percentage of 13.05, compared to 0 MCQs with clinical vignette as a stem. MCQ with a single concept subject to reflection: the total number of statements containing a single reflective concept in the "2012-2015" study sample was 85 MCQs with a percentage of 92.08%. We did not find any ambiguity in stem or options writing.

- Example of MCQ with ambiguity in stem Writing:
The patient was seen after one year with obstructive acute kidney injury: bilateral hydronephrosis (preserved parenchyma), elevated creatinemia, eupnea and normal Kalemia.

What you should do (only one answer is correct)?

A/ A cystectomy

B/ An anterior pelvectomy

C/ BCG therapy

D/ An endovesical chemotherapy

E/ A systemic chemotherapy

III-Review of the place of MCQs as a method of assessment of medical students:

III-a- Multiple-choice questions on medical exams and their pedagogical implications: Writing educational objectives

Good assay question writing begins with identifying the most important information that is to be learned. Test items should come directly from the objectives and focus on important, relevant content [3]. Controversial test items should be avoided [8]. Key facts can be written as declarative sentences, creating a clear picture of what the student should learn [9].

Objectives should define important knowledge or skills and should be supported by the instruction provided through the educational program. Examples of measurable terms are *state*, *explain*, *list*, *identify*, and *compare*. Immeasurable terms include *know*, *learn*, or *become familiar with* [10].

III-b-Defining Levels of Learning

Cognitive knowledge measurement based on Bloom's taxonomy:

Bloom's taxonomy is based on a model of cognitive objectives developed in the late 1950s by Benjamin Bloom [11], it provides a useful structure on which to base the description and writing of learning objectives. It supports educators in developing essays that build from simple to increasingly complex forms of knowledge [12]. Figure 1 is illustrating a description of each of Bloom's six levels of instructional objectives in the cognitive domain, each in terms of the essential question with which it is associated.

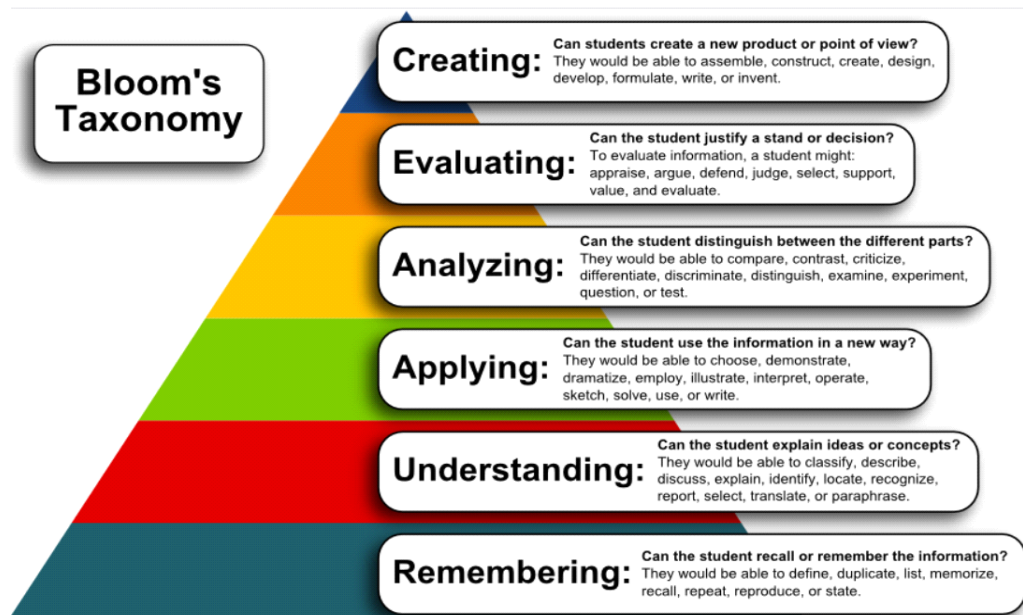


Figure 1: The different six levels of Blooms's Taxonomy
<https://courses.dcs.wisc.edu/design-teaching/PlanDesign>

Some have simplified Bloom's taxonomy into three general levels instead of six [14]. The three levels include 3 categories as shown in table 3.

Table 3
Modified Bloom's taxonomy

Levels	Cognitive domain
Level I:	Knowledge: recall of information
Level II:	Comprehension and application: understanding and being able to interpret data.
Level III:	Problem-solving: use of knowledge and understanding in new circumstances.

A MCQ should test at the same level of learning as the objective it is designed to assess. The program's objectives and test questions should reflect different levels of learning. It is preferable in an item to measure one attribute at a time [15].

A question that provides specific patient information and imaging data (patient vignette) and that asks the learner to choose the most appropriate management is an example of an item that tests problem-solving ability, such questions increase the validity of the examination [10].

The critics concerning the Place of MCQ in assessment of higher ordered thinking argue that many will contain item writing flaws and most will only test factual recall. Studies have shown that this does not necessarily have to be the case, if educators are aware of MCQ writing guidelines, aware of the necessity to have their MCQ peer reviewed to ensure a good quality, educators have to understand that a good MCQ is difficult to write, and should devote all the necessary time and energy to construct a reliable MCQ assessment. The effect of cueing is usually positive and can lead to a higher mean score [16].

Studies suggest that assessment with well-written MCQs has more value than a casually produced open ended assessment. It is possible to use a combination of open questions and MCQs in a sequential problem solving process [17].

IV-Guidelines of redaction in writing MCQs:

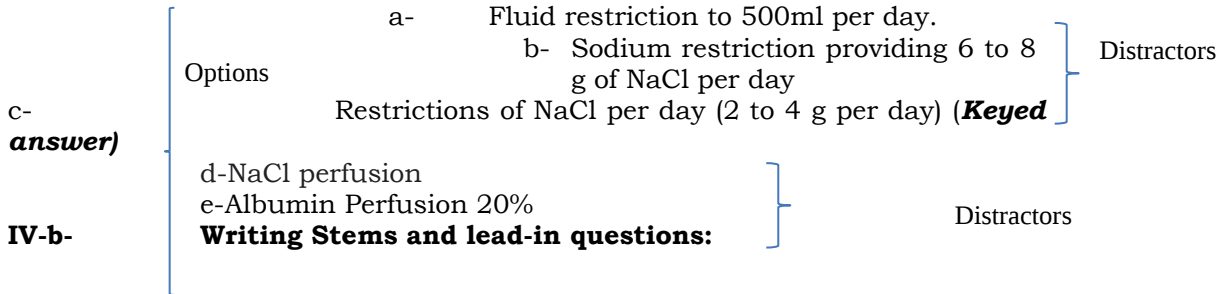
IV-a- Components of a MCQ:

The **item** is the entire test question, ideally, it is composed of a **stem** in which the context of the item is described so that a question can be asked. The question that follows the stem is called **lead- in**. The **options** include a correct alternative (the **keyed response**) and the false alternatives, the (**distractors**).

In the item below you can see the parts of a MCQ item (patient vignette type):

A 26-year-old woman with oedema of the lower limbs reported a nephrotic syndrome. She has gained 4 kg in the past month. Her natremia level is normal.
(**stem**)

What therapeutic measure do you prescribe among the listed options? (**lead- in question**)



The stem is usually composed first and is best written as a complete sentence or question. Direct questions are clearer than sentences completions, research has shown that the use of incomplete stems lowers the students' correct response rate by 10%– 15% [18]. When writing a MCQ, it is a helpful strategy first to conceive the item as a stem ending with an open lead-in question which requires an open short answer. A stem can incorporate maps, diagrams, graphs, or radiologic images, but these should be accompanied by a complete statement. When examinees with enough knowledge can answer an item without looking at the alternatives, it is possible to affirm that the item has a focused stem and lead-in question. If an examinee cannot answer an item without reading the options, this item is considered to be “unfocused” or “stemless” [19]. Examples from our study of stemless items:

- The diagnosis of an obstructive anuria is based on:
- About Membranous nephropathy, what is it true to affirm?

One of the main problems with an item being stemless is that opens the door for the occurrence of the majority of the other technical item writing flaws [20].

The stem should contain all relevant information and as much of the item as possible. If a phrase is stated in the stem, it should not be repeated in the options. As a general guide, students can complete between one and two MCQs per minute [21]. In our study in the semiology assays all the MCQ were stemless with a direct question or a sentence completion.

Use of patient vignette is a good way to test application of knowledge in order to improve the content validity of the examination. Clinical vignettes can provide the basis for the question, beginning with the presenting problem of a patient; they may include the history, physical findings, results of diagnostic studies, initial treatment [19], in our study in the pathology assays 43% of MCQ had a clinical vignette as a stem.

The stem should be stated so that only one of the options can be substantiated and that option should be indisputably correct. When more than one option has some element of accuracy but the keyed response is the best, the stem should ask the student to select the *best* answer rather than the correct answer (it is an entity called one-best MCQ) [22].

Negative questions tend to be less effective and more difficult for the examinee to understand. Negatively posed questions are recognizable by phrases such as “which of the following is not true” or “all of the following except” [8].

Vague frequency terms such as “commonly”, “rarely” and “occasionally” should be avoided as they might become targeted by judicial litigations due to their ill-defined nature. On the other hand, absolute frequency terms such as “all”, “none”, “always” and “never” should be avoided [23].

IV-c-Writing Options:

The number of options should be determined by the educational need, the best number of is three to five. The literature suggests that certain alternatives have such a high level of implausibility that subsequent analyses reveal them to be non-functional distractors or with a low percentage of selection [24].

Although it is quite frequent that in MCQ-based tests all items have the same number of options (usually four or five), there is no rule that the number of options needs to be uniform [21]. In one examination, some items may have four options and some may have five, in our faculty the number of options is standardized to five.

It is recommended to write parallel alternatives regarding degree of complexity and content [15], even though *the keyed response must be unequivocally right and the distractors must be absolutely wrong*, all options should be directed to the same aspect or attribute, they should either be diagnoses, tests, treatments, prognoses, or disposition alternatives. The distractors should preferably be similar in terms of length “*Good items have stems that are longer than the alternatives*”, so, it is better to avoid long, heterogeneous alternatives.

The best distractors are statements that are accurate but do not fully meet the requirements of the problem and incorrect statements that seem right to the examinee [23]. Implausible distractors should not be used. Ideal distractors represent errors commonly made by examinees.

The distractors should not stand out because of their phrasing. Grammatical cues, (when one or more options don’t follow grammatically from the stem) can lead the examinee to the correct option [10].

Options should not include the phrases *none of the above* or *all of the above*. *None of the above* is problematic in items in which judgment is involved and in which the options are not absolutely true or false [25]. Here is an example from our study of a stemless item with heterogeneous alternatives in length and content, and with the mention (no answer is right):

-Overactive bladder:

A/ Is the obligation to urinate in the seconds following the perception of the need to urinate.

B/ Is observed in particular in urinary tract infection (cystitis) or bladder irritation from a bladder tumor.

C/ Is necessarily frequent urination.

D/ Is painful urination.

E/ No answer is right.

Options with repeated elements lead to confusion, unnecessary workload as well as a higher probability of a successful guess due to a “logical convergence strategy” by the test-wise examinee [19]. Also, the options should be classified in logical order: from benign to the most serious, from the smallest to the largest or in alphabetical order: Example of an item with options classified in logical order:

-At what stage(s) of chronic kidney disease do the biological disorders appear?

A/ Stage 1

B/ Stage 2 (mild renal failure)

C/ Stages 3B (moderate to severe renal failure)

D/ Stage 4 (severe renal failure)

E/ Stage 5 (end-stage renal disease)

Options in one item should not reveal information that allows the examinee to automatically know the correct answer to another item. This error in writing MCQs is referred to as “cueing,” when an option in one item provides a hint to the answer for another item [10]: Here is an example of stemless item, with options that are different in length, D and E are longer than the stem, A and B have repeated words:

What is/are the exact proposition's concerning microalbuminuria?

A/ It is positive if it is > 30mg/g.

B/ It is positive if > 3 mg/mmol

C/ It results in a 2+ positive test strip.

D/ It requires annual monitoring upon diagnosis in type 2 diabetes.

E/ Associated with microscopic hematuria, it indicates the existence of diabetic nephropathy

V- Item Analysis (Quality Management Across The Assessment Cycle)

Good quality assurance around test construction is of paramount importance to the validity of test scores, several item-writing principles have been investigated for their effects on test psychometric indices [4]. Downing evaluated the validity of a classroom achievement test in medical education that contained flawed test items. He found that flawed items caused nearly one-quarter more students to fail than unflawed items, because poorly crafted test questions add artificial difficulty to the test scores [26]. Studies demonstrated that flawed MCQ items affected the performance of high-achieving students more than borderline students [24].

Authors of MCQs should review their test items for accuracy and appropriate formatting. It can be very beneficial to have a colleague read and respond to the MCQs and offer feedback. Many medical schools have committees of medical education that can analyze the quality of test items for faculty. It is best for these committees to be interdisciplinary, an example of interdisciplinary review committees is the International Progress Test Review Committee [15], which it is a European committee, composed of medical educators with broad interdisciplinary expertise. Item writers with different backgrounds, and coming from different institutions, send their items to the Committee, that checks the appropriateness of the received items regarding content, language, relevance and absence of item writing flaws, according to the criteria recommended by the European Board of Medical Assessors (EBMA), item writers might receive questions as well as feedback about the items sent to the Committee (Figure 2).

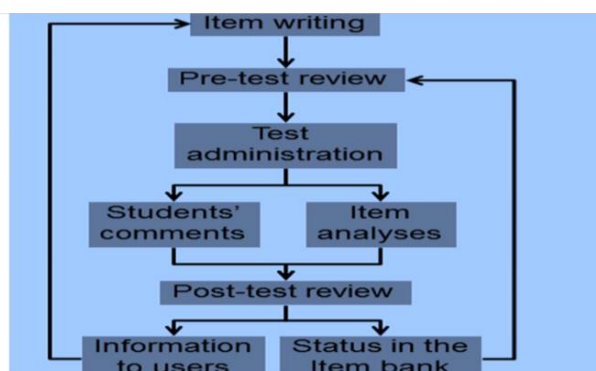


Figure 2: Example of a quality management strategy followed by the International Progress Test Review Committee (Extracted from the EBMA guidelines for item writing version (2017), with permission from the author “Collares CF”)

At the FMPM, the MCQs are written by the teachers in charge of the Module, each teacher asks questions about the course being taught. A meeting for the validation of the items is held with verification of the Check-List issued by the faculty (figure 3).

Fiche technique

Aide-mémoire pour l'élaboration des QCM

- **A Suivre les cinq étapes dans l'élaboration d'une question.**

- 1. Choisir le sujet de la question**
 Les sujets sont les connaissances spécifiques que la question est sensée vérifier. Ils ont trait aux compétences que l'examen devrait mettre à l'épreuve.
- 2. Décider du contexte dans lequel la réalisation de l'objectif est vérifiée**
 Le contexte est la situation clinique (interprétation des données, diagnostic, prise en charge) qui détermine les renseignements qui devraient être fournis dans la question.
- 3. Rédiger le corps de la question**
 Le corps devrait utiliser un cas clinique comme fondement de la question. Il doit aussi comporter tous les renseignements nécessaires pour répondre à la question et se terminer par une question bien précise.
- 4. Créer cinq options de réponse**
 Choisir une (ou plusieurs) option comme étant la réponse correcte et celle-ci devrait l'être de manière évidente. Les autres options, appelées éléments de distraction, doivent être clairement incorrects, mais plausibles à un candidat plus faible.
- 5. Mettre la question à l'épreuve auprès d'un collègue**
 D'autres personnes peuvent plus facilement déceler des problèmes qui peuvent avoir échappés à l'auteur dans l'élaboration de sa question.

Figure 3: Technical sheet of the checklist for MCQs writing procedure at the FMPM (Original "French")

Item analysis is a procedure performed after the examination that provides information regarding the reliability and validity of a test item, it is relatively simple, especially with computer assistance using special software that do processing of the tests' results [28]. Item analysis also tells how difficult or easy the questions were, the difficulty index also called ease index (EI), and whether the questions were able to discriminate between students who performed well on the test, from those who did not, the discrimination index (DI) [29]. Another important technique is analysis of distractors, that provides information regarding the individual distractors and the key of a test item. Using these tools, the examiner is able to modify or remove specific items from subsequent exams. EI describes the percentage of students who correctly answered the item. It ranges from 0 - 100%. The higher the percentage, the easier the item. The recommended range of difficulty is from 30 - 70%, with the goal being to construct a test that contains only a few items that more than 90% or less than 30% of students answer correctly [23]. Items having EI below 30% and above 70% are considered difficult and easy items respectively. Very easy items should usually be placed either at the start of the test as 'warm-up' questions or removed altogether. The difficult items should be reviewed for possible confusing language, areas of controversy, or even an incorrect keyed response. Inclusion of very difficult items in the test depends upon the educational needs; they may be included in order to identify top scorers. DI, also called point biserial correlation, describes the ability of an item to distinguish between high and low scorers. It ranges between -1.00 and +1.00. It is expected that the high-performing students select the correct answer for each item more often than the low-performing students [30]. If this is true, the assessment is said to have a positive DI (between 0.00 and +1.00), indicating that students who received a high total score, chose the correct answer

for a specific item more often than the students who had a low overall score. If, however, the low performing students got a specific item correct more often than the high scorers, then that item has a negative DI (between -1.00 and 0.00). Ideally, each item will have a ratio of at least +0.5 [23]. The ease and discrimination indices are often reciprocally related. However, this may not always be true. Questions having high EI (easier questions), discriminate poorly; conversely, questions with a low EI (harder questions) are considered to be good discriminators [31].

Analysis of distractor is another important part of item analysis. The distractors are important components of an item, as they show a relationship between the total test score and the distractor chosen by the student. Student's performance depends upon how the distractors are designed. Any distractor that has been selected by less than 5% of the students is considered to be a non-functioning distractor, ideally, low-scoring students, should choose the distractors more often, whereas, high scorers should choose the keyed response. By analysing the distractors, it becomes easier to identify their errors, so that they may be revised, replaced, or removed [32].

VI-Scoring modes :

Binary mode is the most commonly used method, including our faculty, in other words, one mark is given for correct answer, and zero mark for wrong answer. This process has the advantage of simplicity of use. However, it lacks subtlety, especially when there is more than one answer among the alternatives [33]. While in the pro-rata weighting mode, the more correct the answers, the more marks the candidate scores. For example, two correct answers (out of 2) equal one point, one correct answer equals half a point, no correct answer equals zero mark. It can also be suggested that some wrong answers will be sanctioned with negative marks (-1), for example if the examinee chooses a very implausible distractor as an answer. There is also variable weighting mode, which means that it is possible to give a different weighting to each correct answer depending on the educational objectives.

VII- Specification table:

Planning for an examination should begin with the development of a test blueprint or "specification table". The test blueprint is defined as a guideline that is used to determine the content and skills to be assessed. It is a double-entry table that indicates the educational objectives or concepts to be assessed and their relevance in (percentage) with a precise scoring of the content of each component based on cognitive levels: memorization, interpretation and application. Test blueprints promote content validity and ensure stability of test content over time (table 4). Thus, validity is the degree to which the evaluations or judgment educators make about students can be trusted based on the quality of gathered evidence [34].

Table4
Types of MCQs : Example of a Specification table

Objectives		Recall of Knowledge: Remember 10%	Comprehension: Interpret 45%	Application: Demonstrate 45%	Number of questions 100%
Describe the becoming of the drug	40%	2,4	10,8	10,8	24,0
Calculate specific pharmacokinetic parameters of an active ingredient from patient data	30%	1,8	8,1	8,1	18,0
Calculate significant changes in pharmacokinetic parameters induced by disease or a disease process	2,5 %	0,2	0,7	0,7	1,5
Use pharmacokinetic parameters to calculate dosage	25%	1,5	6,8	6,8	15,0
Integrate basic pharmacokinetic knowledge with other pharmaceutical disciplines	2,5 %	0,2	0,7	0,7	1,5
Number of questions	100 %	6,0	27,0	27,0	60

VIII-Synthesis :

MCQs can be used as a method of assessment in health science education, at all levels (undergraduate or post graduate or even for continuous education), at the FMPM, this type of test is now regularly used to assess medical students' knowledge. The MCQ format allows for test reliability and for a high degree of test validity, the unfortunate reputation of MCQs of testing only a low cognitive level is a challenge to overcome by writing suitable educational objectives and unflawed items.

Studies argued that MCQs are unsuitable for assessing high ordered thinking, this subject remains controversial. Recent studies suggest that assessment with well-constructed peer-reviewed MCQs has more value than a casually produced

open ended assessment, this type of evaluation can also discriminate the performance levels of the cohort being tested, in addition the low cost is also an advantage that should be taken into consideration.

From our point of view of improving the quality of assessment using MCQs in the FMPM, we suggest that the following recommendation will be taken into consideration:

- Ensure educators continuous training in different assessment techniques, especially in guidelines of writing MCQs.
- Organize seminars to brief students about assessment tools.
- Create an interdisciplinary committee of MCQ validation.
- Do item analysis after each examination.
- Focus on the assessment of medical skills rather than knowledge (a well-made rather than a well-filled head).

References

1. Bloom BS, ed. Taxonomy of educational objectives. Vol I: Cognitive domain. New York, NY: McKay 1956.
2. Bonnie R. Rush, David C. Rankin and Brad J. White. The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. BMC Medical Education 2016; 16:250.
3. Braddom CL. A brief guide to writing better test questions. Am J Phys Med Rehabil 1997; 76: 514 – 516.
4. Case, S. M., & Swanson, D. B. *Constructing written test questions for the basic and clinical sciences* (3rd ed). Philadelphia, PA: National Board of Medical Examiners 2000.
5. CHRETIEN F. Etude docimologique des questions à choix multiples de l'épreuve d'histologie du concours de PCEM1 à la faculté de médecine de Créteil/Paris XII, DER d'Histologie-Embryologie. Faculté de médecine de Créteil-Paris XII ; 2006.
6. Collares CF. GUIDELINES FOR WRITING MULTIPLE-CHOICE QUESTIONS (European Board of Medical Assessors) "EBMA" 2017.
7. Collins J. Education Techniques for Lifelong Learning: Writing Multiple-Choice Questions for Continuing Medical Education Activities and Self-Assessment Modules RadioGraphics 2006; 26: 543–551.
8. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. Collegian 2005; 12: 19-24.
9. Cox KR, Bandaranayake R. How to write good multiple choice questions. Med J Aust 1978; 2: 553–554.
10. Davis BG. Multiple-choice and matching tests. In: Tools for teaching. San Francisco, Calif: Jossey- Bass; 262–271; 1993.
11. Downing SM, Baranowski RA, Grosso LJ, Norcini JJ. Item type and cognitive ability measured: the validity evidence for multiple true-false items in medical specialty certification. Appl Meas Educ 1995; 8: 189 –199.

12. Fowell SL, Southgate LJ, Bligh JG. Evaluating assessment: the missing link? *Med Educ* 1999; 33: 276-81.
13. Fuhrmann BS, Grasha AF. A practical handbook for college teachers. Boston: Little, Brown 1983.
14. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines. *Appl Meas Educ* 2002; 15: 309 –333.
15. Hingorjo MR and Jaleel F, Analysis of one-best MCQs. the difficulty index, discrimination index and distractor efficiency. *J Pak Med Assoc* 2012; 62(2) : 142-7.
16. <https://courses.dcs.wisc.edu/design-teaching/PlanDesign>
17. Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew H. The quality of in-house medical school examinations. *Acad Med* 2002 ; 77 : 156 – 161.
18. Kemp JE, Morrison GR, Ross SM. Developing evaluation instruments. In: *Designing effective instruction*. NewYork, NY: MacMillan College Publishing 1994: 180 –213.
19. Kent TH, Jones JJ, Schmeiser CB. Some rules and guidelines for writing multiple choice test items. Iowa City, Iowa: University of Iowa College of Medicine and American College Testing Program 1974.
20. Laugier H, Weinberg D. « Recherche sur la solidarité et l'interdépendance des aptitudes intellectuelles d'après les notes d'examens écrits du baccalauréat », Paris, CHantenay 1938.
21. LECHEVALLIER J, [1] Règles d'élaboration d'une évaluation par Questions à Choix Multiples. [2]
22. Leclercq, D. La conception des Questions à Choix Multiple. Bruxelles : Labor ; 1986.
23. Newble D, Cannon R. Curriculum planning. In: *A handbook for teachers in universities and colleges: a guide to improving teaching methods*. 3rd ed. London, England: Kogan 1995.
24. 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* 2007; 7: 49.
25. Piéron, H. « Examens et docimologie » 1969 Paris, PUF, J.O. du 27/03/92 : 4231-4233.
26. Schuwirth LWT, van der Vleuten CPM, Donkers HHLM. A closer look at cueing effects in multiple-choice questions. *Med Educ* 1996; 30: 44-49.
27. Sim SM, Rasiah RI. Relationship between item difficulty and discrimination indices in true/false - type multiple choice questions of a para-clinical multidisciplinary paper. *Ann Acad Med Singapore* 2006; 35: 67-71.
28. Srinivasa DK, Adkoli BV. Multiple choice questions: how to construct and how to evaluate? *Indian J Pediatr* 1989; 56 : 69 –74.
29. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). Get vaccinated when it is your turn and follow the local guidelines. *International Journal of Health Sciences*, 5(3), x-xv. <https://doi.org/10.53730/ijhs.v5n3.2938>
30. Tan LT, McAleer JJ. Final FRCR Examination Board. The introduction of single best answer questions as a test of knowledge in the final examination for fellowship of the Royal College of Radiologists in Clinical Oncology. *Clin Oncol (R Coll Radiol)* 2008; 20: 571-6.
31. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distracters in multiple-choice questions: a descriptive analysis.

- BMC Med Educ 2009 ; 9: 40. ^[17]_[SEP]
32. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. *Med Educ* 2008; 42: 198-206.
 33. Tarrant, M., & Ware, J. A comparison of the psychometric properties of three- and four- option multiple-choice questions in nursing assessments. *Nurse education today*, 30(6), 539-543; 2010.
 34. The National Board of Medical Examiners. Constructing written test questions for the basic and clinical sciences. Available at: <http://www.nbme.org/about/itemwriting.asp>. Accessed May 25; 2005.
 35. Wolming, S. & Wikstrom, C. The concept of validity in theory and practice. *Assessment in Education: Principles, Policy & Practice* 2010; 17: 117-132.