

## Evaluation of multiple-choice and short essay questions in pharmacology education

Ali I. Banigesh<sup>1\*</sup>  , Abdulsalam O. Elfowiris<sup>2</sup>   and Abdulsalam Sughir<sup>3</sup>  

<sup>1</sup> Department of Pharmacology and Toxicology, Faculty of Pharmacy, University of Benghazi, Benghazi,

<sup>2</sup> Department of Pharmacology and Toxicology, Faculty of Pharmacy, Omar Al-Mukhtar University, Bayda,

<sup>3</sup> Libyan Academy for Postgraduate Studies, Tripoli, Libya

\* Author to whom correspondence should be addressed

**Received:** 18-03-2023, **Revised:** 14-04-2023, **Accepted:** 22-04-2023, **Published:** 30-06-2023

**Copyright**© 2023. This open-access article is distributed under the *Creative Commons Attribution License*, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### HOW TO CITE THIS

Banigesh et al. (2023) Evaluation of multiple-choice and short essay questions in pharmacology education. *Mediterr J Pharm Pharm Sci.* 3 (2): 13-18. [Article number: 110]. <https://doi.org/10.5281/zenodo.7869133>

**Keywords:** Cognitive level, IWFs, Libya, MCQs, questions, SEQs

**Abstract:** Multiple choice questions (MCQs) and short essay questions (SEQs) are common methods of the assessment of medical students in courses of pharmacology. Poorly constructed test items (questions) are a widespread problem resulting in failure to assess learning objectives. It has been reported that there are 36.0% to 65.0% flawed test items in medical education assessment tools. Thus, the objective of this study was to evaluate MCQs by determining the item writing flaws (IWFs) and to evaluate the SEQs by determining the cognitive level of each item. Four pharmacology tests were administered to third-year pharmacy students at the Department of Pharmacology, Faculty of Pharmacy, Omar Al-Mukhtar University, Bayda, Libya. These were evaluated by determining the IWFs and the level of the cognitive domains. Based on Buckwalter's modification of Bloom's taxonomy cognitive level, for the SEQs, 30.0% of the questions were attempted to check recall of information, 26.0% were attempted to evaluate understanding and interpretation of data and 43.0% of the questions were attempted to check the application of knowledge for solving a particular problem. For the MCQs, 94.6% of the questions were attempted to evaluate the understanding and interpretation of data. For the IWFs, there were more than 40.0% of flawed questions. The most common writing flaws were the negative stem (47.4%), unfocused item (16.0%), non-homogenous in grammar and contents (10.0%), all the above (10.0%) and clang association (05.0%). In a short essay, the SEQs were of excellent quality because they were equally distributed among the three levels of cognitive (level I, II and III). On the other hand, the most common mistakes IWFs of the MCQs were the negative stem (47.0%) and the idea was not clearly and concisely stated in the stem (16.0%). This study concludes that questions in SEQs are valid to measure the learning objective but MCQs were not in pharmacology courses in Libya.

### Introduction

The main objectives of any educational program are to enable learners to develop different cognitive abilities, such as recalling the fundamentals or the principles (lower cognitive order) or problem-solving or clinical reasoning. To ensure that the learners acquired the intended performance or competency, evaluation or assessment must be conducted and aligned with the learning objective, during or at the end of the education program. In pharmacology education, assessment(s) is an overly critical tool that evaluates the learners'



knowledge, skills and attitude consequently learners' outcomes or educational quality. Different type of assessment demands more emphasis on a certain construct (category) for example objective structured clinical examination (OSCE) uses standardized patients to evaluate counseling, clinical procedure, application and clinical problem-solving. The OSCE, for instance in Pharmacy School, must be passed by examination to be a practice pharmacist. On the other hand, MCQs and SEQs are used to assess basic knowledge and clinical reasoning based on learning objectives resulting either in passing or failing in educational courses [1]. In any assessment, two elements: cognitive level and item writing principles are essential in governing the quality and validity of the assessment.

Bloom specifies cognitive abilities as knowledge (facts, basics and principles), comprehension, application, analysis, synthesis and evaluation. Most educational programs such as medical and pharmacy practice attempt to develop a medical curriculum (both instruction and assessment) that incorporates not only basic information (basic and principles) but also complex cognitive abilities [2]. Thus, quality crafted assessment should reflect the basic and the more complex skills. Students are evaluated based on these cognitive abilities [2]. According to the revised Bloom's taxonomy, most of the learning objectives and the examination questions were classified under the lower level of cognitive domains [3]. Despite the importance of complex cognitive skills, most educational programs measure just the recall or basic or fundamental information [2]. Measuring just basic knowledge results in learners with poor learning outcomes. Recently, modified Bloom's cognitive levels categorize the cognitive levels into level I (recall of information), level II (understanding and application) and level III (problem-solving). In addition, for writing MCQ guidelines, there were no recognized guidelines for writing well-constructed MCQs. Nevertheless, Haladyna and Downing [4] developed guidelines for writing well-constructed MCQs. Following the standard items of writing principles results in well-crafted questions. However, failing or violating the guidelines or principles of writing MCQs results high frequency of flawed items in the tests studied (item writing flaws, IWFs) [5], pausing the challenge to learners. Further, flawed items in MCQs lead to incorrectly classifying students as failed when they should have been classified as passed. MCQs consist of the stem (posing question), 3-5 alternative options: one correct answer (key) and distractors (2-4 wrong answers). With the ease of scoring, higher reliability and validity, MCQs is commonly used either alone or in combination to measure the basic knowledge or synthesis, application and problem-solving skills [6] as described by the learning objectives. On the other hand, essay-type assessment is time-consuming but can be used alone or in combination to measure the ability of learners to recall facts and to apply or solve problems [6]. One of the most common problems of MCQs is the IWFs, i.e., unfocused stem or implausible distractors. Also, the deficiency of the SEQs is the focus on the low-order cognitive domain, such as the ability to recall facts. Flaw items or low-order cognitive questions fail to assess learning objectives [6]. In short, the SEQs may emphasize one level of the modified Bloom's cognitive levels while the MCQs may have flawed questions such as negatively worded items or unfocused items. Thus, the purpose of this study was to evaluate MCQs and SEQs items in undergraduate students of the Pharmacology course by determining IWFs in MCQs and cognitive levels of each item in MCQs and SEQs in Libya.

## Materials and methods

This analytical study was conducted at the Department of Pharmacology, Faculty of Pharmacy, Omar Al-Mukhtar University, Bayda, Libya. The study was conducted between the academic years (2020 and 2022). All the examinations were performed by qualifying lecturers of pharmacology with experience in teaching for more than five years. Different methods of assessments such as MCQs and SEQs were used as modules of assessments. The Department developed all the assessment questions in multiple-choice or short essay questions. Four pharmacology examinations involving MCQs and SEQs according to **Tables 1** and **2**. Each question in the MCQs was analysed to evaluate for IWFs (**Table 2**). Furthermore, MCQs and SEQs were

analysed to evaluate the cognitive level (**Table 1**). The cognitive levels of the assessment tools were analyzed using the modification of Bloom's taxonomy [7]: Level I includes questions that attempt to check the recall of information. Level II includes questions that attempt to evaluate the understanding and interpretation of the data. Level III includes questions that attempt to evaluate the application of knowledge for solving a particular problem (**Table 1**). For determining types of IWFs standard criteria given by previous studies [4, 5, 8] were used and commonly occurring violations of item-writing guidelines were identified (**Table 2**). Although there were no humans involved in the descriptive study, we mean the study was assessing the validity of the pharmacology examinations. Ethical consent was obtained from the Department of Pharmacology, Faculty of Pharmacy, Omar Al-Mukhtar University, to evaluate the examinations.

**Table 1:** Modified Bloom's taxonomy

|                  |                               |   |
|------------------|-------------------------------|---|
| <b>Level I</b>   | Knowledge                     | Recall of information                                   |
| <b>Level II</b>  | Comprehension and application | Understanding and interpretation of data                |
| <b>Level III</b> | Problem solving               | Use of knowledge and understanding in new circumstances |

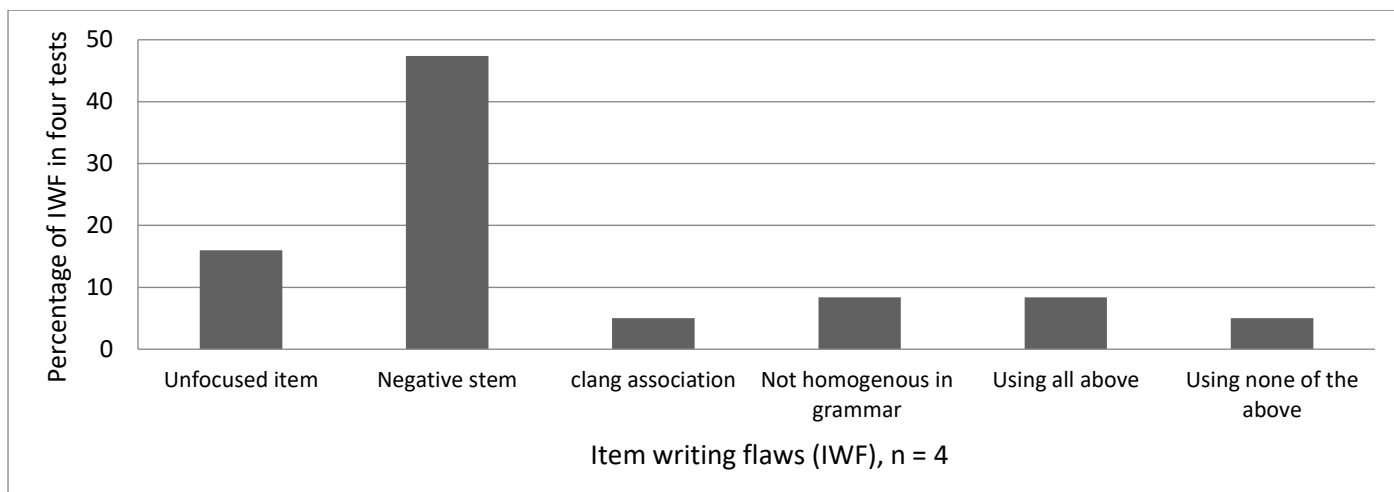
**Table 2:** Judging the MCQs according to the presence of IWFs

|                     |  |
|---------------------|--|
| <b>Type of IWFs</b> | Idea is not clearly and concisely stated in the stem (unfocused) |
|                     | Negative stem  |
|                     | Clue to the right answer (clang association)                     |
|                     | Not homogenous in grammar or in content structure                |
|                     | Using all above  |
|                     | Using none of the above  |

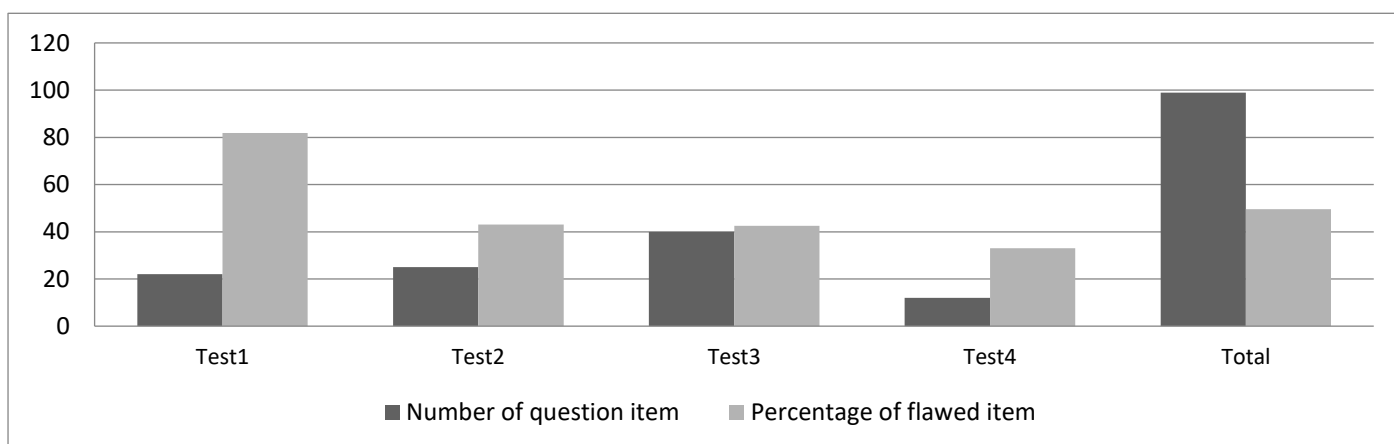
*Statistical analysis:* This is a quantitative study and a descriptive statistic has been used to analysis the percentage of the cognitive level in MCQs and in SEQs, and the percentage of flawed items in the four tests. Based on Haladyna and Downing who have developed guidelines for writing quality crafted questions [9], therefore, a prediction can be obtained if MCQs or SEQs are well constructed and can measure the learning outcomes.

## Results

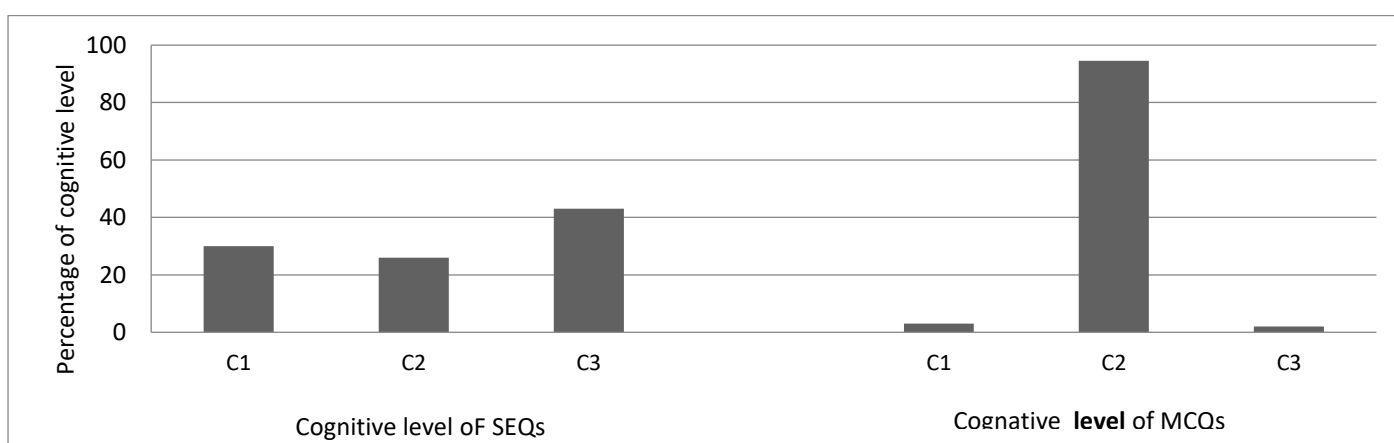
In this study, the examinations are designed to assess lower cognitive skills such as recall of information and higher cognitive abilities such as problem-solving quality. Items in the MCQs or in the SEQs are designed to test the student's ability to recall, understand, to apply what the students have learned. To assess students' academic successes, these assessments should not have any IWFs and the questions should be evenly distributed between the different cognitive levels. To verify if the items in MCQs or SEQs, assess the student's ability in lower and upper cognitive skills and free of any IWFs, four pharmacology examinations were selected and analyzed for the level of cognitive ability and the occurrence of writing flaws. For the IWFs flaws, (**Figures 1 and 2**) of the one hundred questions, there were more than 40.0% flawed questions. *What are the most common writing flaws?* The most common writing flaws were the negative stem (45.0%), unfocused item (15.0%), non-homogenous in grammar and in contents (10.0%), all the above (10.0%), clang association (5.0%) and none of the above (5.0%). For the level of cognitive ability, in SEQs, **Figure 3**, the questions were scattered testing the recall of information (25.0%), application (20.0%) and problem-solving (40.0%). Nevertheless, in MCQs, **Figure 3**, only 3.0% of the question were evaluating the recall of information, 2.0% were evaluating the problem solving while many of the questions (94.6%) were testing the understanding and interpretation. In short, in SEQs, there was a scattered distribution of the cognitive levels while in MCQs, most of the MCQs were intended for the understanding and interpretation of the data (94.6%).



**Figure 1:** Percentage of IWFs in four tests of Pharmacology course



**Figure 2:** Percentage of flawed items in four tests of Pharmacology course



**Figure 3:** Percentage of different cognitive levels in four pharmacology tests

Where C1, C2 and C3 are represent the cognitive level I, II and III, respectively.

## Discussion

Assessment is an essential part of the learning process. It usually spreads to cover the lower cognitive abilities such as assessing learners to recall factual knowledge or assessing higher cognitive skills for instance learner's ability to analyse, apply or solve problems [6]. In addition, the occurrence of writing flaws or uneven distribution of the cognitive levels in writing the test reduces the test validity or quality in assessing the learners' academic achievements. This study found that almost half of the flawed questions were in four different pharmacology tests in Libya. Poorly crafted test items are common problems resulting in the wasting



of time, resources and money [10]. In medical education, 36.0%-65.0% of test items were flawed and due to flawed MCQs, 10.0%-15.0% of the students who failed should have passed [6]. As a result, the flawed tests are not valid to measure the students learning [9]. Also, the most typical of writing flaws were negative stems (45.0%), unfocused stems (15.0%), non-harmonized grammar (10.0%), all of the above (10.0%) and none of the above (05.0%). Data from the current study positively correlate to the previous study in which found that most MCQs flaws are negatively worded stems, unfocused-item stems, and use of the all of the above and none of the above [6].

The multiple assessment formats, MCQs and SEQs, are used not only to assess the lower order cognitive skills (level I): recall of facts but also to assess the higher cognitive abilities (level II): understanding and application (level III) and problem-solving [6, 11]. In MCQs, the current study shows that only 3.0% of the question items assessed recalling facts, 02.0% of the questions assessed problem-solving and 94.6% of questions assessed understanding and application. On the other hand, in SEQs, 30.0% of the questions were in level I (recall of facts), 26.0% were in understanding and application (level II) and 43.0% were in the problem-solving (level III). Present findings indicated that questions in MCQs and SEQs, using Bloom's taxonomy cognitive levels were unevenly distributed between cognitive levels. It is a potential that the questions measure the entire span of Bloom's cognitive level. The current study also showed that a greater proportion of SEQs (30.0%) than MCQs (3.0%) were testing lower levels of cognitive skills. Similarly, greater SEQs (26.0% level II and 43.0% level III) than MCQs were evaluating the higher cognitive skills. However, SEQs were evenly distributed between cognitive levels. In correlation with the literature, Palmer et al. [11] showed that a greater proportion of questions tested lower levels of cognitive skills. This means that the assessment (i.e., MCQs) did not cover the spectrum of Bloom's cognitive level resulting in inferior quality or validity assessment. The present study showed that a greater proportion of SEQs (30.0%) were testing lower levels of cognitive skills. In line with the literature, Palmer et al. [11] showed that more proportion of questions tested lower levels of cognitive skills, what does this mean? This intended that the assessment (i.e., MCQs) did not cover the spectrum of Bloom's cognitive level resulting in inferior quality or validity assessment. Further, it is a trend to find most of the questions focus on testing low-order thinking skills such as what are the side effects of propranolol? Or which of the following medications are used for pregnancy hypertension? These types of questions assess students on recalling or remembering facts, they are easy to design. In contrast, it is not common to find questions that assess higher order thinking skills, for instance, compare the effects of high-dose (1000 mg/d) and low-dose (100 mg/d) aspirin on a long-term cardiovascular patient?

*Conclusion:* This study indicates that questions in SEQs are evenly distributed between cognitive levels while in MCQs, the questions are unevenly distributed in Libyan medical universities. Also, the assessment has 50.0% flawed items, and most item writing flaws are negatively worded stem, unfocused stem, non-homogeneous in grammar, using all the above and none of the above. Thus, it can be concluded that SEQs are valid to measure the learning objective but MCQs are not in Libyan medical universities.

**Author contribution:** AB conceived, designed the study and performed the analysis. AE & AS contributed to collecting and analysis of data. All authors have drafted and revised the manuscript as well as approved the final version of the manuscript and agreed to be accountable for its contents.

**Conflict of interest:** The authors declare the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Ethical issues:** Including plagiarism, informed consent, data fabrication or falsification and double publication or submission were completely observed by the authors.

**Data availability statement:** The raw data that support the findings of this article are available from the corresponding author upon reasonable request.

**Author declarations:** The authors confirm that all relevant ethical guidelines have been followed and any necessary IRB and/or ethics committee approvals have been obtained.



## References

1. Downing SM (2003) Validity: On the meaningful interpretation of assessment data. *Medical Education*. 37 (9): 830-837. doi: 10.1046/j.1365-2923.2003.01594.x
2. Buckwalter JA, Schumacher R, Albright JP, Cooper RR (1981) Use of an educational taxonomy for evaluation of cognitive performance. *Journal of Medical Education*. 56 (2): 115-121. doi: 10.1097/00001888-198102000-00006
3. Kozikoğlu İ (2018) The examination of alignment between national assessment and English curriculum objectives using revised Bloom's taxonomy. *Educational Research Quarterly*. 41 (4): 50-77. doi: Nil.
4. Haladyna TM, Downing SM (1989) A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*. 2 (1): 37-50. doi: 10.1207/s15324818ame0201\_3
5. Downing SM (2005) The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. *Advances in Health Sciences Education*. 10 (2): 133-143. doi: 10.1007/s10459-004-4019-5
6. Baig M, Ali SK, Ali S, Huda N (1969) Quality evaluation of assessment tools: OSPE, SEQ & MCQ. *Pakistan Journal of Medical Sciences*. 30 (1): 3-6. doi: 10.12669/pjms.301.4458
7. Betts SC (2008) Teaching and assessing basic concepts to advanced applications: Using Bloom's taxonomy to inform graduate course design. *Academy of Educational Leadership Journal*. 12 (3): 99-107. doi: Nil.
8. Boland RJ, Lester NA, Williams E (2010) Writing multiple-choice questions. *Academic Psychiatry*. 34 (4): 310-316. doi: 10.1176/appi.ap.34.4.310
9. Michael C, Rodriguez MC (2005) Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*. 24 (2): 3-13. doi: 10.1111/j.1745-3992.2005.00006.x
10. Nedeau-Cayo R, Laughlin D, Rus L, Hall J (2013) Assessment of item-writing flaws in multiple-choice questions. *Journal for Nurses in Professional Development*. 29 (2): 52-57. doi: 10.1097/NND.0b013e318286c2f1
11. Palmer EJ, Duggan P, Devitt PG, Russell R (2010) The modified essay question: Its exit from the exit examination. *Medical Teacher*. 32 (7): e300-7. doi: 10.3109/0142159X.2010.488705