Original Article

Errors in MCQ banks: A need to improvise the quality of multiple choice questions

Prashant E. Natekar *, Fatima DeSouza, Nandini Karapurkar

Department of Anatomy, Goa Medical College, Bambolim, Goa, India

ABSTRACT

Introduction: Assessment of learning is an important component of the education system. The multiple choice question is the commonest method used in written tests for the undergraduate, postgraduate and high stake examinations. Hence, the prevalent tests should be reliable and valid as they attribute quality and efficacy to the assessment system.

Methods: Forty students of First year Bachelor of Dentistry course attended MCQ test in the subject of Anatomy at Goa Medical College, Bambolim, Goa. 25 MCQs of the best single response type with four options were selected from the available textbooks (question bank) in anatomy. Each item comprising of one mark had to be answered in 50 s. There was no negative marking and 50% score was considered as pass. Pre-validation of MCQs was not done considering that the items may have been already pre-validated.

Post-validation of the MCQs was done by item analysis for difficulty index, discrimination index and distractor effectiveness.

Results: The mean score of the difficulty index (*p*) showed that 14 (56%) of the items were within the acceptable limits, 11 (44%) were too difficult and too easy were nil; discrimination index (*d* value) showed that 7 (28%) needs to be discarded, 5 (20%) of the items were acceptable, 4 (16%) were good and 9 (36%) were excellent; distractor effectiveness (de) showed that 2 (8%) items were non-functional distractors, 3 (12%) items had negative discrimination and 20 (80%) were functional distractors.

Discussion: Item analysis can tell us if an item was too easy or too difficult, how it can be discriminated between high and low achievers and also whether all the distractors were effective.

© 2016 Published by Elsevier, a division of Reed Elsevier India, Pvt. Ltd on behalf of Anatomical Society of India.

1. Introduction

Multiple choice questions (MCQs), also known as "test items", are an effective and efficient way to assess the various levels of learning outcomes. Appropriately constructed multiple choice questions result in objective testing that can measure knowledge, comprehension, application and analysis.[1]

Well-formulated MCQs assess cognitive, affective and psycho-motor domains and are preferred over other methods because they ensure objective assessment, minimal effect of examiners bias, comparability and cover a wide range of subjects.[2]

Properly constructed MCQs can assess higher cognitive processing (Blooms taxonomy) such as interpretation, synthesis and application of knowledge instead of just testing recall of isolated facts.[3]

Usually, educators develop test items by themselves or sometimes rely on item test bank as a source of questions. The possibility of error is more in case of test banks if their staff members are not well educated and professionally trained enough for the development of test items.[4]

A well-constructed item is capable of testing the higher levels of cognitive reasoning and can efficiently discriminate between high and low achievers.[5,6] Very often we refer to the published MCQs and select various items as per the curricula in our formative and summative examinations because framing MCQs is tedious, time consuming, involves hard work, cooperative efforts or due to lack of faculty development programmes. The aim of this study is to assess the quality of MCQs from the published text books in anatomy by item analysis, its learning outcomes and to utilize them as a question bank for future use

---

* Corresponding author. Tel.: +91 9158864325.
E-mail address: drpenatekar@gmail.com (P.E. Natekar).

in formative and summative assessment and also for its use in preparation of All India Postgraduate entrance examinations in Dentistry.

2. Materials and methods

Forty students of First year BDS course attended 25 MCQ test on gross anatomy of head face and neck during their formative examination at Goa Medical College, Bambolim, Goa. The MCQs were best single response type with four options, each item comprising of one mark to be answered in 50 s for 25 marks. There was no negative marking and 50% score was considered as pass. Post-validation of the MCQs was done by item analysis. The correct answer was referred as "Key" and there was only one correct answer for each MCQ.[7,8]

These MCQs were obtained from the textbooks available in gross anatomy and hence pre-validation of MCQs was not done considering that the items might have been already pre-validated.

3. Evaluation

Post-validation of the MCQs was done by item analysis for difficulty index, discrimination index and distractor effectiveness.

The papers were evaluated and students were ranked in the order of merit. These papers were arranged in descending order according to their scores. The top one-third were labelled as high achievers group (HAG) and the lower one-third were labelled as low achievers group (LAG). The middle third were not considered for the study.

A table was prepared for each item as follows: Correct Key. **C.**

Options	No. of students selecting option amongst HAG (H)	No. of students selecting option amongst LAG (L)	Total response N%
A			
B			
C			
D			
Not attempted			
Total	13	13	26

Post-validation for each item was analyzed for the following.

3.1. Difficulty index (P)

Percentage of students who selected the correct response. Whether the item had appropriate level of difficulty. It helps to understand about the learning that happened in the classroom.[9] $P = H + \frac{L}{N} \times 100$

where H, number of students in HAG answered correctly; L, number of students in LAG answered correctly and N, total number of students.

3.1.1. Interpretation of difficulty index (P): p value

<30	30	40	50	60	70	>70
Too difficult			Good			Too easy reject
			Acceptable			

3.2. Discrimination index (d)

Whether the item is capable of discriminating between knowledgeable and ill-informed students. The maximum discrimination index that an item can have is 1.

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

3.2.1. Interpretation of discrimination index (DI): d value

<0.15	0.15–0.24	0.25–0.34	>0.35
Discard	Acceptable	Good	Excellent

3.3. Distractor effectiveness (DE): de

Effectiveness of the option. Any distractor attracting less than 5% of the total response is said to be non-functional. It is useful to get feedback on effectiveness/functionality of each alternative, since poor alternatives would lead to greater possibility of guessing the correct answer.[10]

4. Results

In the present study, 25 items selected from the available textbooks were analyzed for difficulty index, discrimination index and distractor effectiveness.

The mean score of the difficulty index (p) showed that 14 (56%) of the items were within the acceptable limits (P value 30–70%), 11 (44%) were too difficult (P value <30%) and too easy (P value >70%) were nil, as shown in Table 1.

The mean score of discrimination index (d value) showed that 7 (28%) needs to be discarded (d value <0.15), 5 (20%) of the items were acceptable (d value 0.15–0.24), 4 (16%) were good (d value 0.25–0.35) and 9 (36%) were excellent (d value >0.35), as shown in Table 2.

In analysis of distractor effectiveness (de), 2 (8%) items were non-functional distractors, 3 (12%) items had negative discrimination and 20 (80%) were functional distractors, as shown in Table 3.

Feedback from trained teaching faculty and students was obtained by open-ended questions and their responses were rated as per Likert scale.

4.1. Open-ended feedback responses from faculty and students

	Agree (%)	Undecided (%)	Disagree (%)
Feedback from faculty			
Tests cognitive domain	76	–	24
Can be administered in short period	55	18	27
Can be assessed by computer	100	–	–
Detect technical flaws	68	22	10
Aids in selection of valid MCQs	73	11	16
Feedback from students			
Chances of guess work	87	–	4
Tests Knowledge accurately	48	04	48
Easy/too difficult	15	63	2
Gets confuse with other options	76	17	7

Table 1
Difficulty index of MCQ.

Total number of items 25		
Parameters (*p* value)	Items	Percentage
<30% (too difficult)	11	44
30–70% (acceptable)	14	56
>70% (too easy)	Nil	Nil

Table 2
Discrimination index of MCQ.

Total number of items 25		
Parameters (*d* value)	Items	Percentage
<0.15 (discard)	7	28
0.15–0.24 (acceptable)	5	20
0.25–0.35 (good)	4	16
>0.35 (excellent)	9	36

Table 3
Distractor effectiveness of MCQ.

Total number of items 25		
Parameters (de value)	Items	Percentage
Functional distractors	20	80
Non-functional distractors	2	8
Negative discrimination	3	12

5. Discussion

The assessment (summative and formative) by means of MCQs forms an important component of evaluation in teaching learning process in addition to long assay and short assay questions. MCQs help in assessing the students wherein the syllabus is very vast so as to rank them in University as well as high stake examination. Pre-validation of properly constructed MCQs is very tedious, time consuming, and cooperative efforts are needed not only to test the standard or quality but also levels of knowledge. The good quality of MCQs can only be possible if they are subjected to item analysis.

Designing good MCQs is a complex challenging and time consuming process. Due to time constraint and lack of cooperative efforts and invariably lack of faculty development programme, one always tends to select readymade MCQs published and available in the textbooks for formative as well as summative examination.

Students normally try to learn topics which they believe that it may appear in the examination. Hence, proper designing of MCQs will have an impact on education wherein it will stimulate their higher levels of thinking skills and help them in making proper decisions. MCQs are a better tool for assessment of cognitive domain, but not a preferred tool for assessing the psychomotor and affective domains.

An earlier study conducted on item analysis reported that 61% difficulty index was in acceptable range (*p* 30–70%), 24% (*p* > 70%) as too easy and 15% items were too difficult (*p* < 30%).[10] A study also reported that difficulty index showing 62% items in acceptable range (*p* 30–70%), 23% were too easy (*p* > 70%) and 15% were too difficult (*p* < 30%).[12]

Another study also revealed that difficulty index showed 80% of items were in acceptable range (*p* 30–70%) and 20% in unacceptable range (*p* > 70% and *p* < 30%) whereas discrimination index (*d*) showed 40% items >0.35, 42% between 0.2–0.34 and 18% <0.20.[13] Similarly, the difficulty index of the 62% items was in acceptable range (*p* 30–70%), 32% (*p* > 70%) too easy and 6% too difficult

(*p* < 30%) whereas the discrimination index showed that 52% items were >0.35, 18% between 0.2–0.34 and 30% items had <0.2.[14] The negative discrimination which has been reported in 20% was probably due to wrong key, ambiguous framing of question or generalized poor preparation of the students.

When distractor effectiveness was analyzed in other study, it was observed that 52.2% were functional distractors (FD), 35.1% were non-functional distractors (NFD) and 10.2% were not chosen by any student.[15] Studies also revealed that 1.1–8.4% were FD and 38% were NFD,[8] similarly 18.16% were FD and 35.33% NFD and 46.01% had nil response.[10]

In our earlier study on the properly constructed and pre-validated items, the difficulty index (*p*) was 80% in acceptable range, 10% was too easy and 10% was too difficult. The discrimination index (*d*) was good 50%, excellent 30% and acceptable 20%.[16]

We observed that the mean score of the difficulty index (*p*) showed that 14 (56%) of the items were within the acceptable limits (*p* value 30–70%), 11 (44%) were too difficult (*p* value <30%) and too easy (*p* value >70%) were nil, as shown in Table 1.

An item with DI = 1 indicates that the item perfectly discriminates between high achievers and low achievers. Discrimination index helps us in identifying the flaws in the item, providing improvement in options, misconceptions in learning so that learning can be improved.

In our study, it was observed that the mean score of discrimination index (*d* value) showed that 7 (28%) needs to be discarded (*d* value <0.15) and 5 (20%) of the items were acceptable (*d* value 0.15–0.24); hence, these items need to be revised, 4 (16%) were good (*d* value 0.25–0.35) and 9 (36%) were excellent (*d* value >0.35) as shown in Table 2.

A distractor is said to be functional only when it is attracted by at least 5% of the total response in the high achievers group and in the low achievers group. However, if more number of students from the lower group are answering an item correct when compared to higher group, we call it negative discrimination. The reasons are often due to an ambiguous question or an answer key that was wrongly marked. In the present study, analysis of distractor effectiveness (de), it was observed that 2 (8%) items were non-functional distractors, 3 (12%) items had negative discrimination and 20 (80%) were functional distractors as shown in Table 3. The non-functional distractor is an indicator which provides us an opportunity to replace it by a functional distractor.

It may also be possible that since there was no negative marking, there is a possibility for guessing the option.

6. Conclusion

Item analysis can tell us if an MCQ was too easy or too difficult and how it can be discriminated between high and low achievers and also whether all the distractors were effective. In the present study, we have observed that whatever items available in published books need to be post-validated in formative examinations and should be carefully selected in University and High stake examination by forming a question bank so that the assessment is valid and reliable to assess their cognitive domain and rank the students properly.

Conflicts of interest

The authors have none to declare.

References

1. Kemp JE. The multiple choice test: writing the questions. *Nurse Educ.* 1989;14: 10–12.

2. Gajjar S, Sharma R, Kumar P, Rana M. Item and test analysis to identify quality multiple choice questions (MCQs) from an assessment of medical students of Ahmadabad, Gujarat. *Indian J Community Med.* 2014;39:17–20.

3. Case SM, Swanson DB. Extended matching items: a practical alternative to free response questions. *Teach Learn Med.* 1993;5:107–115.

4. Verhoeven BH, Verwijnen GM, Scherpbier A, Schuwirth LW, van der V. Quality assurance in test construction: the approach of a multidisciplinary central test committee. *Educ Health.* 1999;12:49–60.

5. Downing SM. Assessment of knowledge with written test forms. In: Norman GR, Van der Vleuten CP, Newble DI, eds. In: *International Handbook of Research in Medical Education.* Dorcrecht, Great Britain: Kluwer Academic Publishers; 2002:647–672.

6. Schuwirth LW, van der Vleuten CP. Different written assessment methods: what can be said about their strength and weaknesses? *Med Educ.* 2004;38:974–979.

7. Issacs G. *Multiple Choice Testing: A Guide to the Writing of Multiple Choice Tests and to Their Analysis.* Campbell Town, NSW: Higher Education Research and Development Society of Australia; 1994.

8. Haladyna TM, Downing SM. How many options are enough for a multiple choice question test item? *Educ Psychol Meas.* 1993;53:999–1010.

9. Haladyna TM. *Developing and Validating Multiple Choice Test Items.* 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 1999.

10. Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty and discrimination indices of MCQs in formative exam in physiology. *South East Asian J Med Educ.* 2013;7:45–50.

12. Karelia BN, Pillai A, Vegada BN. The levels of difficulty and discrimination indices and relationship between them in four response type multiple choice questions of pharmacology summative tests of the year II MBBS students. *JSME.* 2013;6:41–46.

13. Patel KA, Mahajan NR. Itemized analysis of questions of multiple choice question (MCQ) exam. *Int J Sci Res.* 2013;2:279–280.

14. Mehta G, Mokhashi V. Item analysis of multiple choice questions – an assessment of the assessment tool. *Int J Health Sci Res.* 2014;4(7):197–202.

15. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and nonfunctioning distractors in multiple choice questions: a descriptive analysis. *BMC Med Educ.* 2009;9:40.

16. Natekar PE, De Souza F. Item analysis (post validation) of MCQs in anatomy for the first MBBS course. *Indian J Anat.* 2016;5(1):29–32.