

## Research Article

# Item analysis of multiple choice questions from an assessment of medical students in Bhubaneswar, India

Surya Kumar Namdeo\*, Bandya Sahoo

Department of Paediatrics, Kalinga Institute of Medical Science, Bhubaneswar, Odisha, India

**Received:** 18 April 2016

**Accepted:** 25 April 2016

### \*Correspondence:

Dr. Surya Kumar Namdeo,

E-mail: [drsuryanamdeo@gmail.com](mailto:drsuryanamdeo@gmail.com)

**Copyright:** © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## ABSTRACT

**Background:** Multiple choice questions (MCQs) are usually used to assess students in different educational streams. However, the MCQs to be used should be of quality which depends upon its difficulty index (DIF I), discrimination index (DI) and number of Non-functional distracter (NFD). Objective of the study is to evaluate the quality of MCQs, for creating a valid question bank for future use and to identify the low achievers, whose problems can be corrected by counselling or modifying learning methods. This study was done in Kalinga Institute of Medical Science (KIMS) Bhubaneswar.

**Methods:** A part completion test in the department of pediatrics was done. Total 25 MCQs and 75 distracters were analyzed. Item analysis was done for DIF I and DI and presence of number of NFD.

**Results:** Difficulty index of 14 (56%) items was in the acceptable range (p value 30-70%), 8 (32%) items were too easy (p value >70%) and 2 (8%) items were too difficult (p value <30%). Discrimination index of 12 (48%) items was excellent (d value >0.35), 3 (12%) items was good (d value 0.20-0.34) and 8 (32%) items were poor (d value <0.2%). Out of 75 distracters, 40 (53.4%) NFDs were present in 22 items. 3 (12%) items had no NFDs, whereas 8 (32%), 10 (40%), and 4 (16%) items contained 1, 2, and 3 NFD respectively.

**Conclusion:** Item analysis is a simple and feasible method of assessing valid MCQs in order to achieve the ultimate goal of medical education.

**Keywords:** Difficulty index, Discrimination index, Item analysis, Multiple choice questions, Non-functional distracter

## INTRODUCTION

The ultimate aim of medical education is to improve the health and the health care of the population.<sup>1</sup> The outcomes of all medical education programs, in general, are focused on this aim. So Assessments become necessary to measure accurately the students' progress towards achievement of this outcomes.<sup>1</sup> Test with multiple choice questions (MCQ) and analyzing their options have become the choice of many examiners in medical colleges.<sup>2</sup> Haladyna reviewed the validity of taxonomy of MCQ tests and wrote the guidelines for them.<sup>3</sup> Gajjar S examined the quality of MCQ tests and

emphasized that a good MCQ truly assess the knowledge and was able to differentiate the students of different abilities.<sup>4</sup> While Sharif M concluded that MCQ was an efficient tool for measuring the achievement of learners.<sup>5</sup> Even Vyas and Supe suggested that MCQs with 3 alternatives should be preferred than the 4 or 5 options.<sup>6</sup>

There are 3 components of a MCQ, direction (instruction to the students), stem (the question) and choices (alternatives). The correct alternative is called as answer and the other alternatives are called as distracters.<sup>7</sup> To assess the different domain it is important to have a good item. Item analysis is a process which assesses the quality

of those items and of the test as a whole.<sup>8</sup> It can tell us if an item or question was too easy or too difficult, how well it discriminated between high and low scores on the test and all of the alternatives functioned as intended or not.<sup>10</sup> The three numerical indicators of an item analysis are Item difficulty, Item discrimination and distracter analysis.

Despite the fact that preparation of a good item is very much essential to produce a valid MCQ hardly any attempt has been devoted to examine the contents of a test. So keeping this in view, present study has been undertaken with an objective to evaluate MCQs or items and develop a valid question bank for future use and also to identify the low achievers whose problems can be corrected by counseling or modifying learning.

**Settings:** The Study was conducted in Kalinga Institute of Medical Science (KIMS) Bhubaneswar.

## METHODS

A part completion test in the Department of Pediatric was conducted in December 2015 which was attended by 76 out of 100 MBBS students. The test comprised of 25 'Best response type' MCQs with 75 distracters. All MCQs collected from guide book, text book and peers had single stem with four options/responses. To avoid possible copying from neighboring student two invigilators were appointed with front and back camera in the examination room with a minimum distance of 3 feet between two students ahead, back and sideways.

## Data analysis

Data obtained was entered in MS Excel 2007 after taking informed consent from each student and analyzed. Score of 76 students was entered in descending order and whole group was divided in three groups. The group consisting of higher marks was considered as higher ability (H) and other group consisting of lower marks was considered as lower ability (L) group. Out of 76 students, 25 were in H group and 25 in L group; rests (26) were in middle group and not considered in the study.

Based on the data, various indices like difficult index (DIF I), discrimination index (DI) and Distracter analysis were calculated. DIF I describe the percentage of students who answered the item correctly and ranges between 0 and 100%.<sup>9</sup> It was calculated as  $P = (H+L/N) \times 100$ , where P was the item difficulty index, H was the number of students answering the item correctly in the higher ability group, L was the number of students answering the item correctly in the lower ability group and N was the total number of students. An item was considered difficult when the difficulty index value was less than 30% and considered easy when the index was more than 70% and the value between 30-70% was acceptable (between 50-60% are ideal).<sup>10</sup>

The item discrimination index (DI) is the ability of an item to differentiate between students of higher and lower abilities and ranges between 0 and 1.4 It was calculated using the formula  $d = (H-L/N) \times 2$ . Items with a discrimination index between 0.25-0.35 were considered good; those with indices more than 0.35 were excellent, between 0.20-0.24 were acceptable and below 0.20 were poor.<sup>10</sup>

An item contains four options including one correct (key) and three incorrect (distracter) alternatives. Non-Functional distracter (NFD) is an option (s) selected by <5% of students; alternatively functional or effective distracters are those selected by 5% or more participants.<sup>2,8,17</sup> Items were categorized as poor, good or excellent and actions such as discard/revise and store were proposed based on the values of DIF I, DI and distracter analysis.

## RESULTS

Total 25 MCQs and 75 distracters were analyzed. Means and standard deviations (SD) for DIF I (%) and DI were  $65.92 \pm 22.2\%$  and  $0.33 \pm 0.23$  respectively (Table 1).

**Table 1: Assessment of 25 items based on various indices among 76 students.**

Parameter	Mean	Standard deviation
Difficult index	65.92%	22.2
Discriminating index	0.33	0.23

**Table 2: Distribution of item in relation to difficult index.**

Difficult index	Number of items	Percentage	Interpretation
50-60%	1	4%	Good to excellent
30-70%	14	56%	Acceptable
>70%	8	32%	Too easy, Require modification
<30%	2	8%	Too difficult, Require modification

Out of 25 items, one had 'good to excellent' level of difficulty (DIF I = 50-60%) whereas 14 items (56%) were within the range of acceptable DIF I (DIF I = 30 -70%) and 10 (40%) items which were either too easy or too difficult (Table 2).

14 items (56%) had good to excellent discrimination power (DI  $\geq 0.35$ ) whereas 8 items (32%) kept for revision. (Table 3) When these two were considered together, there were 17 (68%) items as ideally acceptable which were included in question bank for future use.

**Table 3: Distribution of item in relation to discriminating index.**

Discriminating index	Number of items	Percentage	Interpretation
>0.35	12	48%	Excellent
0.34-0.25	3	12%	Good
0.24-0.20	2	8%	Acceptable
<0.20	8	32%	Require modification

**Table 4: Distracter analysis.**

Distracter analysis	
No. of items	25
No. of total distracter	75
Functional distracter	35(46.6%)
Non- Functional distracter	40(53.4%)

**Table 5: Frequency distribution of non-functional distracters (NFD) according to selection.**

Number of items with NFD	Frequency	Percentage	Cumulative
0 NFD	3	12%	12 %
1 NFD	8	32%	44 %
2 NFD	10	40%	84%
3 NFD	4	16%	100%

Out of 75 distracters, 40 (53.4%) NFDs were present in 22 items (Table 4). 3 (12%) items had no NFDs whereas 8(32%), 10(40%), and 4 (16%) items contained 1, 2, and 3 NFD respectively (Table 5). That means 32% of the questions were three-choices, 10% were two-choices and 16% were one-choice questions, respectively.

## DISCUSSION

The assessment tool of any examination should be designed according to the objective. If properly designed One-best MCQs are one of the best assessment tool that quickly assess any level of cognition according to Bloom's taxonomy.<sup>11</sup>

DIF I in our study was 65.92±22.2% with 56% items in acceptable range (p 30-70%), 32% items very easy (p>70%) and 8% items very difficult (p<30%). Karelia, Pillai & Vegada showed a range of mean±SD between 47.17±19.77 to 58.08±19.33 in their study.<sup>12</sup> They showed 61% items in acceptable range (p 30-70%), 24% items (p>70%) and 15 % items (p<30%). Singh JP found in their study Difficulty index of 11 (55%) items were in the acceptable range (p value 30-70%), 9 (45%) items were too easy (p value >70%) and no any items were too difficult (p value <30%).<sup>15</sup> Gajjar S showed Means and standard deviations (SD) for DIF I (%) were 39.4±21.4%, respectively.<sup>4</sup> Mehta G also showed similar p value of 31

(62%) items were in the acceptable range (30-70%), 16 (32%) items >70% and 3 (6%) items <30%.<sup>13</sup>

DI is an index which differentiates high ability and low abilities student. It is obvious that a question which is either too difficult (attempted wrongly by everyone) or too easy (response correctly by everyone) will have nil to poor DI. Mean DI in present study was 0.33±0.23 though it was not an excellent DI (>0.35) but it was good (0.24-0.34) and acceptable. Gajjar S, reported the items in his study had a very low discrimination index (DI) with mean DI of 0.14±0.19.4 In another study 46% of the 20 MCQ items had a discrimination index of >0.35 (Excellent items), 22% items had a discrimination index between 0.25-0.35 (Good items), 10% items had a discrimination index between 0.20-0.24 (acceptable items), while 22% items had a discrimination index of <0.20 (Poor items).<sup>12</sup> In an earlier study done by Mehta G, the mean of DI was 0.33±0.18. Items with DI >0.35 were 26 (52%), DI between 0.2 and 0.34 were 9 (18%) and DI<0.2 were 15 (30%), 13 while study done by Singh J P show, the items with DI >0.35 were 10 (50%), DI between 0.2 and 0.34 were 4 (20%) and DI <0.2 were 6 (30%).<sup>14</sup>

Designing of plausible distracters and reducing the NFDs is important aspect for framing quality MCQs. Presence or absence of NFDs in an item also affect the discriminative power. More NFD in an item makes it easy and conversely item with more functioning distracters makes the item difficult. In Our study among 75 distracters, 40 (53.4%) NFDs and 35 (46.4%) FDs were present. 12% of the all the distracters were sufficiently attractive to be selected whereas 32% had one, 40% had two and 16% had three nonelected distracters. In another study done by Sharif et al showed 34.6%, 38.1%, 15.3% of items had one, two and three NFDs respectively, Whereas 12% items had no NFDs as similar to our study.<sup>5</sup>

Items analyzed in our study were neither too easy nor too difficult (mean DIF I = 65.92%) which was excellent but the overall DI was good (mean DI 0.33). Therefore, items were acceptably difficult and good at differentiating higher and lower ability students. Those items which provide good index of discrimination & difficult index with all functioning alternatives should be retained and placed in a question bank for further use.<sup>10</sup> So in our study 17 items were good (DIF I 30-70% and DI > 0.25) which were retained in question bank and rest 8 items were revised. Most of the distracters (total 40) present in 22 items were not good distracters and were modified.

## CONCLUSION

Item analysis is a valuable procedure performed after the examination which provides information regarding the reliability and validity of an item or test.<sup>15</sup> It aids in detecting specific technical flaws and thus provides information for improving test item.<sup>10</sup>

Thus we conclude analysis of items strengthen the future question bank. Also discussion of the analysis's result with the faculties helps in modification of teaching methodology and outcome of learning. Therefore item analysis is a simple and feasible method of assessing valid MCQs in order to achieve the ultimate goal of medical education.

*Funding: No funding sources*

*Conflict of interest: None declared*

*Ethical approval: Not required*

## REFERENCES

1. Chandratilake M, Davis M, Ponnampereuma G. Evaluating and designing assessments for medical education: the utility formula. *Intern J Med Edu.* 2009;1:1-7.
2. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distracter efficiency. *J Pak Med Assoc.* 2012;62:142-6.
3. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item- writing guidelines for classroom assessment. *Appl Measurement Edu.* 2002;15(3):309-34.
4. 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 Ahmedabad, Gujarat. *Indian J Comm Med.* 2014;39(1):17-20.
5. Sharif MR, Rahimi SM, Rajabi M, Sayyah M. Computer software application in item analysis of exams in a college of medicine. *J Sci Tech.* 2014;4(10):565-9.
6. Vyas R, Supe A. Multiple choice questions: a literature review on the optimal number of options. *The National J India.* 2008;21(3):130-3.
7. Kehoe J. Writing multiple-choice test items. *Pract Assess, Res Evaluation.* 2008;4(9).
8. Singh T, Gupta P, Singh D. Test and item analysis. *Principles of Medical Education.* 3rd ed. New Delhi, Jaypee Brothers Medical Publishers (P) Ltd; 2009: 70-77.
9. 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;9:1-8.
10. Manual of basic workshop in medical education technologies cuttack: regional training centre. SCB medical college cuttack. 2015;63-5.
11. Bloom S, Hastings T, Modays J. Handbook of formative and summative evaluation of students learning. New York, McGraw Hill; 1971:103.
12. Karelia BN, Pillai A, Vegada BN. The levels of difficulty and discrimination indices and relationship between them in four responses type multiple choice questions of pharmacology summative tests of year II MBBS students. *IEJSME.* 2013;6:41-6.
13. Singh JP, Kariwal P, Gupta SB, Shrotriya VP. Improving multiple choice questions through item analysis: an assessment of the assessment tool. *Int J Basic Sci Appl Res.* 2014;1(2):53-7.
14. Mehta G, Mokhasi V. Item analysis of multiple choice questions: an assessment of the assessment tool. *Int J Health Sci Res.* 2014;4(7):197-202.
15. Botti MJC, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. *Collegian.* 2005;12: 19-24.

**Cite this article as:** Namdeo SK, Sahoo S. Item analysis of multiple choice questions from an assessment of medical students in Bhubaneswar, India. *Int J Res Med Sci* 2016;4:1716-9.