

DIFFICULTY AND DISCRIMINATION ANALYSIS OF END OF TERM MULTIPLE-CHOICE QUESTIONS AT COMMUNITY MEDICINE DEPARTMENT, WAH MEDICAL COLLEGE

Musarat Ramzan, Khola Waheed Khan, Saana Bibi, Shezadi Sabah Imran

Wah Medical College, Wah/National University of Medical Sciences (NUMS) Pakistan

ABSTRACT

Objective: To perform post analysis of multiple-choice questions given in the 2nd term and send up examinations of the years 2016 to 2018, to establish relationship between difficulty (DF) and discrimination indices (DI) and to find out significant mean difference between the two.

Study Design: Cross sectional study.

Place and Duration of Study: Community Medicine Department, Wah Medical College, Wah, from Nov 2018 to Mar 2019.

Methodology: A total of 390 Multiple-Choice Question of second term and send-up were taken for the study from the year 2016, 2017 and 2018. The response sheets were assessed by Optical Machine Reader (OMR) and the level of difficulty, power of discrimination and reliability were obtained. The data was entered in SPSS version 22.

Results: A total of 315 test items were included. Results of the study showed that the reliability (KR20) for all the examined items was in the acceptable range i.e. ≥ 0.7 and there was no association was found between difficulty index and year $p=0.310$. The mean difficulty index was found to be 0.48 ± 0.22 and discrimination index as 0.24 ± 0.14 .

Conclusion: The analysis of 390 test items showed that most of the questions were acceptable in terms of difficulty and discrimination. There is still a need to modify and improve the testing ability of the MCQs with negative discrimination and higher difficulty index.

Keywords: Assessment, Difficulty index, Discrimination index, Multiple choice questions.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Assessment is an integral part of student evaluation that can give an insight about their learning and competencies. For assessment a valid, reliable and objective tool should be used¹. A good tool can assess cognitive, affective as well as psychomotor domains². This is also a need of the hour to transit from using tools that test only rote memory to those that inculcate problem solving³.

The MCQs as a tool of assessment have gained popularity in this regard. On one hand the MCQs can assess a many part of the curriculum in short time, on the other hand they have more objectivity, comparability and minimum bias^{1,2}.

It is also a fact that a good quality MCQ that can assess higher order cognitive process needs a lot of time and effort to construct^{1,4}. It is therefore essential that MCQS be handled very carefully from its construction till its interpretative stage. Once constructed, MCQs should be trial tested and the scores generated can be further subjected to statistical item analysis⁵.

Item analysis is a procedure of collecting and summarizing students' responses to individual test items for the assessment of the quality of test items⁶. It

is a technique that assesses the MCQ tool validity and reliability. The students' performance is analyzed to determine the item fate, whether it should be kept, discarded or reviewed¹.

Many statistic and metric tools can be used for item analysis e.g. Difficulty index (DIF), discrimination index (DI), reliability of test and score distribution^{2,7}.

The DIF or ease index designates the proportion of students who answered the items correctly. The recommended range is 30-70%⁸. The DI (point biserial correlation) defines the ability of an item to distinguish between students with high scores from students with low scores with a view that a high performing student selects the correct answer for each item more often than the low performing students⁴. The recommended range of DI is 0.25-0.48.

The department of community medicine strives to enhance the relevance of assessment given to the students during the years of training in order to strengthen the MCQ banks and raise the standard of assessment. Time is a commodity which requires to be spent sagaciously. The perpetuity of the process of looking into the assessments indicates designing any in-home test with in shortest possible time, effectively.

The current study was carried out to perform post analysis of MCQs items given in the 2nd term and send up examinations of the years 2016 to 2018, to establish

Correspondence: Dr Musarat Ramzan, Department of Community Medicine, Wah Medical College, Wah Pakistan

Received: 30 Jul 2019; revised received: 20 Jan 2020; accepted: 24 Jan 2020

relationship between difficulty (DF) and discrimination indices (DI) and find out significant mean difference across three years between the two. The rationale of the study was to furnish information on the fitness of the items to be retained in the question bank of the department.

METHODOLOGY

The cross-sectional study was carried out in the Community Medicine department Wah Medical College, Wah, from November 2018 to March 2019 after approval from Dean of the college. A total of 130 MCQ items each of second term and send-up were taken for the study from the year 2016, 2017 and 2018. This makes a total of 390 test items (convenient sampling). During the study period 315 students solved these MCQ items.

MCQ test items used in all six end of term examinations were of one best type. The MCQs were formed by different faculty members of the department according to the topic they taught. These MCQs were than pre hoced in the department. Finally, the head of department vetted these questions for the content, semantics and construct validity. The responses were taken on a specific response sheet and were assessed by optical machine reader (OMR). The data on difficulty indices (*p*-value), power of discrimination and reliability were obtained by OMR. Further the data was entered in SPSS-22 and mean and standard deviation were calculated for difficulty and discrimination indices.

MCQ items that lie between 30-70% were considered acceptable for difficulty while the acceptable range for DI is 0.25-0.48. Chi square test was applied on difficulty and discrimination indices to find significant mean difference across three years. The *p*<0.05 was considered significant.

RESULTS

A total of 315 test items were included. Results of the study showed that the reliability (KR20) for all the examined items was in the acceptable range i.e. ≥ 0.7 and there was no association was found between diffi-

culty index and year *p*=0.310 as shown in table-I.

In year 2017, there was 77 (59%) frequency 0.2-0.4 followed by 8 (6.15) negative frequency in 2018, there was significant association of discrimination index and Year, *p*=0.032 shown in table-II.

DISCUSSION

The assessment of the content taught and desired outcome of the student is an essential component of the evaluation of effective curriculum delivery. In addition, an assessment either formative or summative affects learning. An appropriate assessment tool serves the purpose. From all the assessment tools, MCQs are the best because they evaluate any level of knowledge domain^{9,10}.

In our study of post hoc analysis of the last three years end of term MCQs items the majority of items were of the acceptable difficulty i.e. 58.6%. In India similar researches were carried out in which the difficult questions comprised of 85% and 65% of the test items respectively^{11,12}. The item difficulty determines that whether the students have learned the concept being tested. It is also important that those questions that are less difficult are not useless they can be use in the beginning of the paper as warm up questions and very difficult questions can be revisited for their quality and content coverage¹³.

Of all the MCQs majority of the items were in acceptable range for discrimination i.e. 49.33%. However, the number of MCQs with negative and poor discrimination was high in 2016 as compared to other two years. The research carried out by Mukherjee *et al*, also showed that the mean discrimination index was 0.31 ± 0.27 meaning that most of the questions were reasonably good¹⁴.

These results are also consistent with a study carried out in India in which 50% items were in acceptable range of difficulty and almost all questions are in acceptable range of discrimination¹⁵.

The reliability (KR20) measured for all the examined items was in the acceptable range i.e. ≥ 0.7. A

Table-I: Association of difficulty index and year

Years	<30, n (%)	30-70, n (%)	>70, n (%)	<i>p</i> -value
2016	36 (27%)	73 (56%)	21 (16%)	0.310
2017	29 (22%)	76 (58%)	25 (19%)	
2018	23 (17%)	81 (62%)	26 (20%)	

Table-II: Association of discrimination index and Year.

Year	Negative, n (%)	0.01-0.19, n (%)	0.2-0.4, n (%)	≥0.4, n (%)	<i>p</i> -value
2016	10(7.69%)	43(33%)	60(45%)	17(13%)	0.032
2017	8(6.15%)	27(20%)	77(59%)	17(13%)	
2018	8(6.15%)	34(26%)	58(44%)	30(23%)	

study in west Bengal India showed the KR20 to be 0.914. However, a research carried out by Donald S Christian the KR (20) ranged from 0.29-0.5216.

The curve estimation analysis between difficulty and discrimination index for all the three years separately showed a linear relationship and the Pearson correlation showed highly significant association between the two. However, a study carried out by Si-Mui-Sim found out that the relationship between discrimination and difficulty index was not linear but dome-shaped. Initially discrimination power increases with increase in difficulty until a plateau is reached (at DI of 51-70% and DIF of 40-74%) and then began to decline with further increase in difficulty¹⁷. A study carried out by Karkal *et al*, in India on item analysis of MCQs found out a negative correlation between DF and DI without statistical significance¹⁸.

The items going below x-axis indicate negative or poor discrimination with high level of difficulty but they are less in number in the year 2018 showing a gradual improvement in the item quality. The items with lower values should be scrutinized again to look into any flaws in their construction, distractors or key¹³.

The result showed that MCQs are a good assessment tool for formative as well as summative assessment. These MCQs should be of average difficulty and high discrimination index. When such MCQs are used for assessment, not only the students' cognition but also the teacher's ability to design a good and relevant assessment can be checked. Assessment and in turn teaching can be made more valid by incorporating post hoc analysis¹⁹.

CONCLUSION

Item analysis is a valuable tool for quality assurance that leads to the strengthening of MCQ bank. The analysis of items showed that a good percentage of questions were in acceptable range of difficulty and discrimination. There is still a need to modify and improve the testing ability of the MCQs with negative DI and higher DIF indices.

CONFLICT OF INTEREST

This study has no conflict of interest to be declared by any author.

REFERENCES

1. Kheyami D, Jaradat A, Al-Shibani T, Ali FA. Item analysis of multiple choice questions at the department of paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos Univ Med J* 2018; 18(1): 68-74.
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 Ahmedabad, Gujarat. *Ind J Commun Med* 2014; 39(1): 17-20.
3. Park IS, Suh YO, Park HS, Kang SY, Kim KS, Kim GH, et. al. Item development process and analysis of 50 case-based items for implementation on the Korean Nursing Licensing Examination. *J Educ Eval Health Prof* 2017; 14(1): 20-25.
4. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *J Pak Med Assoc* 2012; 62(2): 142-47.
5. Odukoya JA, Adekeye O, Igbinoaba AO, Afolabi A. Item analysis of university-wide multiple choice objective examinations: the experience of a Nigerian private university. *Qual Quant* 2018; 52(3): 983-97.
6. Quaigrain K, Arhin AK. Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Educ* 2017; 4(1): 1301013.
7. Hermi A, Achour W. Item analysis of examinations in the faculty of medicine of tunis. *Tunis Med* 2016; 94(4): 247-52.
8. Sahoo D, Singh R. Item and distracter analysis of multiple-choice questions (MCQs) from a preliminary examination of undergraduate medical students. *Int J Res Med Sci* 2017; 5(12): 5351-55.
9. 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(5): 1716-19.
10. Patil R, Palve SB, Vell K, Boratne AV. Evaluation of multiple-choice questions by item analysis in a medical college at Pondicherry, India. *Int J Community Med Public Health* 2017; 3(6): 1612-16.
11. Rao C, Prasad HK, Sajitha K, Permi H, Shetty J. Item analysis of multiple-choice questions: Assessing an assessment tool in medical students. *Int J Educ Psychol Res* 2016; 2(4): 201-204.
12. Kolte V. Item analysis of multiple-choice questions in physiology examination. *Ind J Basic Applied Med Res* 2015; 4(4): 320-26.
13. Patel RM. Use of Item analysis to improve quality of Multiple-Choice Questions in II MBBS. *J Educ Technol Health Sci* 2017; 4(1): 22-29.
14. Mukherjee P, Lahiri SK. Analysis of multiple-choice questions: Item and test statistics from an assessment in a medical college of Kolkata, West Bengal. *IOSR J Dent Med Sci* 2015; 14(12): 47-52.
15. Khan GA, Ishrat N, Khan AQ. Using item analysis on essay type's questions given in summative examination of medical college students: facility value, discrimination index. *Int J Res Med Sci* 2015; 3(1): 178-82.
16. Christian DS, Prajapati AC, Rana BM, Dave VR. Evaluation of multiple-choice questions using item analysis tool: a study from a medical institute of Ahmedabad, Gujarat. *Int J Com Med Public Health* 2017; 4(6): 1876-81.
17. 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 Singap* 2006; 35(2): 67-71.
18. Karkal YR. Item analysis of multiple-choice questions of undergraduate pharmacology examinations in an international medical school in India. *J NTR Univ Health Sci* 2016; 5(3): 183-86.
19. Panchal P, Prasad B, Kumari S. Multiple choice questions - role in assessment of competency of knowledge in anatomy. *Int J Anat Res* 2018; 6(2,1): 5156-62.