

ORIGINAL ARTICLE

Item Analysis of Multiple-Choice Questions at the Department of Community Medicine, Wah Medical College, Pakistan

Musarat Ramzan¹, Shezadi Sabah Imran², Saana Bibi³, Khola Waheed Khan⁴, Imrana Maqsood⁵

ABSTRACT

Objective: The objective of the study was to assess the quality of multiple-choice questions (MCQs) of three different assessments in the subject of Community Medicine by computing the difficulty index, discrimination index and reliability and to estimate the relationship between difficulty and discrimination indices.

Study Design: Retrospective observational study.

Place and Duration of Study: Department of Community Medicine at Wah Medical College from August to December 2018.

Materials and Methods: Three sets of MCQs were included in the study. Mean and standard deviation of difficulty and discrimination indices were calculated and one-way analysis of variance and Kruskal Wallis test were applied on difficulty and discrimination indices. The association was determined by Pearson correlation and considered significant at p value of < 0.05 .

Results: The mean difficulty index of first term, second term and send-up examination were 41.5, 48.8 and 51.9 respectively. Mean discrimination indices were 0.28, 0.27 and 0.26 and reliability were 0.83, 0.81 and 0.79. In the study, 72% MCQs of the first term, 61.5 % of the second term and 63% of the send-up examinations were in the range 30-70% of difficulty. There was a significant difference in the difficulty indices of the three question papers. The correlation between discrimination and difficulty indices was curvilinear and positively correlated.

Conclusion: It is concluded that all three question papers have acceptable reliability, more than 65% MCQs have acceptable difficulty index and about 69% have good discriminatory power.

Key Words: Assessment, Difficulty Index, Discrimination Index, Multiple Choice Questions, Reliability.

How to cite this: Ramzan M, Imran SS, Bibi S, Khan KW, Maqsood I. Item Analysis of Multiple-Choice Questions at the Department of Community Medicine, Wah Medical College, Pakistan. *Life and Science*. 2020; 1(2): 60-63. doi: <http://doi.org/10.37185/LnS.1.1.53>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license.

(<https://creativecommons.org/licenses/by-nc/4.0/>). Non-commercial uses of the work are permitted, provided the original work is properly cited.

Introduction

Assessment is a crucial component of both learning and teaching. Continuous improvement in assessment tools has significant impact on the quality of learning and it also enhances the capability

of learners.¹ Written assessment can be done by number of tools including multiple-choice questions (MCQs). A type of MCQs are the most common method of assessment in medical colleges.² The A type MCQ provides four to five options, and students are supposed to select the best answer. These questions not only assess memory and understanding but also higher levels of learning i.e application, analysis and evaluation of problems. Thus, many aspects of the subject can be examined with MCQs of high reliability in a short time period.⁴ The answering and scoring of MCQs is time effective and logistically efficient.⁵ However construction of a flawless MCQ needs time and expertise.^{6,7}

The quality control of items is an important factor to improve the quality of assessments and item analysis is a process which assesses not only the quality of items but the entire assessment simultaneously. It

^{1,2,4}Department of Community Medicine/ Medical Education³
Wah Medical College

National University of Medical Sciences, Rawalpindi
Department of Pharmacology

⁵Fazaila Medical College

Air University, Islamabad

Correspondence:

Dr. Shezadi Sabah Imran

Associate Professor, Community Medicine
Wah Medical College

National University of Medical Sciences, Rawalpindi

Email: sabah_imran00@yahoo.com

Funding Source: NIL; Conflict of Interest: NIL

Received: July 11, 2019; Revised: Feb 24, 2020

Accepted: Feb 05, 2020

analyzes student responses to individual test items, and it is very helpful in recognizing redundant items thus developing a pool of valid items.^{6,8}

In item analysis many indices can be calculated like difficulty index, discrimination index, reliability of the test, distractor functionality and test score distribution. Difficulty index also called *p value* depicts the proportion of students who get the item right; a high percentage indicates easy and low percentage indicates difficult items. The recommended range is 30-70% and the items outside this range have very little power of discrimination. Discrimination index distinguishes between high and low achievers. Its value ranges from 0 to 1. Positive discrimination index means high performers choose the right answer of each item more frequently than the low achievers who achieve lower overall score. Negative value means low performing students get the item right. This may be due to vague stems or an error in the answer key.^{1,6,8} Difficulty and discrimination indices have positive association except when the difficulty index is either too high or too low.^{2,9} Reliability is also a measurement of internal consistency of assessment and it ranges from 0 to 1, the higher the value the more is the reliability.^{1,3} The Optical Machine Reader (OMR) uses software programs to find out the item statistics including *p value*, point biserial, reliability and standard error.

The objective of this study was to assess the quality of MCQs of three different assessments of Community Medicine by computing the difficulty index, discrimination index and reliability and to estimate the relationship between difficulty and discrimination indices. The results of such analysis identify the items that can be retained or modified, revalidated or discarded. This helps in the development of a valid question bank for assessment in the subject of Community Medicine.

Materials and Methods

The study was carried out at the Department of Community Medicine at Wah Medical College from August to December 2018. It was a retrospective observational study and three sets of MCQs in Community Medicine were included in the study. Each question paper comprised of 65 MCQs, one best type with one correct answer (key) and four distracters. A true response to an item was awarded

one mark and zero was given to the wrong answer. The response sheets were assessed by an OMR which calculated the difficulty index, discrimination index of each question and reliability of three sets of MCQs. SPSS version 22 was used for further analysis. Mean and standard deviation of difficulty and discrimination indices were calculated and one-way analysis of variance (ANOVA) was applied on difficulty indices after assessing the normality of distribution using Kolmogorov-Smirnov test and Kruskal-Wallis test on discrimination indices of the three examinations. Furthermore, post-hoc analysis using Tukey HSD test was also done to determine which question paper was different from others regarding difficulty index. The relationship between difficulty and discrimination indices of 195 items was determined by Pearson correlation and it was considered significant at *p value* of < 0.05. The items with index less than 30% were considered difficult and more than 70% were regarded easy. The discrimination index of more than 0.3 considered as excellent discrimination, 0.2-0.29 as good. 0-0.19 as poor and negative value as defective MCQ.

Results

There were 65 MCQs in all three sets of questions. The reliability of first term was 0.83, second term was 0.81 and send-up examination was 0.79. Mean with standard deviation and ranges of difficulty and discrimination indices regarding first term, second term and send-up examinations are shown in the Table 1.

Table 1: Descriptive analysis of difficulty index (P) and discrimination index

| Item analyses parameters | 1 st term (n=94) | | 2 nd Term (n=94) | | Send Up (n=100) | |
|-----------------------------|-----------------------------|---------------|-----------------------------|---------------|-----------------|---------------|
| | Mean± SD | Range | Mean± SD | Range | Mean ± SD | Range |
| Difficulty index | 41.5 ± 17.7 | 7 to 84 | 48.8 ± 21.0 | 2 to 94 | 51.9± 20.5 | 14 to 95 |
| Discrimination index | 0.28 ± 0.19 | -0.21 to 0.59 | 0.27 ± 0.18 | -0.11 to 0.58 | 0.26± 0.14 | -0.13 to 0.58 |

The percentages of easy, difficult and acceptable items are presented in Figure 1 and the percentages of items having poor, acceptable and excellent discrimination are illustrated in Figure 2.

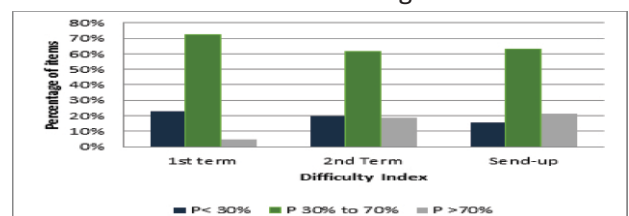


Fig 1: Difficulty index of MCQ papers of first term, second term and sendup assessments

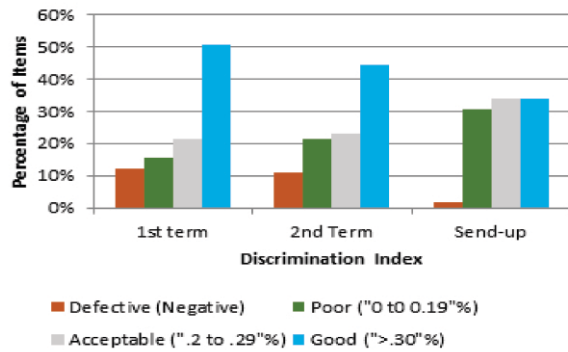


Fig 2: Discrimination index of MCQ papers of first term, second term and sendup assessments

The mean difficulty indices of the three internal examinations were significantly different from each other and post hoc analysis showed that the send-up examination was significantly different from the first term examination (Table 2) while mean discrimination indices of the three examinations were not significant (*p value*=0.181) after applying Kruskal-Wallis test.

Table 2: Mean comparison of difficulty index across three internal examinations

| Measures | | Mean | Standard Deviation | 95% CI for Mean | | <i>P value</i> |
|------------------|-----------------------------|-------|--------------------|-----------------|-------------|----------------|
| | | | | Lower Bound | Upper Bound | |
| Difficulty index | 1 st Term (n=65) | 41.55 | 17.64 | 37.18 | 45.93 | .010* |
| | 2 nd Term (n=65) | 48.78 | 20.92 | 43.60 | 53.97 | |
| | Send-up (n=65)** | 51.91 | 20.56 | 46.81 | 57.00 | |

**P*<0.05 was considered significant using one-way ANOVA.

**Send-up results found significantly different from 1st term results by using Tukey HSD test.

The scatter diagram showed curvilinear relationship between difficulty and discrimination indices of 195 items was made. The discrimination index correlated positively with the difficulty index (*r*= 0.438 *p*<0.001) and maximum discrimination i.e more than 0.4 was corresponding with the items having difficulty index of 30 to 78 percent (Figure 3).

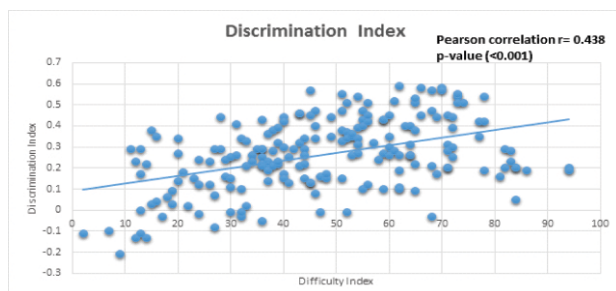


Fig 3: Correlation between difficulty and discrimination indices of three assessments

Discussion

MCQs are a valuable and efficient tool to assess student cognition however, their effectiveness depends on the quality of MCQs which is best appraised by item analysis. We aimed to perform item analysis of three question papers of our Department.

The mean difficulty indices of first term, second term and send-up examination were 41.5, 48.8 and 51.9, respectively. This showed that all three assessments had questions with moderate difficulty. The results were comparable with the studies carried out by Mozaffal et al⁹ and Sushua et al¹⁰ but contradictory to Joao Pais et al¹¹ and Sanju Gajjar et al.²

The mean discrimination indices of the three examinations were 0.28, 0.27 and 0.26 showing that all three assessments had good discrimination. The results were similar to the studies conducted by Sushua et al¹⁰ and Pais et al¹¹ but dissimilar with the Gajjar et al² and Sook Park et al.¹²

In the current study, 72% MCQs of the first term, 61.5% of the second term and 63% of the send-up examinations were in the acceptable range of difficulty. The rest of the MCQs were either very easy or very difficult needing modification or substitution. This result was almost similar with the prior study of Manish et al.¹³ Regarding discrimination index, 72% MCQs of the first term and 67.6% MCQs of the second term and send-up examination respectively were in the acceptable range and had a tendency to discriminate between the high and low achievers.¹⁴ As 69% of total MCQs had acceptable discriminatory power therefore these three papers were considered to be good question papers.¹⁵ The MCQs having poor or negative discrimination were either very easy or very difficult¹⁶ and required to be reviewed by experts for identifying possible reasons of poor discrimination like ambiguous questions, wrong key, implausible distracters and need to be reconstructed as valid items. The reliability of all three assessments was good and acceptable as they were low stakes examinations.

The difficulty index of the send-up examination was significantly different from other assessments and this might be because of better preparation of students for send-up examinations. MCQs having difficulty index between 30 to 78 percentage had more discriminatory power as compared to MCQs

that were not within this range. Similar results have been observed in different studies. Sim et al showed non-linear relation between both the indices of the three examinations of year two medical students.¹⁷ Deena et al showed significant dome shaped correlation ($r=0.162$) between difficulty and discrimination indices.¹⁸ Mitra et al showed significant negative correlation between difficulty and discrimination indices (-0.325).¹⁹ This means that with increasing difficulty index, there was decrease in discrimination index. Habib et al showed negative correlation (-0.453) between the two indices.²⁰

Recommendations

This study emphasizes the need of modification and reconstruction of MCQs having less discriminatory ability. The items having negative discrimination index should be evaluated to determine flaws either in the vignette, lead in or in the key. Based on the results of this study it is recommended that post hoc analysis of all examinations should be performed to increase the validity and authenticity of question bank for future use. This analysis would also help the educators to identify their inadequacy in making assessments.

Conclusion

It is concluded that all three question papers have acceptable reliability which were 0.83, 0.81 and 0.79, respectively. More than 65% MCQs have an acceptable difficulty index and about 69% have good discriminatory power. The items which do not meet the acceptable criteria may be altered or discarded.

REFERENCES

1. Islam Zu, Usmani A. Psychometric analysis of Anatomy MCQs in Modular examination. *Pak J Med Sci.* 2017; 33: 1138-43.
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. *Indian J Community Med.* 2014; 39: 17-20.
3. Rahma NAA, Shamad MMA, Idris MEA, Elfaki OA, Elfakey WEM, Salih KMA. Comparison in the quality of distractors in three and four options type of multiple choice questions. *Adv Med Educ Prac.* 2017; 8: 287-91.
4. Ramakrishnan M, Sathe AB, Vinayak A. Item analysis: a tool to increase MCQ validity. *Indian J Basic Appl Med Res.* 2017; 6: 67-71.
5. Abdulghani HM, Ahmad F, Irshad M, Khalil MS, Al-Shaikh GK, Syed S, et al. Faculty development programs improve the quality of Multiple Choice Questions items' writing. *Sci Rep.* 2015; 5: 9556.
6. Kaur M, Singla S, Mahajan R. Item analysis of in use multiple choice questions in pharmacology. *Int JAppl Basic Med Res.* 2016; 6: 170-3.
7. Rush BR, Rankin DC, White BJ. The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. *BMC Med Edu.* 2016; 16: 250.
8. Namdeo SK, Sahoo B. Item analysis of multiple choice questions from an assessment of medical students in Bhubaneswar, India. *International Journal of Research in Medical Sciences.* 2016; 4: 1716-9.
9. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *J Pak Med Assoc.* 2012; 62: 142-7.
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 Journal of Medical Education.* 2013; 7: 45-50.
11. Pais J, Silva A, Guimarães B, Povo A, Coelho E, Silva-Pereira F, et al. Do item-writing flaws reduce examinations psychometric quality? *BMC Res Notes.* 2016; 9: 399.
12. 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: 20.
13. Badkur M, Suryavanshi G, Abraham AK. The correlation between the acceptable range of difficulty and discrimination indices in four response type multiple choice questions in physiology. *Indian Journal of Basics and Applied Medical Research.* 2017; 6: 695-700.
14. Tejinder S, Piyush G, Daljit S. Principles of medical education. 3rd ed. New Delhi: Jaypee Brothers. 2013. p. 70-1.
15. Hermi A, Achour W. Item analysis of examinations in the Faculty of Medicine of Tunis. *Tunis Med.* 2016; 94: 247-52.
16. Borwn FG. Principles of educational and psychological testing. 3rd ed. New York: Rinehart and Winston. 1983. p. 55-6.
17. Sim SM, RasiyahRI. 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.
18. 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: 68-74.
19. Mitra NK, Nagaraja HS, Ponnudurai G, Judson JP. The levels of difficulty and discrimination indices in type a multiplechoice questions of pre-clinical semester 1, multidisciplinary summative tests. *IJSME.* 2009; 3: 2-7.
20. Habib MA, Talukder HK, Rahman MM, Ferdousi S. Post-application Quality Analysis of MCQs of Preclinical Examination Using Item Analysis. *Bangladesh Journal of Medical Education.* 2016; 7: 2-7.