Review Article

Refine the multiple choice questions tool with item analysis

M. Senthil Velou[1*], E. Ahila[2]

[1]Assistant Professor, Department of Physiology, AIIMS, Mangalagiri, Andhra Pradesh, India
[2]Senior Lecturer, Department of Periodontology, Sri Venkateshwaraa Dental College, Ariyur, Puducherry, India
[*]Corresponding author email: **senthil.velou@aiimsmangalgiri.edu.in**

International Archives of Integrated Medicine, Vol. 7, Issue 8, August, 2020.
Available online at http://iaimjournal.com/
ISSN: 2394-0026 (P) **ISSN: 2394-0034 (O)**
Received on: 17-07-2020 **Accepted on:** 26-07-2020
Source of support: Nil **Conflict of interest:** None declared.
How to cite this article: M. Senthil Velou, E. Ahila. Refine the multiple choice questions tool with item analysis. IAIM, 2020; 7(8): 80-85.

Abstract

Assessment facilitates and improves learning outcomes. Multiple choice questions (MCQs) is one of the important tools in the assessment of learning outcomes and used most often due to its reliability and simplicity. It requires more time and effort of the educators in order to construct effective MCQs that serve its purpose. Work is only half done if we do not perform Item analysis after the completion of the MCQ test for assessment. Item analysis assesses the quality of each item (questions) in term of its difficulty level (Index of Difficulty), its ability to differentiate between high performers and low performers (Index of Discrimination) and how correctly the distractors are used in each item (Distractor Efficiency). The results of Item analysis helps to improve the strength of the MCQs, as good items can be retained whereas bad items are revised, replaced or removed.

Key words

Item analysis, Index of Difficulty, Index of Discrimination, Distractor efficiency.

Introduction

Assessment is an essential and integral component of learning process as the former drives the latter. The goal of assessment is to measure the learning outcomes but at the same time it also facilitates and improves learning outcomes. Teaching also must undergo appropriate modifications based on the results of assessment, so that teaching aligns itself with the goals of the curriculum that propound the learning outcomes. The theoretical component of the knowledge can be tested by various written format tools viz, long answer questions, short answer questions, multiple-choice questions, modified long essay questions, key feature questions, concordance script test etc., among of the gamut of methods available. MCQs are one of the canonical recipes in the assessment of knowledge due to its reliability and simplicity. It

stands out among the assessment tools as it can evaluate enormous load of curriculum, both its breadth and depth within a short span of time, though the examiner who constructs the questions must invest huge amount of his time and effort to come out with effective MCQs. It can be used to test the whole spectrum of knowledge ranging from recall types of memory to higher-order thinking and application based on how the MCQ is constructed and its target audience. It is increasingly used in the assessment of medical undergraduates during their formative as well as summative assessments in addition to various screening tests that qualify the candidates to various medical education programs throughout the country. It can be a critical segment in the comprehensive competence assessment game plan when it is combined with other tools as it is efficient, objective and discriminative when due importance is given for its construction [1]. Usually MCQs are made to two parts- the first part is called the stem which contains the problem or statement and the second part, called the response which contains a list of probable explanations. The list contains a single best answer called key, along with inappropriate options which are wrong answers called distractors [2]. The distractors also must be logical and meaningful so that they provide a tough competition while choosing the correct option. Though it is extensively used in medical curriculum, the antagonistic truth is that most of the professionals involved in making and using the MCQs for assessment purpose are in dire need of proper knowledge and training in constructing and analyzing it for its intended purpose. The work done is only sketchy if the teachers use the MCQs to assess the students' knowledge without assessing the MCQs for its validity and reliability, which is called the Item analysis [3]. Thus the aim of this article is provide adequate knowledge on how to assess the MCQs for its validity and reliability, which is called Item analysis in simple terms.

Invest quality time in the construction of effective MCQs

When preparing MCQs for assessment, do not pick it directly from the MCQ books or from the end of the chapters given in certain textbooks. The strength of the assessment tool depends on how it is made. If adequate thought process goes into the making of the MCQs, an efficient tool emerges that provides valuable insights to both the learner and the instructor. The time spent in making effective MCQs is worth for the following reasons.

1. It helps to identify the strengths and weakness in students understanding of a particular topic, while providing guidelines to teachers [4].
2. Higher-order cognitive processing, as per Blooms taxonomy, such interpretation of data, application of learned information and synthesis of new knowledge can be tested, instead of just pure recall of sheer facts [5, 6].
3. The assessment is completely objective with consistent scoring and grading in comparison to other tools of assessment [7, 8].
4. It conjures complex thought process as students have to deliberate on different units of information with each other to arrive at a decision [9].
5. Due to its high objectivity, results can be released immediately after the examination, as it can be corrected by anybody (including machines) who has the keys [10, 11].
6. As correction can be done by a machine (computer), large number of students can be assessed within a short span of time [12, 13].
7. A context-rich question hones the problem-solving ability of the student, which help him to encounter real situation in his practice with confidence [9].

Item analysis

As mentioned earlier, the utilization of the MCQ as an assessment tool to analyze the knowledge of the students is only work half done. The remaining important part of the work, called Item

analysis, is to analyze the MCQ itself in terms of its level of difficulty and its ability to discriminate between high performing students and low performing students. It assesses the student responses to individual test items in order to assess the quality of each item and of the test as a whole. Item analysis makes use of statistical methods and expert judgments to gauge a test based on the quality of individual items, item sets, and the entire set of test items, as well as the relationships among the test items [14]. The information provided by this analysis can be used to improve the quality and accuracy of the individual item as well as the test as a whole [15]. Items with ambiguous distractors or misleading information in the stems can be eliminated from the MCQ pool. Through item analysis, the instructors can improve their skill in constructing valid MCQs in the future. In addition it also directs the curriculum administrators to identify specific areas of the course content that needs revision or further clarification as evidenced by poor mastery of the subject in that area by the students [16]. The three important indices discussed in this article are index of difficulty, index of discrimination and distractor efficiency.

Index of difficulty

Any item analysis starts with the analysis of index of difficulty. Simply it states the level of difficulty in answering the questions by the respondents during their test. It indicates the level of difficulty of the items in relation to the cognitive ability of the respondents [17]. It is defined as the proportion or percentage of respondents who have correctly answered the questions [18-20].

How to calculate the Index of Difficulty?

The scores of the respondents are arranged in the order of merit so that the upper 27% forms the higher achievers (HA) and the lower 27% forms the lower achievers (LA). To calculate the Index of Difficulty we need three parameters. They are a) number of higher achievers, b) number of lower achievers (LA) and c) total number of respondents (N) and the formula to calculate the Index of Difficulty is given by Kelly method [21]:

$$\text{Index of Difficulty} = \frac{\text{HA+LA}}{\text{N}}\text{X}100$$

*HA – number of respondents in the Higher Achievers group who have correctly answered the item.

*LA - number of respondents in the Lower Achievers group who have correctly answered the item.

*N – total number of respondents in HA and LA groups

Table - 1: Ranges of index of difficulty.

Sr. No.	Range	Difficulty level
1	20 & below	Very difficult
2	21-40	Difficult
3	41-60	Average
4	61-80	Easy
5	81 & above	Very easy

Ranges of index of difficulty are depicted in **Table – 1**. Index of Difficulty describes the percentage of respondents who have answered an item correctly and it can vary from 0 to 100% [21]. Actually it is a misnomer as higher the value easier is the item and lesser the value more difficult is the item, and because of this some authors designated this as Ease Index [22]. When selecting MCQs for a test, the very difficult level (20 & below) and very easy level (81 & above) should generally be avoided. Before avoiding, particular attention must be paid to very difficult level items as it indicates that the students have not understood the concepts well, the particular topic was not given adequate attention or teachers have not taught that particular topic properly. Item moderation should follow item analysis for very difficult items, and if it cannot be moderated either it should be discarded or replaced [23]. In a test the easy, average and difficult items should be in the proportion of 25%, 50% and 25% respectively [24].

Index of discrimination

After analyzing index of difficulty, next go for the determination of Index of Discrimination. It tells about the capacity of an item to differentiate higher achiever group from lower achiever group [25]. It determines whether those who have performed better throughout the test also performed better on each item. Like in the Index of Difficulty, here also the higher achiever group and lower achiever group is compared with one another.

How to calculate Index of Discrimination?

$$\text{Index of Discrimination} = \frac{\text{HA}-\text{LA}}{\text{N}} X\ 2$$

*HA, LA, and N are same as in Index of Difficulty.

Table – 2: Classification of items based on the index of discrimination [26].

Sr. No.	Index of Discrimination levels	Description
1	0.40 & above	Very good items
2	0.30–0.39	Reasonably good
3	0.20–0.29	Marginal items (subject to improvement)
4	0.19 Or less	Poor items (to be rejected or improved by revision)

Classification of items based on the index of discrimination [26] is shown in **Table – 2**. It ranges between -1.00 and +1.00. The index gives positive values ((between 0.00 and +1.00) when HA group respondents choose correct option in each item than the LA group respondents. It gives negative values (between -1.00 and 0.00) when LA group respondents got a specific item correct more often than the HA group respondents [27]. In other words, negative Discrimination index signifies that more knowledgeable respondents are selecting wrong options whereas the less knowledgeable respondents are selecting correct options more often. Index of Difficulty and Index of Discrimination are often related reciprocally to each other, but not always. Items having high Index of Difficulty value (easier questions), discriminate poorly whereas questions with a low Index of Difficulty value (harder questions) are considered to be good discriminators [28].

Distractor efficiency

Distractors are the wrong options in the response segment of the MCQs. Analysis of the strength of these distractors in confusing students while selecting the correct option is called Distractor Efficiency. There are two types of distractors based on how frequently it is chosen by the respondents. Functional Distractors (FD) are the distractors that are chosen by more than 5% of the respondents whereas the Non-functional Distractors (NFD) is the distractors that are chosen by less than 5% of the respondents [22].

How to calculate the Distractor Efficiency?

Analyze each item in terms of how many times each distractor was selected by the respondents.

$$\text{Distractor Efficiency (DE)} = \frac{\text{Number of times an option was selected}}{\text{N}} \text{X}100$$

N is the total number of respondents.

DE ranges from 0% to 100%. If an item contains three or two or one or nil NFDs, then DE would be 0, 33.3%, 66.6%, and 100%, respectively [29]. Once identified, these NFDs must be revised, replaced, or removed [30] as erroneous MCQs affect the performance of the HA group than the other group of respondents [31]. It is equally important in making a MCQ tool more efficient when compared to Index of Difficulty and Index of Discrimination. The performance of the respondents is influenced by the way the distractors are designed [32]. Ideally, if the distractors are properly designed, it should lead to LA group selecting these options more often than the HA group [27].

Summary

MCQs are effective tools to assess the learning outcomes of a teaching programme. Item analysis makes this tool more powerful by assessing the difficulty level of the question,

differentiating high achievers from low achievers and pointing out the areas of course content that needs further clarification or modification.

References

1. AM Brady. Assessment of learning with multiple-choice questions. Nurse Educ Pract., 2005; 5: 238-42.
2. Cizek GJ, O'Day DM. Further investigation of nonfunctioning options in multiple-choice test items. Educational Psychol Measurement, 1994 Dec; 54(4): 861-72.
3. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple-choice questions for use in nursing research and education. Collegian, 2005; 12: 19-24.
4. Tan LT, McAleer JJ. Final FRCR Examination Board. The introduction of single best answer questions as a test of knowledge in the final examination for fellowship of the Royal College of Radiologists in Clinical Oncology. Clin Oncol (R Coll Radiol), 2008; 20: 571-6.
5. Carson J, Delpierre G, Masters K. Designing and managing MCQs: Appendix C: MCQs and Bloom's taxonomy. (Online) 2011.
6. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences, National Board of Medical Examiners, 3rd edition, (Online) 2010.
7. Öztürk M. Multiple-Choice Test Items of Foreign Language Vocabulary. Eğitim Fakültesi Dergisi, 2007; 20(2): 399-426.
8. Brown H. D. Language Assessment: Principles and Classroom Practices. White Plains, NY: Pearson Education, 2004.
9. Schuwirth LW, Verheggen MM, van der Vleuten CP, Boshuizen HP, Dinant GJ. Do short cases elicit different thinking processes than factual knowledge questions do? Med Educ., 2001; 35: 348–56.
10. Premadasa IG. A reappraisal of the use of multiple-choice questions. Medical Teacher, 1993; 15(2-3): 237–42.
11. Lowe D. Set a multiple choice question (MCQ) Examination. BMJ, 1991; 302(6779): 780–2.
12. Edward M. Multiple choice questions: their value as an assessment tool. 6. Vol. 14. Lippincott Williams & Wilkins, Inc; 2001, p. 661–666.
13. Hammond EJ, McIndoe AK, Sansome AJ, Spargo PM. Multiple-choice examinations: adopting an evidence-based approach to exam technique. Anesthesia, 1998; 53(11): 1105–8.
14. McCowan R., McCowan Sh. Item Analysis for Criterion-referenced Tests, 1999. Retrieved on June 20, 2020.
15. Ary D., Jacobs L.C., Razavieh A. Introduction to Research in Education, 6th edition, Wadsworth, California, 2002.
16. Suruchi, Rana S.R. Test item analysis and relationship between difficulty level and discrimination index of test items in an achievement test in biology. Indian J. Res., 2015; 3(6): 56–58.
17. Boopathiraj C., Chellamani K. Analysis of test items on difficulty level and discrimination index in the test for research in education. Int. J. Soc. Sci. Interdiscip. Res., 2013; 2: 2.
18. Reynolds C. R., Livingston R. B. Mastering Modern Psychological Testing: Theory and Methods. Upper Saddle River, NJ: Pearson Education, 2012.
19. Fulcher G., Davidson F. Language Testing and Assessment: An Advanced Resource Book. New York: Routledge, 2007.
20. Zajda J. Learning and Teaching. Australia: James Nicholas Publisher Pty Ltd., 2006.
21. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice

questions: A descriptive analysis. BMC Med Educ., 2009; 9: 1–8.
22. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and distracter efficiency. J Pak Med Assoc., 2012; 62: 142–7.
23. Jonathan A. Odukoya, Olajide Adekeye, Angie O. Igbinoba, Afolabi A. Item analysis of university-wide multiple choice objective examinations: the experience of a Nigerian private university. Qual Quant, 2018; 52: 983–997.
24. Ananthakrishna N. The item analysis: In: Medical education principles and practice; by Anathakrishnan N, Sethukumaran KR, Kumar S, editors. ch. 20. Pondicherry, India: Alumni Association of National Teacher Training centre, JIPMER; 2000, p.131-7.
25. Fowell SL, Southgate LJ, Bligh JG. Evaluating assessment: the missing link? Med Educ., 1999; 33: 276-81.
26. Ovwigho B .O. Empirical Demonstration of Techniques for Computing the Discrimination Power of a Dichotomous Item Response Test. IOSR Journal of research and Method in Education, 2013; 3(2): 12-17.
27. Muzaffar RH, Farhan J. Analysis of One-Best MCQs: the Difficulty Index, Discrimination Index and Distractor Efficiency. J Pak Med Assoc., 2012; 62(2): 142-47.
28. Carroll RG. Evaluation of vignette-type examination items for testing medical physiology. Am J Physiol., 1993; 264: S11-5.
29. Rao C, Kishan Prasad HL, 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: 201-4.
30. Gronlund NE, Linn RL. Measurement and evaluation in teaching. 6th edition, New York: Macmillan Publishing Co; 1990.
31. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. Med Educ., 2008; 42: 198-206.
32. Dufresne RJ, Leonard WJ, Gerace WJ. Making sense of student's answers to multiple-choice questions. Phys Teach., 2002; 40: 174-80.