CAN ITEM ANALYSIS OF MCQs ACCOMPLISH THE NEED OF A PROPER ASSESSMENT STRATEGY FOR CURRICULUM IMPROVEMENT IN MEDICAL EDUCATION?

By

YOGESH R. PAWADE *   DIPTI S. DIWASE **

* Assistant Professor, Department of Biochemistry, Seth G.S. Medical College & K.E.M. Hospital, Mumbai, India.
** Assistant Professor, Department of Information Technology, K.J. Somaiya College of Engineering, Vidyavihar, Mumbai, India.

ABSTRACT

Item analysis of Multiple Choice Questions (MCQs) is the process of collecting, summarizing and utilizing information from students' responses to evaluate the quality of test items. Difficulty Index (p-value), Discrimination Index (DI) and Distractor Efficiency (DE) are the parameters which help to evaluate the quality of MCQs used in an examination. This study has been postulated to investigate the relationship of items having good p-value and DI with their DE and their utility to frame 'ideal questions'. This study further evaluates the MCQs as a tool of assessment so as to improve the curricula in Medical Education. In this study, 20 test items of 'Type A' MCQ tests of assessment were selected. The p-values, DI and DE were estimated. The relationship between the p-value and DI for each test item was determined by Pearson correlation analysis. Mean p-value and DI of the test were 66.53 ± 16.82% and 0.41 ± 0.16% respectively. Only 20% of total test items crossed the p-value of 80% indicative of their easy difficulty level. 95% of the test items showed acceptable (> 0.2) DI. 12 out of 20 test items showed excellent DI (≥ 0.4). 8 (40%) test items were regarded as 'ideal' having p-value from 30-70, and DI > 0.24. Correlation studies revealed that, DI associated with p-value (r = -0.288; P = 0.219). Mean DE of the test was 76.25 ± 22.18%. The DE was directly related to the DI. Items with good and excellent DI had DE of 66.67 ± 14.43% and 83.33 ± 19.46% respectively. In conclusion, an acceptable level of test difficulty and discrimination was maintained in the type A MCQ test. The test items with excellent discrimination tend to be in the moderately difficult range. There was a consistent spread of difficulty in type A MCQ items used for the test. Much more of these kinds of analysis should be carried out after each examination to identify the areas of potential weakness in the type A MCQ tests to improve the standard of assessment.

Keywords: Item Analysis, Difficulty Index, Discrimination Index, Distractor Efficiency, Type A MCQ.

INTRODUCTION

In multi-disciplinary integrated curriculum like medical education, Multiple Choice Questions (MCQs) are used mostly for comprehensive assessment at the end of a semester or academic sessions and provide feedback to the educators on their academic performance. Scheming MCQs is a multifarious and time consuming process as compared to the descriptive questions. After the assessment, a medical academician needs to know the effectiveness of the test questions in resounding students' learning related performance in the course. Because of versatility in the assessment, MCQs are one of the best and commonly used assessment tool to gauge the knowledge competencies of medical students. Appropriately, constructed MCQs evaluate higher-order cognitive processing of Bloom's taxonomy such as interpretation, synthesis and application of knowledge, more than just testing recall of isolated facts [1], [2].

Among the different types of MCQs used in the medical field, the most frequently used type is the single best-response type (type A MCQ) with four choices [3]. These test questions were taken from the subject of Biochemistry. The examination questions had been formulated by the content experts who taught the respective syllabi and

scrutinized by the senior academicians of the department. Item analysis is the process of assembling, summarizing and using information from students' responses to assess the quality of test items [4]. The item statistics can help to determine effective items and those that need improvement or omission from the question bank. It allows any aberrant items to be given attention and revised. One of the most widely used methods in investigating the reliability of a test item has been Classical Test theory (CT) item analysis [4], [5]. This type of item analysis essentially determines test homogeneity. The more similar are the items given in the test; the more likely they measure the same kind of intended aptitude and therefore attaining higher reliability.

In CT, item difficulty index (p-value), also called as "ease index" is the first item characteristic to be determined [5]. It is described as the percentage of the total group of students selecting the correct answer to that question. It ranges from 0 – 100%. The higher the percentage, the easier the item. The recommended range of difficulty is from 30 – 70%. Items with p-values <30% and above 70% are considered difficult and easy items respectively [2].

$$p\ value = \frac{Frequency\ of\ marking\ correct\ key}{Total\ No.\ of\ students}$$

It is very obligatory, as the reliability of the tests to measure students' performance are often questioned due to nonconformity of item difficulty with the ability of the students. Very easy items should usually be placed either at the start of the test as 'warm-up' questions or removed altogether. The difficult items should be reviewed for possible perplexing language, areas of disagreement, or even an inappropriate key. Inclusion of very difficult items in the test depends upon the target of the teacher, who may want to include them in order to identify top scorers.

Along with difficulty index, item Discrimination Index (DI), also called as "point biserial correlation" is another important guide [4]. This provides information on the efficacy of the items in a given test to discriminate between students with higher and lower abilities [6].

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

where, H and L are the number of correct responses in the high and low groups respectively. N is the total number of students in both high and low groups.

It ranges between -1.00 and +1.00. It is expected that, the high-scorers select the correct answer for each item more often than the low scorers. If this is true, the assessment is said to have a positive DI (between 0.00 and +1.00), indicating the total high scorers, chose the correct answer for a specific item more often than the overall low scorers. If, nevertheless, the low scorers got a specific item correct more often than the high scorers, then that item has a negative DI (between -1.00 and 0.00). Culpabilities in structuring test items logically affects the values of discrimination index. Items with poor discrimination ability should be inspected for potential deficiencies [5].

The difficulty and discrimination indices are associated reciprocally. However, this may not always be true. Questions having high p-value (easier questions), discriminate poorly; conversely, questions with a low p-value (harder questions) are considered to be good discriminators [7].

Another convincing technique is, analysis of distractors that provides information regarding the individual distractors and the key of a test item. Using these tools, the examiner is able to modify or remove specific items from subsequent exams [1]. The distractors are important components of an item, as they show a relationship between the total test score and the distractor chosen by the student. Students' performance depends upon how the distractors are designed [8]. Distractor Efficiency (DE) is one such tool that tells us whether the item was well constructed or failed to perform its purpose. Any distractor that has been selected by less than 5% of the students is considered to be a non-functioning distractor (NF-D) [1]. Distractor efficiency ranges from 0 - 100% and is determined on the basis of the number of NF-Ds in an item. Four NF-Ds: DE = 0%; 3 NF-Ds: DE = 25%; 2 NF-Ds: DE = 50%; 1 NF-D: DE = 75%; No NF-D: DE = 100%. Ideally, low-scoring students, who have not mastered the subject, should choose the distractors more often, whereas, high scorers should discard them more frequently while choosing the correct option. By analysing the distractors, it becomes easier to identify their errors, so

that they may be revised, replaced, or removed.

Tarrent and Ware confirmed that, flawed MCQ items affected the performance of high-achieving students more than borderline students [9]. Construction of a balanced MCQ, therefore, addresses the concerns of the students of getting an acceptable average grade and that of the faculty to have an appropriate spread of the score [10].

Hence, the present research study was envisioned with an objective to analyze the quality of the MCQs (type A) used in the assessments of first year medical students in the preclinical phase and to test the quality of framed MCQs for the subsequent tests. The authors have aimed to find out the relationship between the item difficulty and item discrimination indices of these MCQs with their distractor efficiency and the effect of non-functioning distractors on these indices.

Materials and Methods

Data Collection

MCQs (given in Appendix) were taken from the assessment test papers from the years 2010 – 2014 (each year having one cohort). Each of these examinations was held during the first six months of the preclinical phase and the test paper was based only upon the syllabus assigned for the examination. A total of 20 test items were selected for the item analysis. Each type A MCQ consisted of a stem and four choices and the students were to select one best answer from these four choices. A correct answer was awarded 1/2 mark and there were no negative marks for the incorrect answer.

Item Analysis

The result of the examinees' performance in the test was used to investigate the p-value, DI and DE of each MCQ item. The p-value of an item is calculated as the percentage of the total number of correct responses to the test item [11], [12]. It is calculated using the formula,

$$P = R/T$$

where P is the item difficulty index, R is the number of correct responses and T is the total number of responses (which includes both correct and incorrect responses). An item is considered difficult and easy when the difficulty index value is < 30% and > 70% respectively.

The item DI measures the difference between the percentages of students in the upper group with that of the lower group who obtained the correct responses. At first, top and bottom 27% of the total number (n) of students were counted [12], [13]. The total number of students who obtained the correct response in the Upper Group 27% (UG) and the Lower Group 27% (LG) was counted. The respective percentage of the number of students in upper group (PU) and lower group (PL) is calculated. The discrimination index was calculated using the formula, DI = PU - PL. The higher the discrimination index, better is the ability of the test item to discriminate between students with higher and lower test scores. Based on Ebel's (1972) guidelines on CT item analysis, items were categorized depending on their discrimination indices [4], [14]. The item with negative Discrimination Index (DI) was considered to be discarded; DI: 0.0 – 0.19: poor item: to be revised; DI: 0.20 – 0.29: acceptable; DI: 0.30 – 0.39: good;DI: ≥ 0.40: excellent.

Statistical Analysis

All data were expressed as mean ± SD of the total number of items. The relationship between the item difficulty index and discrimination index for each test item was determined by Pearson correlation analysis using IBM SPSS 22. The correlation is considered significant at 0.01 level (2-tailed).

Results

A total of 180 students appeared for the test consisting of 20 type A MCQs (single-best response MCQs). The mean score achieved was 6.65 ± 1.64 (maximum 10 marks). Mean scores according to groups were: lower 4.41 ± 0.73; middle 6.48 ± 0.75; upper 8.56 ± 0.56. Students were ranked in the order of merit from the highest score of 9.5 to the lowest score of 2.

Table 1 shows the categorization of the first 27% students

	Score	
	Mean	SD
Above 63%	8.56	0.56
27- 63%	6.48	0.75
Below 27%	4.41	0.73
Total	6.65	1.64

Table 1. Scores of Upper and Lower Performance Groups in the Test

MCQ	p-value	DI	DE (%)
1	82.78	0.42	50
2	73.33	0.51	50
3	81.11	0.38	50
4	78.89	0.42	75
5	65.00	0.54	75
6	94.44	0.07	25
7	27.22	0.21	75
8	78.89	0.43	75
9	58.89	0.24	100
10	64.44	0.56	100
11	52.78	0.63	100
12	78.33	0.23	50
13	53.89	0.62	100
14	81.11	0.25	75
15	76.67	0.43	100
16	78.89	0.35	75
17	50.56	0.31	75
18	42.78	0.62	100
19	47.22	0.41	75
20	63.33	0.64	100

Table 2. Distribution of Difficulty Index (p-value), Discrimination Index (DI) and Distractor Efficiency (DE) of 20 MCQs in the tests

p-value	Interpretation	p-value (Mean ±SD)	DE (%) (Mean ±SD)	No of items
> 70	Easy	80.44 ±5.57	62.50 ±21.25	10 (50%)
30 –70	Average	55.43 ±7.97	91.67 ±12.50	9 (45%)
< 30	Difficult	27.22 ±0.00	75.00 ±0.00	1 (5%)
Total		66.53 ±16.82	76.25 ±22.18	20 (100%)

Table 3. Distribution of Difficulty Index (p-value) level in the tests and their Interpretation

DI	Interpretation	DI (Mean ±SD)	DE (%) (Mean ±SD)	No of items
= 0.40	Excellent	0.52±0.09	83.33±19.46	12 (60%)
0.30 –0.39	Good	0.35±0.04	66.67±14.43	3 (15%)
0.20 –0.29	Marginal	0.23±0.02	75.00±20.41	4 (20%)
< 0.20	Poor	0.07±0.00	25.00±0.00	1 (5%)
Total		0.41±0.16	76.25±22.18	20 (100%)

Table 4. Distribution of Discrimination Index (DI) in the tests and their Interpretation

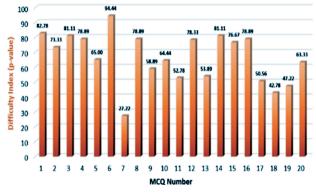Figure 2. Discrimination Indices of MCQS

Figure 1. Difficulty Indices (p-Value) of MCQS

(UG) and the last 27% (LG). The distribution of p-value and DI of the 20 MCQs is depicted in Table 2. The mean p-value of the test was found to be 66.53 ± 16.82% which indicates relatively easy test paper. Only 20% items (1/5th of 20 test items) in this study crossed the p-value of 80% as shown in Table 3 and Figure 1.
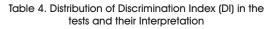
The mean DI of the test was 0.41 ± 0.16 demonstrating the acceptable discrimination quality. 19 (95%) test items were found with DI ≥ 0.2 discriminating good and weak students. 12 (60%) items showed excellent DI ≥ 0.4 as depicted in Table 4 and Figures 2 and 3 depict DE of MCQs.

Figure 4 (a) and 4 (b) reveals the correlation between
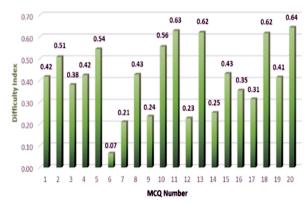

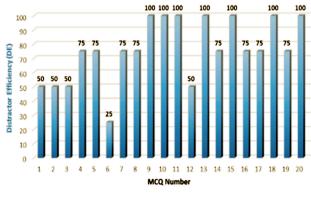
Figure 3. Distractor Efficiency of MCQS

individual item's p-value and DI score. 40% of the test items with DI< 0.4 had the p-value ranging between 27 – 95%. Out of 12 items with excellent DI (≥ 0.4), 58.33% had the p-value between 50 – 80%. Correlation study between p-
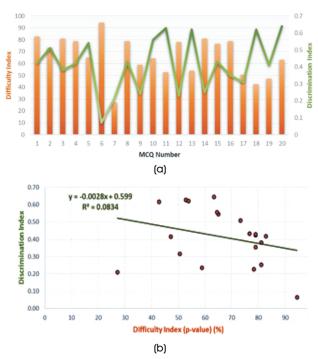
(a)

(b)

Figure 4. (a). Correlation of Difficulty Index with Discrimination, (b). Correlation of Difficulty Index with Discrimination (with trend)

p-value	DI = 0.30	DE (%) (Mean ±SD)	DI = 0.30	DE (%) (Mean ±SD)	No of items
> 70	3 (15%)	50.00 ±25.00	7 (35%)	67.86 ±18.90	10 (50%)
30-70	1 (5%)	100.0 ±00.00	8 (40%)	90.63 ±12.94	9 (45%)
< 30	1 (5%)	75.00 ±0.00	0 (0%)	0.00 ±0.00	1 (5%)
Total	5 (25%)	65.00 ±28.50	15 (75%)	80.00 ±19.36	20 (100%)

Table 5. Categorization of Difficulty Index (p-value) and Discrimination Index (DI) of 20 MCQs in the tests

value and DI showed that DI correlated negatively with p-value ($r = -0.288$; $P = 0.219$), but it was statistically insignificant. Negative correlation signifies that, with increasing p-values, there is decrease in DI. As the items get easy (above 75%), the level of DI decreases consistently.

Corresponding DE of 20 MCQs was also worked out, details of which are given in Tables 2 and 6. Two items having lowermost DIs, with item 6 being the easiest one (p = 94.44; DI = 0.07; DE = 25), and item 7, most difficult (p = 27.22; DI = 0.21; DE = 75). 45% items were of average (recommended) difficulty with a mean p-value of 55.43 ± 7.97%. Similarly, majority of items (60%) had excellent DI (0.52 ± 0.09), with few items with marginal (20%) and poor (5%) DI. This revealed that, 8 (40%) items were 'ideal' having a p-value from 30 – 70, and DI > 0.24, as evident from Table 5. The total number of distractors were 80 (4 per item).

Out of the 80 distractors, 19 (23.75%) were NF-Ds. 13 (65%) items had NF-Ds, while 7 (35%) items had effective distracters.

Table 6 explains that, items with 3 NF-Ds had high p-value (94.44%) and poor DI (0.07), whereas items with 2, 1 and no NF-Ds had recommended p-values 78.89 ± 4.13, 63.47 ± 19.90 and 58.97 ± 10.71, and excellent DI 0.39 ± 0.12, 0.37 ± 0.11 and 0.53 ± 0.15 respectively.

Figure 5 shows that, DE varies indirectly with the p-value, with most difficult items having DE of 75.00% and easy items having DE 62.50 ± 21.25%. Items with average difficulty had DE of 91.67 ± 12.50%. Figure 6 reveals that, DE is directly related to the DI. Items with good and excellent DI had DE of 66.67 ± 14.43% and 83.33 ± 19.46% respectively.

DE (%)	No of NF-Ds	p-value (Mean ±SD)	DI (Mean ±SD)	No of items
25	3	94.44 ±0.00	0.07 ±0.00	1 (5%)
50	2	78.89 ±4.13	0.39 ±0.12	4 (20%)
75	1	63.47 ±19.90	0.37 ±0.11	8 (40%)
100	None	58.97 ±10.71	0.53 ±0.15	7 (35%)
Total		66.53 ±16.82	0.41 ±0.16	20 (100%)

Table 6. Distribution of Distractor Efficiency (DE) in the tests and Non-functional Distractors (NF-Ds)

Figure 5. Correlation of Difficulty Index with Distractar Efficiency

Figure 6. Correlation of Discrimination Index with Distractor Efficiency

Distractor analysis gives an opportunity to study the responses made by the students on each alternative of the item. The analysis of 5 questions selected on the basis of p-value and DI gave a varied result.

The MCQ No. 6 was a very easy question (p-value = 94.44%) with lowest DI (0.07), as both upper and lower groups selected, nearly equally, the correct answer. The DE was 25% and 94.44% students selected the correct response, making the rest of the alternatives useless.

The MCQ No. 7 was selected for its lowest p-value (27.22%) making it the most difficult item with more of the higher group choosing the correct response, but paradoxically the DI was poor (0.21). The DE was 75% as the distractor alternative 'd' served no purpose at all, as hardly 5 students selected it, and distractor alternative 'b' was so effective that 88 students selected it, made it most difficult to choose even from the reduced number of choices. Distractor alternative 'b' was considered the right answer by many students both in the upper & lower groups and needs to be revised to discriminate properly from the correct choice.

MCQ No. 13 had a p-value of 53.89, DI of 0.62, and DE of 100%, showing that, it was moderately difficult and being able to differentiate students into different strata. From the upper group, 85.19% students selected the correct response 'c', while 76.92% students of the lower group were distributed among all the distractor choices.

MCQ No. 18 (DI = 0.62) was an excellent discriminatory and of average difficulty level (p-value = 42.78%), as 80% of the students in the upper group chose the correct alternative 'd' and 82% students in the lower group chose the distractors. This question had DE of 100% as the distractors were not clear to many students, making it a very difficult item.

MCQ No. 20 (DI = 0.64) was another excellent discriminatory with DE of 100% and of average difficulty level (p-value = 63.33%), as 93% of the students in the upper group chose the correct alternative 'c' and 28% students in the lower group chose the same correct alternative.

Discussion

The assessment tool is one of the strategies which should be designed according to the objective, to strategize the assessment tool. One-best response MCQs, if properly written and well-constructed, are one of the strategies of the assessment tool that quickly assess any level of cognition according to Bloom's taxonomy [1].

The mean difficulty index scores of the test was 66.53 ± 16.82%. Only 50% of the total test items had difficulty index scores crossing 70%. This observation was similar to a study in a medical school reported by Sim Si-Mui and Rasiah (2006), who found that, about 40% of the MCQ items crossed difficulty index 70% showing that, the test items were easy for the examinees [12]. Brown (1983) and Algina (1986) have reported that, any DI of 0.2 or higher is acceptable and the test item would be able to differentiate between the weak and good students [15], [16]. In the present study, 95% of the MCQs from the test had DI of more than 0.2. Thus it showed that, most of the MCQs were good or satisfactory questions which would not need any modifications or editing. 12 of the 20 items showed DI equal to or more than 0.4, indicating that, these MCQ items were excellent test items for differentiating between poor and good performers.

Negative correlation between difficulty and discrimination indices indicated that, with increase in difficulty index, there is decrease in discrimination index. As the test items get easier, the discrimination index decreases, thus it fails to differentiate weak and good students. Sim Si-Mui and Rasiah (2006) established that, maximum discrimination occurred with difficulty index between 40 – 74% [12]. In the present study, 76.9% of the test items with difficulty index between 50% and 79% had excellent discrimination index.

For calculation of the DI, the method adopted by Kelley (1939) was used in which upper and lower 27% performers were selected [13]. The only limitation of this test is that it cannot be used for a smaller sample size. But in this study, the sample size was 180 and hence the observed results truly reflect the discriminative power of the test items. One inadequacy of only analysing a question in terms of its difficulty index is the inability to differentiate between students of widely differing abilities. Subjective judgment of item difficulty by item writer and the vetting committee may allow faulty items to be selected in the item bank. Items with poor discrimination index and too low or too high

difficulty index should be reviewed by the respective content experts [17]. This serves as an effective feedback to the respective departments in a medical college about the quality control of various tests. When the difficulty index is very small, indicating difficult question, it may be that, the test item is not taught well or is difficult for the students to grasp. It also may indicate that, the topic tested is inappropriate at that level for the students [18].

In the scatter plot, there is a wide variation in the DIs with similar levels of difficulty index below 75%. Guessing practices by the students might have caused the wide variation in DIs, as the negative marking scheme is presently not implemented in MCQ tests. An acceptable level of test difficulty and discrimination indices appears to be maintained the test. This observation could be due to the fact that the test items went through a series of screening before being selected for the examinations. The quality of test items may be further improved based on action taken in reviewing the distractors by the item writer based on the calculated discrimination and difficulty index values. Few common causes for the poor discrimination are ambiguous wording, grey areas of opinion, wrong keys and areas of controversy [19]. Items showing poor discrimination should be referred back to the content experts for revision to improve the standard of these test items. It is important to evaluate the test items to see their efficiency in assessing the knowledge of the students based on the difficulty and discrimination indices of the test items.

A distractor used by less than 5% of students is not an effective distractor and should be either replaced or corrected as it affects the overall quality of the question. An NF-D makes the question easier to answer, thereby affecting the assessment of the student. Items having NF-D can be carefully reviewed. With some alterations in the distractors, can be given as the initial item on the test, as a 'warm-up' question. However, they would not be able to differentiate among students, if that is the purpose. Assessment of MCQs by these indices highlights the importance of assessment tools for the benefit of both the student and the teacher [20].

The DE of difficult items in our study was 75% – 100% which was expected, as difficult items would require a lot of guesswork on the part of the student, thereby using all the distractors. The numbers of NF-Ds also affect the discriminative power of an item. It is seen that reducing the number of distractors from four to three decreases the difficulty index, while increasing the DI and the reliability [21].

We observed that, items having all four functioning distractors had excellent discriminating ability (DI = 0.53 ± 0.15) as compared to items with any number of NF-Ds. This contradicts other studies favoring better discrimination by three distractors as compared to four [22].

It was also observed that, items having good/average difficulty index (p-value = 30 – 70) and good/excellent DI (> 0.24), considered to be 'ideal' questions, had DE of 90.63 ± 12.94%, which is close to items having no NF-D.

Tarrant and Ware found three-option items performing equally well as four-option items and have suggested to write three-option items as they require less time to be developed [23]. Similar observations were made by literature review conducted by Vyas et. al [24]. This can be because, writing items with four distractors is a difficult task and while writing the fourth distractor, we are mostly trying to fill the gap, allowing it to become the weakest distractor.

Results from this study clearly highlighted the importance of item analysis of MCQs. Items having average difficulty and high discrimination with functioning distractors should be incorporated into future tests to improve the test development and review. This would also improve the overall test score and properly discriminate among the students.

Conclusion

There was a consistent spread of difficulty in type A MCQ items used for the test. The test items that demonstrated excellent discrimination tend to be in the moderately difficult range. Items with all functional distractors performed best in discriminating among the students. Factors other than the difficulty, like the faulty test item constructions, are not significant at the test. The results of this study should initiate a change in the way MCQ test items are selected for any examination and there should be a proper assessment strategy as part of the curriculum

development in medical education. These kinds of analysis should be carried out after each examination to identify the areas of potential weakness in the type A MCQ tests to improve the standard of assessment.

References

[1]. J. Carneson, G. Delpierre, and K. Masters, (2011). *Designing and Managing MCQs: Appendix C: MCQs and Bloom's Taxonomy.* Retrieved from URL:http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html

[2]. S. M. Case, and D. B. Swanson, (2010). *Constructing written Test Questions for the Basic and Clinical Sciences, National Board of Medical Examiners, 3rd ed.* Retrieved from URL: http://www.nbme.org/ publications/item-writing-manual.html

[3]. E.N. Skakun, E. M. Nanson, W.C. Taylor, and S. Kling, (1977). "An investigation of three types of multiple choice questions". *Ann Conf. Res. Med Educ.*, Vol. 16, pp. 111-116.

[4]. A.M. Zubairi, and N.L.A. Kassim, (2006). "Classical and Rasch analysis of dichotomously scored reading comprehension test items". *Malaysian J. of. ELT. Res.,* Vol. 2, pp. 1-20.

[5]. A. Davies, (1990). *Principles of Language Testing.* Cambridge. Oxford: Basil blackwell Ltd.

[6]. S.L. Fowell, L.J. Southgate, and J.G. Bligh, (1999). "Evaluating assessment: the missing link?" *Med Educ.*, Vol. 33, pp. 276-281.

[7]. R.G. Carroll, (1993). "Evaluation of vignette-type examination items for testing medical physiology". *Am. J. Physiol.*, Vol. 264, pp. S11-S15.

[8]. R.J. Dufresne, and W.J. Leonard, W.J. Gerace, (2002). "Making sense of student's answers to multiple-choice questions". *Phys Teach.*, Vol. 40, pp. 174-180.

[9]. M. Tarrant, J. Ware, (2008). "Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments". *Med. Educ.*, Vol. 42, pp. 198-206.

[10]. R.G. Carroll, (1993). "Evaluation of vignette-type examination items for testing medical physiology". *Am. J. Physiol.,* Vol. 264, pp. S11-S15.

[11]. A. Johnston, (2003). *LTSN Physical Sciences Practice Guide Effective Practice in Objective Assessment the Skills of Fixed Response Testing.* Retrieved from: http://www.heacademy.ac.uk/physsci/home/pedagogicthemes/assessment2003

[12]. Sim Si-Mui, and R.I. Rasiah, (2006). "Relationship between item difficulty and discrimination indices in true/false type multiple choice questions of a para-clinical multi disciplinary paper". *Ann. Acad. Med.*, Vol. 35, pp. 67-71.

[13]. T. L. Kelley, (1939). "The selection of upper and lower groups for validation of test items". *J. Educ. Psychol.*, Vol. 30, pp. 17-24.

[14]. R.L. Ebel, (1972). *Essentials of Educational Measurement.* 1st Ed, New Jersey; Prentice Hall.

[15]. F. G. Brown, (1983). *Principles of Educational and Psychological Testing,* 3rd Ed. New York: Holt, Rinehart and Winston.

[16]. L. Crocker, and J. Algina, (1986). *Introduction to Classical and Modern Test Theory.* New York: Holt, Rinehart and Winston.

[17]. Z. Meshkani, and H. Abadie, (2005)."Multivariate analysis of factors influencing reliability of teacher made tests". *Journal of Med. Ed.*, Vol. 6, pp. 149-159.

[18]. R.N. Marso, and F.L. Pigge, (1991)."An analysis of teacher made tests and item construction errors". *J. Contemp. Edu. Psych.*, Vol. 16, pp. 279-286.

[19]. D. Baner, V. Kopp, M.R. Fischer, (2007). "Answer changing in multiple choice assessment change that answer when in doubt and spread the word". *BMC Med. Education.*, Vol. 7, pp. 28.

[20]. J. Pellegrino, N. Chudowsky, R. Glaser, (2001). *Knowing What Students Know: The Science and Design of Educational Assessment.* Washington, DC: National Academic Press.

[21]. M.C. Rodriguez, (2005). "Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research". *Educational Measurement: Issues and Practice.,* Vol. 24, pp. 3-13.

[22]. M. Tarrant, J. Ware, and A. M. Mohammed, (2009). "An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis". *BMC Med. Educ.*, Vol. 9, pp. 40.

[23]. M. Tarrant, and J. Ware, (2010). "A comparison of the psychometric properties of three- and four-option multiple-choice questions in nursing assessments". *Nurse Educ. Today.*, Vol. 30, pp. 539-543.

[24]. R. Vyas, and A. Supe, (2008). "Multiple choice questions: a literature review on the optimal number of options". *Nat. Med. J. India.*, Vol. 21, pp. 130-133.

Appendix

Multiple Choice Questions (MCQs)   Marks: 10

1. Which of the following is <u>NOT</u> true for hydrops fetalis?
   a) It is a α -thalassemia
   b) It is an α -thalassemia
   c) Death occurs soon after birth
   d) All four copies of α-globin genes are mutated

2. The nucleotide present at the 3' end of a t-RNA molecule is
   a) uridylate b) cytidylate c) thymidylate d) adenylate

3. The Michaelis Menten Constant (Km) is
   a) equilibrium constant for the dissociation of EI → E + P
   b) not changed by the presence of non-competitive inhibitor
   c) equal to ½ Vmax
   d) the substrate concentration at ½ Vmax

4. All of the following features are seen in Marasmus <u>EXCEPT</u>
   a) fatty liver b) muscle wasting c) growth retardation
   d) anaemia

5. The quality of protein is assessed by all of the following <u>EXCEPT</u>
   a) protein efficiency ratio (PER) b) biological value (BV)
   c) net protein utilization (NPU) d) total number of aminoacids present

6. If the amount of adenine present in a DNA molecule is 20%, then the amount of guanine will be
   a) 20   b) 30   c) 35   d) 40

7. Transcription of a gene in eukaryotes
   a) always begins at AUG codon
   b) requires a primer RNA
   c) always reads the DNA strand in 3'→5' direction
   d) does not require local unwinding of two DNA strands

8. Succinate dehydrogenase is inhibited by malonate as it
   a) is a structural analogue of succinate
   b) is a structural analogue of fumarate
   c) binds to a site other than the active site
   d) brings about a conformational change in the active site of the enzyme

9. The pigment visual purple contains
   a) 11-cis-retinal
   b) 11-cis-retinol
   c) all-trans-retinal
   d) all-trans-retinol

10. Which of the following coenzymes is not derived from vitamins?
    a) Coenzyme A   b) FAD   c) TPP   d) Coenzyme Q

11. The linkages between the glucose residues of glycogen are
    a) α (1 → 4) & β (1 → 4)
    b) β (1 → 4) & α(1→ 6)
    c) α(1→ 4) & α(1→ 6)
    d) α(1→4) & β (1 → 6)

12. Zein of corn is called an incomplete protein as it lacks
    a) alanine & lysine
    b) leucine & aspartic acid
    c) lysine & tryptophan
    d) methionine & proline

13. All of the following forces may play a role in the formation of the tertiary structure <u>EXCEPT</u>
    a) hydrogen bond
    b) disulphide bridges
    c) hydrophobic interactions
    d) peptide bonds

14. Which one of the following is glycerophospholipid?
    a) Sphingomyelin
    b) Cerebroside
    c) Phosphatidyl inositol
    d) Ganglioside

15. All of the following statements about prostaglandins are true <u>EXCEPT</u>
    a) They are cyclic fatty acids
    b) They are potent biologic effector
    c) They cause uterine contraction
    d) They are synthesized only in the prostate gland

16. Fetal hemoglobin contains the following chains
    a) 2α & 2δ b) 2α & 2γ c) 2α & 2β d) 2α & 2ϵ

17. One of the following is the major storage and transport form of fatty acids
    a) Cholesterol
    b) Triacylglycerol
    c) Phospholipid
    d) Albumin

18. Pyridoxal phosphate is required as a coenzyme in all of the following <u>EXCEPT</u>
    a) heme synthesis
    b) transamination
    c) decarboxylation
    d) glycogen synthesis

19. One of the following points about micro-filaments is NOT true?
    a) They form cytoskeleton with microtubules
    b) They provide support and shape
    c) They form intracellular conducting channels
    d) They are involved in muscle cell contraction

20. The rate limiting enzyme in heme biosynthesis is
    a) ferrochelatase
    b) δ-amino levulinic acid dehydratase
    c) δ-amino levulinic acid
    d) uroporphyrinogen

ABOUT THE AUTHORS

Dr. Yogesh R. Pawade is currently working as an Assistant Professor in the Department of Biochemistry at Seth GSMC & KEM Hospital, Parel, Mumbai. He completed his M.B.B.S. from GMC&Hospital, Nagpur and has completed his Post-Graduation in Bio-chemistry and received M.D. (Biochemistry) from IGGMC & Mayo Hospital, Nagpur. His research interests are Clinical Biochemistry, Medical Genetics and Medical Education Technology.

Dipti S. Diwase is currently working as an Assistant Professor in the Department of Information Technology at K.J Somaiya College of Engineering, Vidyavihar, Mumbai, India. She received her B.E. degree in Computer Science and Engineering from Sant Gadge Baba University and M.E. degree in Embedded System and Computing from G.H. Raisoni College of Engineering, Nagpur. Her research interests are Fuzzy Logic, Web Security and Programming Languages.