Comparison between three option, four option and five option multiple choice question tests for quality parameters: A randomized study

Bhavisha Vegada, Apexa Shukla, Ajeetkumar Khilnani¹, Jaykaran Charan², Chetna Desai³

Abstract:

Background: Most of the academic teachers use four or five options per item of multiple choice question (MCQ) test as formative and summative assessment. Optimal number of options in MCQ item is a matter of considerable debate among academic teachers of various educational fields. There is a scarcity of the published literature regarding the optimum number of option in each item of MCQ in the field of medical education.

Objectives: To compare three options, four options, and five options MCQs test for the quality parameters – reliability, validity, item analysis, distracter analysis, and time analysis.

Materials and Methods: Participants were 3rd semester M.B.B.S. students. Students were divided randomly into three groups. Each group was given one set of MCQ test out of three options, four options, and five option randomly. Following the marking of the multiple choice tests, the participants’ option selections were analyzed and comparisons were conducted of the mean marks, mean time, validity, reliability and facility value, discrimination index, point biserial value, distracter analysis of three different option formats.

Results: Students score more (P = 0.000) and took less time (P = 0.009) for the completion of three options as compared to four options and five options groups. Facility value was more (P = 0.004) in three options group as compared to four and five options groups. There was no significant difference between three groups for the validity, reliability, and item discrimination. Nonfunctioning distracters were more in the four and five options group as compared to three option group.

Conclusion: Assessment based on three option MCQs is can be preferred over four option and five option MCQs.

Key words:
Distracter analysis, item analysis, multiple choice questions, number of options per item

Evaluation is the most important component of teaching-learning set of courses.¹ In medical education; Miller’s pyramid is the most persisting frames for evaluation process.² The base of pyramid is formed by knowledge and understanding and top of pyramid is formed by testing performance in artificial settings and in realistic situations.³ It is very difficult to climb up the pyramid without completing the base, so knowledge forms the foundation of pyramid.⁴ Testing of knowledge through different evaluation methods is an important component of medical education system. There are mainly two types of evaluation methods-written assessment methods and viva voce. Written assessment methods are very popular because they are easy to conduct and administer even in large scale of examinees and with cost effectiveness.⁵ However, the important issue is element of subjective bias.⁶ To overcome this issue, objectivizing evaluation is emphasized for both summative and formative assessment.⁷ Multiple choice questions (MCQs) are one of the most popular written objective assessment methods.⁸ Properly constructed MCQs can assess higher cognitive processing of bloom’s taxonomy in a large group of examinees in a short period of time. Other advantages associated with the MCQs are high reliability, more content coverage, and capacity to discriminate between high and low achievers.⁹,¹⁰ Constructing good MCQ item is a complex, challenging, and time consuming process particularly for finding

How to cite this article: Vegada B, Shukla A, Khilnani A, Charan J, Desai C. Comparison between three option, four option and five option multiple choice question tests for quality parameters: A randomized study. Indian J Pharmacol 2016;48:571-5.
plausible distracters.\cite{8,9} MCQs offer another disadvantage that chances of guessing increases with the decrease in number of options.\cite{9} There are various types of MCQs used in the assessment in education-single best option type/multiple option, extended matching items, patient vignettes, etc., Single best response MCQs can also be divided further based on number of options such as three responses, four responses, and five responses.\cite{10}

Argument related to the optimal number of options in MCQs has lasted for more than 80 years.\cite{10,11} Most of the academic teachers use four or five options in summative or formative assessment because they believe that it improves psychometric quality by limiting the effect of random guessing.\cite{10,11} Traditional four or five options formats have one important drawback, i.e. alternative plausible distracters become difficult to construct.\cite{12} Limiting the number of options enhance item writing by preparation of fewer distracters and save examinee’s time to complete exam.\cite{12,13} In the case of three option MCQs, more questions can be accommodated in the test which can enhance the reliability of the test by more coverage of subject.\cite{13} The optimal number of options in MCQ test is still a matter of considerable debate, but its solution particularly in medical education demands further evidence due to less research work of this field. This study was designed with the aim of comparing three option, four option, and five option MCQs test for the quality parameters – reliability, validity, item analysis, distracter analysis, and time analysis.

**Materials and Methods**

The present study is a randomized study to compare the quality of three, four, and five option MCQ test. A written informed consent of participation and approval of Institutional Ethics Committee (Ethics/Approval/2016/1) were obtained.

**Participants**

This study was planned for the second professional M.B.B.S. students. The topic chosen for the MCQ test was autonomic nervous system (ANS). The students were informed about the date of the examination 15 days prior to the examination. The students’ participation in the study was informed and voluntary.

According to the university marks obtained in first professional examination, the students ($n = 132$) were divided into three groups of high, mid, and low achievers. Then by using randomization software, three groups (Groups A, B, C) were created from the equal selection from the high, mid, and low achievers groups. After randomization, scores of first professional university examination were compared between three groups by one-way analysis of variance (ANOVA). There was no significant difference between three groups.

**Instrument**

The subject experts (Investigator 1, 2) prepared a 30 item MCQ test initially with five options. Keeping the order of the questions same, one distracter was excluded randomly from the distracters to obtain a four option test. Similarly, using the five option questions, two distracters were excluded randomly from the distracters to obtain a three option test. Exclusion of two options in Group A would give the test with three options, exclusion of one option in group B would give the test with four options, and group C with original five options test.

The prepared sets were sent to ten subject experts for validation. Seven members responded and their suggestions were incorporated to finalize the three sets. Total time of the test was decided 45 min based on the opinion of subject expert.

**Examination Administration**

On the day of the examination, the three groups were seated in three different rooms and arrangements were made to ensure similar environment to each group. One student was absent from Group A. All students were informed that they could return papers if completed before time. For all students, the time taken to complete the test was noted at the time of submission of the question paper by an investigator.

**Statistical Analysis**

The descriptive statistics was reported in the form of mean, standard deviation, confidence interval, median, frequency, and percentages. Distribution of data was checked by Shapiro–Wilk test and Kolmogorov–Smirnov test. Concurrent validity was checked by Pearson correlation Coefficient. Comparisons of correlation coefficients between three groups were done by online available software (http://www.vassarstats.net/rdiff.html). Reliability was measured by Cronbach’s alpha coefficients. The comparisons of Cronbach’s alpha coefficients between three groups were done by online available software Croncor comparing Cronbach’s alphas (http://www.comparingcronbachalphas.org/). Calculation of facility value, discrimination index, and point biserial correlation coefficient was done according to standard formulas.\cite{14} Comparison of marks, time, facility index, discrimination index, and Z-transformed point biserial correlation coefficient between three groups was done by one-way ANOVA followed by Tukey’s post hoc test. Distracter analysis was conducted according to standard formulas by Microsoft Office Excel 2007.\cite{15} Except the specific calculation done by online software mentioned above, other comparisons and analysis were done by SPSS Statistics for Windows, version 17.0., SPSS Inc., Chicago, IL, USA.

**Results**

**Characteristics of Scores and Time Taken by Examinees in Three Multiple Choice Question Groups (Based on the Number of Options)**

In this study, it was found that the difference between mean scores of three groups was significant ($P = 0.000$). Post hoc analysis showed that marks obtained in three option MCQ test was significantly more as compared to four option and five option test [Table 1]. Mean time taken by examinees was 32.80, 35.98, 35.65 min in three option, four option, and five option groups, respectively. Mean time was significantly less in three option MCQ test as compared to 4 and five option MCQ test.

**Concurrent Validity and Reliability of Three Multiple Choice Question Groups**

For the concurrent validity, correlation between internal test score and score obtained in three, four, five option groups were all “moderate” and statistically significant. These correlation coefficients were not significantly different from...
Vegada, et al.: Comparison of three option, four option and five option multiple choice question tests

Table 1: Descriptive statistics of the scores obtained in multiple choice question tests

|                     | Three-option group (n=44) | Four-option group (n=44) | Five-option group (n=43) | P       |
|---------------------|---------------------------|--------------------------|--------------------------|---------|
| Mean (95% CI)       | 16.64 (15.41-17.86)       | 14.52 (13.18-15.86)      | 11.67 (10.22-13.12)      | 0.000*  |
| SD                  | 4.02                      | 4.39                     | 4.71                     |         |
| Median              | 16.00                     | 15.00                    | 10.00                    |         |
| Range of scores     | 8.00-25.00                | 6.00-27.00               | 4.00-23.00               |         |
| CV                  | 0.24                      | 0.30                     | 0.40                     |         |
| Skewness            | 0.10                      | 0.20                     | 0.60                     |         |
| Kurtosis            | −0.16                     | 0.38                     | −0.42                    |         |

n=number of students one-way ANOVA followed by the Tukey test *P<0.05 between three option and four option group, between three option and five option group. CI=Confidence interval, SD=Standard deviation, CV=Coefficient of variation

Table 2: Comparison of validity and reliability between three multiple choice question tests

|                     | Three-option group (n=44) | Four-option group (n=44) | Five-option group (n=43) |
|---------------------|---------------------------|--------------------------|--------------------------|
| Validity            |                           |                          |                          |
| Mean MCQ test score (out of 30) | 16.64              | 14.52                     | 11.67                     |
| Mean internal score (out of 40) | 16.36                 | 16.89                     | 16.19                     |
| Correlation coefficient (r, P) | 0.407, <0.01          | 0.492, <0.01              | 0.550, <0.01              |
| Reliability         | Cronbach’s α            | 0.61                      | 0.67                      | 0.75*   |

Table 3: Comparison of item analysis between three multiple choice question tests

|                     | Three-option group (n=30) | Four-option group (n=30) | Five-option group (n=30) | P       |
|---------------------|---------------------------|--------------------------|--------------------------|---------|
| Item difficulty     |                           |                          |                          |         |
| (facility index value)  | 55.45 (17.34)            | 48.48 (19.76)            | 39.05 (19.09)            | 0.004*   |
| Item discrimination |                           |                          |                          |         |
| (discrimination index) | 0.21 (0.16)              | 0.23 (0.12)              | 0.26 (0.16)              | 0.455   |
| Item discrimination |                           |                          |                          |         |
| (point biserial correlation coefficient) | 0.29 (0.17) | 0.33 (0.13) | 0.34 (0.19) | 0.994  |

n=Number of students, MCQ=Multiple choice question

Discussion

This study was done with the aim of comparing three options, four options, and five options MCQ test for the quality of assessment. It was observed that student scores more and took less time for completion of three options as compared to four options and five options groups. Facility index was more in three options group. There was no significant difference between three groups for the validity, reliability, and item discrimination. Nonfunctioning distracters were more in the four and five options group as compared to three option group.

For scoring pattern, there was progressive reduction of scores from three options to five options groups. Such type of scoring pattern may reflect more chances of guessing with fewer distracters, so issue of guessing may be raised if MCQs with less number of options are used. For blind guessing in MCQ test, quality of options would be responsible rather than number of options. Meta-analysis by Rodriguez suggested that examines might be engaged in the educated guessing rather than blind guessing when elimination of least plausible distracter would be done. Because of our study design, we eliminated distracters in random manner, so this significant difference might be due to the method of elimination of distracters and quality of distracters.

This study showed that mean time taken by examinees in three option group was less than mean time taken by examinees in four and five option groups. Hence, mean time per item decreased with decrease in number of alternatives per item. Overall saving time per item in three option groups was 6 s
as compared to four option and five option groups. This time difference would allow completion of more questions in time limited three option MCQs exams which will increase the validity.\cite{11,12,13} Fewer options also have distinct advantages for test takers that they need less time for item development and item administration.\cite{1,11}

In a manner consistent with findings of previous research, our study showed that there was no significant difference in reliability and validity between three option, four option and five option group.\cite{17,18} Haladyna and Downing showed that reliability increased with increased number of options, but this difference was small when more than three options were used.\cite{19} As per Rodriguez, moving from five options to four options items reduces reliability by 0.35 on average and moving from five options to three options items does not affect reliability.\cite{10} Change in reliability also depends on distracter elimination method. For random distracter elimination, reliability drops 0.6 on average and no change in reliability if ineffective distracters are deleted.\cite{10} Vyas and Supe concluded in literature review that reliability might be increased with three option format because total number of items increased to balance decreased number of alternatives.\cite{1}

Mean item facility values in all three different option groups had an acceptable level, but there was significant difference in facility index values between three option and five option groups.\cite{14} Rodriguez suggested that a small increase in facility index was observed when reducing the options from four to three, while there was a large increase in the facility index (items were easier) if the options were reduced to two.\cite{10} In this study, small increase in facility value was observed when reducing from five options to four options, but there was a large increase in facility value if the options were reduced from five to three. Previous studies concluded that there was no significant difference in item facility after reduction of number of options, but reducing the number of options may increase the probability of random guessing to choose correct answer.\cite{15,17,18,20} The reason for such findings in previous researches may be because of use of nonfunctional distracter elimination method to minimize the effects on difficulty level. Motivated examinees rarely resort to random guessing when they have sufficient time and appropriate difficulty level.\cite{11} In this study, significant difference between three and five options facility values might be due to option deletion method. In this study, all the test groups had mean discrimination index values and point biserial values within acceptable range and there were no significant difference in discrimination index and point biserial values and also consistence with previous studies.\cite{17,18}

For the distracter analysis, in this study, we found that most of the questions had two functioning distracters irrespective of number of choices. Only one item from five option group had four functioning distracters. These findings are consistence with other studies despite of our method for option deletion was randomized process rather than removal of nonfunctioning distracters.\cite{15,18} Haladyna and Downing also concluded that approximately two-thirds of all four option items had only one or two functioning distracters and none of the five option items had four functioning distracters.\cite{10} It is difficult for the teachers to develop three or more equally plausible distracters, additional distracters have role of “fillers” only and also fail to discriminate between poor and good examinees.\cite{20} Contrary to other studies, in this study, we removed 1 and 2 distractor randomly from five option MCQ test to prepare four option and three option MCQ test respectively. There are chances that some good distractor may have been removed during this process. Here, it should be noted that in the case of randomization both known and unknown confounders distributed equally in all groups. Hence, even if there is chance that good distractor may be removed, still random removal is better because the chance will be equally distributed in all groups and will not affect the results and conclusion of the study. Some other methods like “Horst statistics” could be used for the distractor analysis but the calculation of distractor analysis done by us is adequate as per the standard criteria and available published studies.

Result from our study highlights that comparison of score variability, time, validity, facility value, discrimination ability, and distracter performance produced only minor differences on the impact of change of number of options in MCQ test in pharmacology assessment [Table 5].

Table 4: Comparison of distracter analysis in three multiple choice question tests

| Distracters | Three-option group (n=30) | Four-option group (n=30) | Five-option group (n=30) |
|-------------|---------------------------|--------------------------|--------------------------|
| Number of distracters | 60 | 90 | 120 |
| With frequency <5% | 2 (3.33) | 15 (16.67) | 22 (18.33) |
| Discrimination ≥0 | 14 (23.33) | 35 (38.89) | 45 (37.50) |
| Both | 1 (1.66) | 13 (14.44) | 18 (15.00) |
| Functioning distracters per test | 45 (75.00) | 53 (58.89) | 71 (59.16) |
| Functioning distracters per item | One | 15 (50.00) | 11 (36.67) | 4 (13.33) |
| Two | 15 (50.00) | 15 (50.00) | 12 (40.00) |
| Three | NA | 4 (13.33) | 13 (43.33) |
| Four | NA | NA | 1 (3.34) |
| Mean functioning distracter (SD) | 1.50 (0.51) | 1.77 (0.67) | 2.37 (0.76) |

n=Number of items, NA=Not applicable, SD=Standard deviation

Table 5: Summary of quality parameters of multiple choice question test analysis

| Parameter | Result |
|-----------|--------|
| Characteristic of scores | Mean score obtained in three option group was more than four and five option groups |
| Distribution of scores | Normal distribution followed in all three groups |
| Characteristic of time | Mean time taken by three option group was less than four and five option groups |
| Validity | No significant difference in three groups |
| Reliability | No significant difference in three groups |
| Facility value | Mean facility value was more in three option groups as compared to five option groups |
| Discrimination index | No significant difference in three groups |
| Point biserial coefficient | No significant difference in three groups |
| Functioning distracters | All three groups have two functioning distracters per item |
There are very few studies in medical specialty, especially in pharmacology subject regarding the influences of number of options on teacher generated tests, so we conducted prospective and randomized study to minimize any bias for the quality parameters of MCQs. However, the study was conducted on the second year M.B.B.S. students for ANS pharmacology assessment, and the sample size was also less, the result of this study should be interpreted cautiously for other specialty, subject, and content. We do acknowledge that more items will increase the reliability of the study. Choice of 30 items was based on real life situations where around 20–30 MCQs are used for internal assessment. The MCQ item list was also sent to few experts and all were agreed on number of items. Reliability obtained in the study was also found to be in normal range showing the number of MCQs were adequate. We could not found any association or guideline which relates the number of items with number of students. We could have analyze each items for all parameters of item analysis but our objectives was to analysis three MCQ tests based on three, four, and five option items so for this study that comparison of whole test is more useful looking at our objectives.

**Conclusion**

Assessment based on three option MCQs can be preferred over four option and five option MCQs.

**Financial Support and Sponsorship**

Nil.

**Conflicts of Interest**

There are no conflicts of interest.

**References**

1. Vyas R, Supe A. Multiple choice questions: A literature review on the optimal number of options. Natl Med J India 2008;21:130-3.  
2. Peile E. Knowing and knowing about. BMJ 2006;332:645.  
3. Anshu. Assessment of knowledge: Written assessment In: Singh AT, editor. Principles of Assesment in Medical Education. 1st ed. New Delhi: Jaypee Brothers Medical Publishers (P) Ltd.; 2012. p. 70-9.  
4. Siddiqui NI, Bhavsar VH, Bhavsar AV, Bose S. Contemplation on marking scheme for Type X multiple choice questions, and an illustration of a practically applicable scheme. Indian J Pharmacol 2016;48:114-21.  
5. Karelia BN, Pillai A, Vegada BN. The levels of difficulty and discrimination indices and relationship between them in four-response type multiple choice questions of pharmacology summative tests of year II MBBS students. IeJSME 2013;6:41-6.  
6. Panczyk M, Rebandel H, Gotlib J. Comparison of four and five option multiple choice questions in nursing entrance tests. ICERI 2014 Proceedings; 2014. p. 4131-8.  
7. Mukherjee P, Lahiri SK. Analysis of multiple choice questions (MCQs): Item and test statistics from an assessment in a medical college of Kolkata, West Bengal. IOSR J Dent Med Sci 2015;1:47-52.  
8. Dehnad A, Nasser H, Hosseini AF. A comparison between three-and four-option multiple choice questions. Procedia Soc Behav Sci 2014;98:398-403.  
9. Kilgour JM, Tayyaba S. An investigation into the optimal number of distractors in single-best answer exams. Adv Health Sci Educ Theory Pract 2016;21:571-85.  
10. Rodríguez MC. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. Educ Meas Issues Pract 2005;24:3-13.  
11. Seinhorst G. Are three options better than four. Lancaster University; 2008.  
12. 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;9:40.  
13. Haladyna TM, Downing SM, Rodríguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. Appl Meas Educ 2002;15:309-33.  
14. Ciraj AM. Item analysis and question banking. In: Singh AT, editor. Principles of Assesment in Medical Education. New Delhi: Jaypee Brothers Medical Publishers (P) Ltd.; 2012. p. 116-27.  
15. Schneider SD, Armour C, Park YS, Yudkowski R, Bordage G. Reducing the number of options on multiple-choice questions: Response time, psychometrics and standard setting. Med Educ 2014;48:1020-7.  
16. Bagnato F, Butman JA, Mora CA, Gupta S, Yamano Y, Tasciyan TA, et al. Conventional magnetic resonance imaging features in patients with tropical spastic paraparesis. J Neuropsychiatr Clin Neurosci 2005;11:525-34.  
17. Thanyaporn P, Butmang JA, Mora CA, Gupta S, Yamano Y, Tasciyan TA, et al. Conventional magnetic resonance imaging features in patients with tropical spastic paraparesis. J Neuropsychiatr Clin Neurosci 2005;11:525-34.  
18. Lee H, Winke P. The differences among three-, four-, and five-option-item formats in the context of a high-stakes English-language listening test. Language Testing 2013;30:99-123.  
19. Haladyna TM, Downing SM. A Quantitative Review of Research on Multiple-Choice Item Writing. 1985. Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, 31 March-4 April, 1985). Available from: http://www.eric.ed.gov/?id=ED255580. Last accessed on 2016 May 01].  
20. Shizuka T, Takeuchi O, Yashima T, Yoshizawa K. A comparison of three- and four-option English tests for university entrance selection purposes in Japan. Language Testing 2006;23:35-57.