



The Optimal Number of Choices in Multiple-Choice Tests: A Systematic Review

Zahra Emami Moghadam¹, Alireza Akhlaghi², *Roghaie Khoshkholgh³

¹Assistant Professor, Nursing and Midwifery Care Research Center, Mashhad University of Medical Sciences, Mashhad, Iran.

²MSc Student of Health Education and Promotion, Yasuj University of Medical Sciences, Yasuj, Iran.

³Department of Midwifery, Firoozabad Branch, Islamic Azad University, Firoozabad, Iran.

Abstract

One of the long-term issues faced by test designers and educators is the appropriate number of choices in a multi-choice test. Currently, the usual number of options for multiple-choice questions in the medical field are three to five, and it is generally believed that more options are better. Numerous theoretical and empirical studies have provided evidence in favor of using three-choice questions. According to the findings of these studies, the psychometric properties of three-choice questions are similar to four- or five-choice questions and the validity and reliability of the test or the coefficients of difficulty and differentiation do not change significantly with decreasing the number of options. Therefore, reducing the number of questions can reduce the time needed to design tests and take exams, saving the time and energy of the faculty and students. Most studies have concluded that it is cost-effective to use a three-choice question if it does not change the psychometric properties of the test by reducing the number of options.

Key Words: Exam, Question, Number of Options, Faculty, Student.

*Please cite this article as: Emami Moghadam Z, Akhlaghi A, Khoshkholgh R. The Optimal Number of Choices in Multiple-Choice Tests: A Systematic Review. Med Edu Bull 2021; 2(3): 253-60. DOI: 10.22034/MEB.2021.311998.1031

*Corresponding Author:

Roghaie Khoshkholgh, Department of Midwifery, Firoozabad Branch, Islamic Azad University, Firoozabad, Iran.

Email: roghaiekhoshkholgh@gamil.com

Received date: Aug. 15, 2021; Accepted date: Sep.12, 2021

1- INTRODUCTION

Multiple-choice questions are among the most common questions used in tests. With multiple-choice questions, as many choice options can be offered as needed, but at least two choices must be provided. Each multiple-choice question has the option for multiple answers, randomizing the order of the answers and including a write-in option as one of the choices (1-3). An important question is the optimum number of choices per multiple-choice question. In other words, how many options are needed to take into account the points under discussion while maintaining the test quality. The use of four- and five-choice questions is common in important medical science tests worldwide. Also, most of the available question design guides and instructions are based on four- and five-choice questions. The reason for this is unclear. It could be attributed to the fact that education and exam officials are unaware of the studies conducted on the subject. It has also been suggested that three-choice questions are not used in the medical field as they will increase the probability of guessing the correct answer. Professors usually believe that three-choice questions increase the chance of successful guessing (4-8).

At present, three- to five-choice questions are commonly used for multiple-choice questions in the field of medical sciences, and there is a general belief that with more options, the students' grades will be closer to reality. However, the results of different studies are contradictory. Although there have been many relevant studies, there still is no definitive answer in this regard and it remains a challenge in the field of multiple-choice questions. The aim of the present study is to review the relevant studies and identify the optimum number of options for multiple-choice questions.

2- MATERIALS AND METHODS

2-1. Data sources

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist was used as a template for this review. A systemic search of electronic databases Medline (via PubMed), SCOPUS, Web of Science, ProQuest, Cochrane Library SID, Magiran, CIVILICA, and Google Scholar search engine was conducted with no time limit up to August 2021. The search was performed independently and in duplicate by two reviewers and any disagreement was solved by the supervisor.

2-2. Included studies

Review articles, systematic reviews, case-control studies, cross-sectional studies, qualitative studies, and descriptive and analytical studies were included in the study. Pilot, preliminary, and case report studies were not included due to the limited sample size and higher risk of bias.

2-3. Study selection

Database search was performed for the relevant studies, abstracts of the studies were screened for identification of eligible studies, full-text articles were obtained and assessed, and a final list of included studies was made. This process was performed independently and in duplicate by two reviewers and any disagreement was resolved by a third reviewer. References were organized and managed using EndNote software (version X8).

2-4. Data collection process

A form was developed and followed for each study. The data collected by the two reviewers was combined and compared for accuracy and any discrepancies were solved by a third reviewer.

2-5. Risk of bias

The assessment of the risk of bias was done based on the Cochrane Risk of Bias criteria (9) by two reviewers independently and in duplication and any discrepancies were resolved by a third reviewer.

2-6. Synthesis of results

Due to the difference in the included studies, study designs, lack of control groups in some studies, small sample size, and type of intervention used, and duration of treatment and follow-up, meta-analysis was not conducted.

3- RESULTS

Finally, 23 related articles were selected. Traditionally, multiple-choice test items (MCIs) have been offered with four or five answer options, and measurement textbooks have recommended this. Recently, however, many studies have theoretically and empirically found that three options are just as effective, and three may be the optimal number of options for MCIs.

In a study by Costin (1970), the characteristics of tests composed entirely of two-, three- or four-choice items were investigated given a fixed total number of alternatives across the whole test (Tversky's condition). The results showed that the number of alternatives per item had an inverse relationship with item difficulty but was directly related to item discrimination. Reliability and standard error of measurement of three-choice item tests were equivalent or superior to tests of four- or two-choice items and these results held up after taking into account the testing time (10).

Ralph G. Straton et al. (1980) designed an experiment to further test the efficacy of three-option achievement items. Parallel tests of three- and five-option items were created and distributed randomly to college students. Results showed no differences in mean item difficulty, mean discrimination, or total test score, but a substantial reduction in time spent on three-option items (11). Results of Budescu et al.'s study (1985) showed that a strong and consistently negative relationship between the rate of

performance and the number of options was detected in all tests (12). Owen et al. (1987) designed an experiment to further test the efficacy of three-option achievement items. Parallel tests of three- and five-option items were created and distributed randomly to college students. Results showed no differences in mean item difficulty, mean discrimination, or total test score, but a substantial reduction in time spent on three-option items. It means that content validity may be boosted by writing additional three-option items to tap more content (13). Haladyna et al. (1989) conducted a study on four standardized tests (set of 477 options). The results showed that 38% of the options attracted less than 5% of students (non-functional options) (14).

In a study by Andres et al. (1990), the researchers demonstrate that under certain very general conditions, it is better to set questions with $k = 3$ alternative answers, whilst under other criteria, the value $k = 2$ is to be preferred (15). In a study by Haladyna et al. (1993), results from three different testing programs (three, four, and five options) support the conclusion that test items seldom contain more than three useful options. Therefore, testing program personnel and classroom teachers may be better served by using two- or three-option items instead of the typically recommended four- or five-option items (16). In another study by R. Eric Landrum et al. (1993), students from two consecutive semesters were given multiple-choice tests over five units of an undergraduate psychology course. During the first semester, students were given five 50-question four-option multiple-choice tests, and during the second semester, students were given five 50-question three-option multiple-choice tests. Results indicate that students performed significantly better on three-option items than on four-option items (corrected for chance guessing) (17).

In a study conducted by Michael S. Trevisan et al. (1994), a two-option test was constructed, and options were systematically added to the test by using a taxonomy of item writing rules to guide the process of distractor development. No significant differences ($p=2 .05$) were found among the reliability coefficients, the reliability estimates for the three-, four- and five-option formats being all on the same order of magnitude. These findings continue to provide evidence for the efficacy of the three-option questions (18). In a study by Sidick et al. (1994), it was suggested that given the similarity of the psychometric properties and the likely reduction in the cost of development and administration time, three-alternative multiple-choice items may be preferable to five-alternative multiple-choice items for some testing purposes (19).

James E. Bruno et al. (1995) showed that given a fixed total number of alternatives for a multiple-choice type test, the use of three alternatives at each choice point will maximize discriminability, power, and informativity of a test (20). W. Todd Rogers and Dwight Harley (1999) showed that tests consisting of three-option items are at least equivalent to tests composed of four options in terms of internal consistency score reliability. The difficulty is inversely related to the number of options, and the findings for item discrimination are not conclusive (21).

A Meta-analysis of 80 years of research (1920 - 2000) reviewed 27 studies and concluded that reducing the number of options increases the coefficient of difficulty. The changes were less than four to three and more than three to two if the number of options was reduced. Also, reducing the number of options reduced the coefficient of differentiation and reduced the reliability, except in the case of reducing the options from four to three, which increases them slightly. Reducing the number of options did not change the

validity. The review showed that when the number of options decreased from four to three, the difficulty factor increased slightly (i.e., the test became easier) and when the number of options was reduced to two, the difficulty factor increased significantly (i.e., the test became very easy) (21). Vyas et al. (2008) studied the number of suitable options in multiple-choice exams. After reviewing 23 articles, they recommended the use of three-choice questions for medical exams. Their conclusion was that the three-choice tests are more efficient and easier to run than the four- and five-choice tests, and that it takes less time to write the options, read the options, and run the test (23).

In a study by Tarrant et al. (2010), a test consisting of 50 four-choice questions was taken from 36 undergraduate nursing students. After analyzing the test, the non-functional option of each question was removed and a three-choice test was created from which 41 questions were selected and used for another group of nursing students the following year. Comparing the two tests, the researchers concluded that the mean and range of score changes in the two years were almost the same, but the acceptance rate was slightly lower in the second year. The first test had a higher difficulty coefficient and was easier ($P > 0.05$). In terms of differentiation, the second-year questions had a larger clean coefficient but was not significant. The three-choice test was better than the four-choice test in terms of reliability (24).

In another study, Baghaei et al. (2011) aimed to determine the optimal number of options for multiple-choice test items. They investigated the amount of time and energy needed for developing multiple-choice tests with more distracters and concluded that three options per item are optimal (25). Thanyapa et al. (2014) pointed out that three-option questions are just as reliable and valid as four- or five-

option questions (26). Dehnad et al. (2014) aimed to compare three-option MCQs with four-option in terms of test usefulness. The results suggested that three-option MCQs save time for covering more content and items in the test, thereby increasing test validity and reliability. Three-option MCQs seem to be easier and less demanding for novice and inexperienced teachers as they would not be forced to use implausible and defective distracters when writing items. Three-option MCQs are recommended for novice teachers, classroom-based tests, and achievement tests requiring more content to be covered in a short period of time whereas four-option MCQs could be more appropriate for high-stake tests or whenever it is functional or feasible (27).

Schneid et al. (2014) aimed to examine the effects of reducing four- and five-option MCQs to three-option MCQs on response times, psychometric characteristics, and absolute standard-setting judgments in a pharmacology examination administered to health professions students. Results showed that students answered three-option MCQs an average of five seconds faster than they answered four- and five-option MCQs (36 seconds versus 41 seconds; $p = 0.008$), and the cut scores generated for three-option MCQs using the three-level Angoff (TLA) ratings were eight percentage points higher ($p = 0.04$) (28). Sadeghi et al. (2017) aimed to compare Item Facility and Item Discrimination of MC vocabulary test items and attempted to find whether these indices are affected by the number of options. To this end, four 20-item stem equivalent vocabulary tests (3-, 4-, 5-, and 6-option MC) were administered to 194 pre-intermediate students. Results suggested that the six-option test is the most difficult. Also, the results of the questionnaire revealed test takers' preference towards the use of three-option MC. Their findings demonstrated that

increasing the number of options makes a test more difficult and that choosing the right number of options for MC tests is controversial. Testers are recommended to consider various factors while choosing the right number of options (29). Results of a study by Loudon et al. (2018) showed that four-choice questions and more extend the exam time while having no advantages compared to three-choice questions (30). In contrast, Raymond et al. (2019) found little support for the conventional practice of five options, challenged the recommendation from previous research that three options are optimal, and proposed the continued use of the four-option format (31). Tweed (2019) suggested that options should not be limited to three, four, or five, and other options should be considered as well (32).

4- DISCUSSION

The aim of the present study was to review the relevant studies and to identify the optimal number of options for multiple-choice questions. The results of the present study showed that the three-choice test was better than the four- and five-choice tests in distinguishing between strong and weak students and the psychometric properties of the three-choice questions are similar to the four- or five-choice questions, and that with decreasing the number of options, the validity and reliability of the test or item difficulty and discrimination do not change significantly. Also, a shorter time period is needed for writing and reading options and taking the test. There is a general belief that the more options there are for a question, the better the reliability outcomes will be and the fewer students can choose the correct answer by guessing. However, the results of relevant studies are contradictory. Researchers use different methods to investigate the difference between different numbers of multiple-choice questions. One of these methods is to calculate the attraction percentage of

each option. If a small number of students choose an option in a question, it is called a non-functioning option. Insisting on designing multiple options sometimes leads to non-functional options (14, 24). The question is if the designed options do not produce their intended effect, then why test makers insist on designing items with four or five options. Designing distractors is one of the most difficult steps in creating a multiple-choice question. These options, while not correct, should be able to attract those students with insufficient information. The higher the number of options, the more difficult it is to design distractors. Therefore, it is desirable to achieve the evaluation objectives using a smaller number of options.

On the other hand, the number of options is one of the factors that affect the test time, because reading each option takes time. If the number of options is reduced, students will answer the questions within a shorter time period. Therefore, the number of questions can be increased in the same period of time, thus improving the validity and reliability of the test. On the other hand, if the number of options is reduced, the question designers will feel less stressed, and thus fewer mistakes will be committed. Moreover, faculty members, instead of spending their time designing distractors, will spend the same time designing fewer options but with better quality. Grier et al. (1975) used mathematical formulas to prove that the reliability of a three-choice test is greater when the total number of options in the two tests is the same (23). The results of a review article showed that reliability increases with increasing options, although such an increase was very small when more than three options were used (14). Another meta-analysis showed that reducing the number of options from four to three leads to a slight increase in the reliability coefficient (22). Results of another study (1994) showed that three-

choice questions improve the test validity (33). Another study (1987) also concluded that the three-choice item improves the test validity. Based on the results of the above study, reducing the number of options from four to three in a 100-question test reduced the test time by 17% (34). In a meta-analysis, Rodriguez et al. (2005) found that reducing the number of options did not change the test validity (22). According to the Vyas review study, there was no difference between three- and five-choice questions or between three- and four-choice questions in terms of item difficulty (23). In a review study, Haladyna et al. also showed no significant difference between the three- and four-choice tests in terms of item difficulty (14). In a meta-analysis, Rodriguez et al. showed that when the number of options decreased from four to three, the difficulty coefficient of the item increased slightly (i.e., the test became easier) and when the number of options reached two, the difficulty coefficient of the item increased significantly (i.e., the test became very easy) (22).

However, four- and five-choice questions are still commonly used in important medical science exams worldwide. The reason for this is unclear. It could be attributed to the fact that education and exam officials are unaware of the studies conducted on the subject. It has also been suggested that three-choice questions are not used in the medical field as they might increase the probability of guessing the correct answer. Professors usually believe that three-choice questions increase the chance of successful guessing. However, in a comparative study of three- and four-choice tests, Rogers et al. (1999) showed that the rate of guessing and using answer-guessing techniques was lower in the three-choice question (24). Finally, it seems that more recent studies are trying to address this issue through classical theory and various models of question-

answer theory. Therefore, it can be stated that a definite answer to the question about the optimal number of options in a multiple-choice question is yet to be agreed upon by researchers and used in practice.

5- CONCLUSION

Numerous theoretical and empirical studies have suggested results in favor of using three-choice questions. Based on the findings of these studies, the psychometric properties of three-choice questions are similar to those of four- or five-choice questions, and with decreasing the number of options, the validity and reliability of the test or item difficulty and discrimination do not change significantly. On the other hand, reducing the number of options will reduce the time needed to design questions and arrange an exam.

6- AUTHORS' CONTRIBUTIONS

Study conception or design: ZE, and RK; Data analyzing and draft manuscript preparation: AA, and RK; Critical revision of the paper: ZE, and AA; Supervision of the research: ZE and RK; Final approval of the version to be published: ZE, AA, and RK.

7- CONFLICT OF INTEREST: None.

8- REFERENCES

1. Saeidi M., Alipour-Anbarani M. Students' Challenge in Answer-changing on Multiple-choice Exams; Doubting the Answer or Not? A Systematic Review. *Medical Education Bulletin*, 2020; 1(1): 41-8.
2. Moss E. Multiple choice questions: their value as an assessment tool. *Current Opinion in Anaesthesiology* 2001; 14(6):661.
3. Stirling J. The use of multiple choice tests in the teaching and assessment of undergraduate psychology degree course. *Psychology Teaching Review* 1996; 5(2):83-9.
4. R. McG. Harden, R. A. Brown, L. A. Biran, W. P. Dallas Ross, R. Multiple choice questions: to guess or not to guess. *Medical Education*. 1976;10 (1): 27-32.
5. E. J. Hammond, A. K. McIndoe, A. J. Sansome, P. M. Spargo, Multiple-choice examinations: adopting an evidence-based approach to exam technique. *Anaesthesia*, 1998; 53(11): 1105-8.
6. Budescu, D., Bar-Hillel, M. To guess or not to guess: A decision-theoretic view of formula scoring. *Journal of Educational Measurement*, 1993; 30(4): 277.
7. Thomas J. Prihoda, R. Neal Pinckard, C. Alex McMahan, Anne Cale Jones., Correcting for Guessing Increases Validity in Multiple-Choice Examinations in an Oral and Maxillofacial Pathology Course. *J Dent Educ*. 2006; 70(4): 378-86.
8. Jallili M, Khabaz Mafinejad M, Gandomkar R, Mortaz Hejri S. Principles and methods of student assessment in health professions. *The Academy of Medical Sciences Islamic Republic of Iran*. 2017; Tehran. ISSN: 978-600-6734-07-1.
9. STROBE statement. Available at: <https://www.strobe-statement.org/index.php?id=available-checklists>.
10. Costin, F. The optimal number of alternatives in multiple-choice achievement tests: Some empirical evidence for a mathematical proof. *Educational and Psychological Measurement*, 1970; 30: 353-58.
11. Ralph G. Straton, Ralph M. Catts. A Comparison of Two, Three and Four-Choice Item Tests Given a Fixed Total Number of Choices. *Educational and Psychological Measurement* 1980; 40(2): 357-65.
12. Budescu, D. V. and Nevo, B. Optimal number of options: An investigation of the assumption of proportionality. *Journal of Educational Measurement*, 1985; 22: 183-16.
13. Owen, S. V. and Froman, R. D. What's wrong with three-option multiple-choice items? *Educational and Psychological Measurement*, 1987; 47: 513-22.
14. Haladyna TM, Downing SM. *Applied measurement in education*, 1989 - Taylor & Francis. 1989; 2(1): 51-78.

15. Andrés, A. M., & Luna del Castillo, J. D. Multiple choice tests: Power, length and optimal number of choices per item. *British Journal of Mathematical and Statistical Psychology*, 1990; 43(1):57-71.
16. Haladyna TM, Downing SM. How Many Options is enough for a Multiple-Choice Test Item? *Educational and Psychological Measurement*. 1993;53(4):999-1010.
17. R. Eric Landrum, Jeffrey R. Cashin, Kristina S. Theis. More Evidence in Favor of Three-Option Multiple-Choice Tests. *Educational and Psychological Measurement*. 1993; 53 (3): 771-8.
18. Michael S. Trevisan, Gilbert Sax, Willilam B. Michael. Estimating the Optimum Number of Options per Item Using an Incremental Option Paradigm. *Educational and Psychological Measurement*, 1994; 54(1): 86.
19. Sidick, J. T., Barrett, G. V., & Doverspike, D. Three-alternative multiple choice tests: An attractive option. *Personnel Psychology*, 1994;47(4): 829-35.
20. Bruno, J. E., Dirkzwager, A. Determining the optimal number of alternatives to a multiple-choice test item: An information theoretic perspective. *Educational and Psychological Measurement*, 1995;55(6): 959.
21. Rogers, W. T., & Harley, D. An empirical comparison of three- and four-choice items and tests: Susceptibility to testwiseness and internal consistency reliability. *Educational and Psychological Measurement*, 1999; 59(2): 234-47.
22. Michael C. Rodriguez. Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research. *Educational Mesurment* 2005; 24(2): 3-13.
23. Vyas R, Supe A. Multiple choice questions: a literature review on the optimal number of options. *Natl Med J India*. 2008 May-Jun;21(3):130-3. PMID: 19004145.
24. 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 Jul 7;9: 40. doi: 10.1186/1472-6920-9-40.
25. Baghaei, P., & Amrahi, N. The effects of the number of options on the psychometric characteristics of multiple choice items. *Psychological Test and Assessment Modeling*, 2011;53(2); 192-211.
26. Thanyapa, I., Currie, M. The number of options in multiple choice items in language tests: does it make any difference? Evidence from Thailand. *Language Testing in Asia* 4, 8 (2014). <https://doi.org/10.1186/s40468-014-0008-7>.
27. Dehnad A, Nasser H, Hosseini AF. A Comparison between Three-and Four-Option Multiple Choice Questions, *Procedia - Social and Behavioral Sciences*, 2014; 98: 398-403. <https://doi.org/10.1016/j.sbspro.2014.03.432>.
28. Schneid SD, Armour C, Park YS, Yudkowsky R, Bordage G. Reducing the number of options on multiple-choice questions: response time, psychometrics and standard setting. *Med Educ*. 2014 Oct;48(10):1020-7.
29. Sadeghi, K., Akhavan Masoumi, G. Does number of options in multiple choice tests affect item facility and discrimination? An examination of test-taker preferences. *Journal of English Language Teaching and Learning*, 2017; 9(19): 123-43.
30. Loudon C, Macias-Muñoz A. Item statistics derived from three-option versions of multiple-choice questions are usually as robust as four- or five-option versions: implications for exam design. *Adv Physiol Educ*. 2018 Dec 1;42(4):565-75.
31. Raymond, M.R., Stevens, C. & Bucak, S.D. The optimal number of options for multiple-choice questions on high-stakes tests: application of a revised index for detecting nonfunctional distractors. *Adv in Health Sci Educ* 2019; 24: 141-50.
32. Tweed, M. Adding to the debate on the numbers of options for MCQs: the case for not being limited to MCQs with three, four or five options. *BMC Medical Education*, 2019;19(1). doi:10.1186/s12909-019-1801-x.
33. Tversky A. On the optimal number of alternatives at a choice point. *Journal of Mathematical Psychology*. 1964; 1(1): 386.
34. Owen SV, Forman RD. What's wrong with three-option multiple choice items? *Educ Psychol Meas*, 1987' 47: 513-22.