Are Faculty Members of Paramedics able to Designed Accurate Multiple Choice Questions?

Reza Pourmirza Kalhori[1] & Mohammad Abbasi[2]

[1] Medical Emergency Department, Faculty of Paramedical, Kermanshah University of Medical Sciences, Kermanshah, Iran

[2] Faculty of Nursing and Midwifery, Qom University of Medical Sciences, Qom, Iran

Correspondence: Mohammad Abbasi, Assistant Professor, School of Nursing and Midwifery, Qom University of Medical Sciences, Qom, Iran, Tel: 98-91-2527-0638. Fax: 98-25-3771-5214. E-mail: mohamad_abbasi55@yahoo.com

Abstract

Background and Objectives: Multiple choice questions (MCQ) are one of the assessment instruments in medical sciences. The overall aim of this study was to perform qualitative and quantitative analysis of the multiple choice question MCQ provided by the professors of Kermanshah University of Medical Sciences-Faculty of Medicine in the academic year 2011-2012.

Materials and Methods: In this descriptive-analytic study, 37 tests of the Faculty of Medicine were analyzed. Quantitative data included difficulty coefficient, discrimination coefficient, whole credibility test, standard deviation of the questions, and the qualitative data consisted of taxonomic percent I, II and III, percentage of questions with no structural problems. The data were analyzed using SPSS software version 20.00 while T-test and chi-square test were applied.

Results: The average validity coefficient of the total tests (KR-20) was measured as 0/63, the average difficulty coefficient as 0/58, the average discrimination coefficient as 0/19. The average percentage of the questions without structural problems as 37.1%; all of which were in the acceptable range. The mean Taxonomy I percentage of the questions was 38/36% ($\pm$11.31), Taxonomy II percentage of the questions was 42.46% ($\pm$15.51) with no significant difference in the entire tests. Average percentage of questions with taxonomy III was 20.73% ($\pm$12.83) for which independent t-test showed significant difference in the total tests (P = 0.00). Average percentage of questions without structural problems was measured as 55.23% ($\pm$13.23) for which there was a significant difference in the total tests when independent t-test was used (P = 0.041).

Conclusion: Considering the average validity of the whole test, the mean difficulty coefficient and Taxonomy indexes I, II and III, the tests designed by the professors of the Faculty of Allied Science are within an a acceptable standard range.

Keywords: multiple choice questions, difficulty index, discrimination coefficient

1. Introduction

The aim of educational activities is changing the behavior of the learners (learning). In order to determine the occurrence of the learning, an assessment of the teaching process is essential. This kind of objective evaluation depends on the aim of the learning activities and also their conditions and is known as Measurement (Newble & Cannon, 2001). Measurement is a process that will determine to what extent a person or an object possesses a value or characteristics. Evaluation simply refers to determination of the value for anything, or value judgment; in the training process this tool is called a test (M. Khan & Aljarallah, 2011). A test must possess three characteristics: validity, reliability and feasibility (H. F. Khan, Danish, Awan, & Anwar, 2013). Tests are classified into two categories known as the teacher made tests and the standardized tests (Briggs, Alonzo, Schwab, & Wilson, 2006; Pourmirza kalhori, Darabi, Roshanpour, Rezaei, 2013).

Multiple choice questions (MCQ) are the most common objective ones. Objective tests such as MCQ or correct–wrong tests are those that if be corrected by different people or at different times lead in the same results (Khan & Aljarallah, 2011). Each MCQ have two parts called the root and the proposed choices among which, one is the right or the most correct answer for the root (Briggs et al., 2006). This type of questions are used to assess personal information, medical decisions, statistical interpretations and mental skills such as recall,

recognition and decoding problems (Shah, Woodward, & Smith, 2013). If the questions are well designed, they can assess high levels of knowledge, understanding, perception, applications and information (Rudner, 2009).

MCQ are used in the final and mid-term examination in various fields of medical science (Gillam, Rodrigues, & Myles, 2015). To MCQ, Millman principles must be adhered to. The most important Millman principles in designing multiple choice questions include: inserting most of the question information in the stem of the question, following each learning goal with one question, using simple and clear words, non-application of negative options in the stem of the questions, non-application of "all choices" or "none", non-application of conflicting options, highlighting negative words, independency of the questions from each other, equivalency of the questions in terms of length and lexical structures, non-application of repetitive words, lack of spelling mistakes and inserting the choices vertically. After the test, the questions must be analyzed in terms of quality and quantity. Qualitative aspect refers to the percentage of questions with educational objective based taxonomy which include taxonomy I (applied knowledge), Taxonomy II (recall) and Taxonomy III (concepts application), and also refers to the percentage of questions with no structural problems. Quantitative aspect of the test implies on the whole test validity, the difficulty coefficient, the discrimination coefficient and standard deviation of the test.

Several studies have analyzed teacher made multiple-choice tests in medical science. In radiology residency exams, the results indicated a reliability coefficient and a difficulty coefficient lower than the standard (Paulino & Kurtz, 2008). The analysis of the Anesthesiology tests showed an inappropriate difficulty coefficient and validity of the tests (Burton, 2005). Comparison of the professors' and experts' opinions about multiple choice questions based on achievement to educational aims in medical Sciences showed some problems in designing such questions (Hingorjo & Jaleel, 2012).

The situation is not better in other fields of medical science. Assessment of the written MCQ reflects important note about designation of multiple choice questions; such as used moderate difficulty index and more discrimination index to achieve better results in medical science teaching (Parmelee, Michaelsen, Cook, & Hudes, 2012). In quantitative analysis of the tests, not only attention to structural problems but also view to taxonomic percent I, II and III tests were necessary (Haghshenas et al., 2012). These findings suggest the necessity of educating medical professors on multiple choice tests designation; analysis of the multiple choice questions can have a positive influence on the exams' reliability coefficient (Pourmirza Kalhori, Roshanpour, Rezaei, & Naderipour, 2011). This study was designed with the overall goal of qualitative and quantitative analysis and comparison of MCQ provided by the professors of Kermanshah University of Medical Sciences - Faculty of Allied Science in the academic year 2011-2012.

2. Materials and Methods

In this study, the reports of the analysis of the MCQ designed by the professors of the Faculty of Allied Science of Kermanshah University of Medical Sciences which were sent to the Evaluation Unit of Medical Education and Studies Development Center of Kermanshah University of Medical Sciences for quantitative and qualitative analysis in the academic year 2011-2012 were studied. Convenience sampling method using the available data was applied. Based on the available information, about 50 exams were studied; however, only 35 exams had qualitative and quantitative analysis capabilities. Thus, the qualitative and quantitative analyses were performed just for these 35 samples.

Data collection tools consisted of the analysis reports of the exams designed by the professors of Kermanshah University of Medical Sciences provided by the Evaluation Unit of Medical Education and Studies Development Center of Kermanshah University of Medical Sciences. The analysis software is known as the Unique Software in Exam Analysis. This software works in Windows XP and the results for quantitative variables of difficulty coefficient, discrimination coefficient, the whole exam validity and standard deviation of the questions were extracted from its reports. In order to collect the main qualitative data of the exams, the reports from Medical Education and Studies Development Center on qualitative analysis of each test including taxonomic percentage I, II and III of the questions and the percentage of questions with no structural problems, were used. SPSS software version 16.00 was used for data analysis. For summarization of the data, mean and standard deviation, coefficients and percentages were used. In order to analyze quantitative and qualitative characteristics of the questions, two-dimensional tables including the numbers and percentages used, and for comparison with the standards, t-test and chi square test were applied. In this study, to maintain ethical considerations, the names of the professors are not mentioned and they are not compared with each other. The permission to perform this study was received from the President of the Faculty of Allied Science. We estimated about 50 tests for the qualitative and quantitative analysis, while after sampling we had only 35 qualified exams for quantitative and qualitative analysis; this was the certain limitation of this study.

3. Results

Among the total of 35 analyzed exams, 27 tests (77.1%) were designed by male professors and 8 tests (22.9%) by female professors. The average validity coefficient of the whole test (KR-20) of 20(57.1%) tests was within the acceptable range. Average difficulty coefficient of 20 tests (57.1%) was in the acceptable range, two tests (5.7%) in difficult range and 13 tests (37.2%) in simple range (Figure 1).

Considering mean discrimination coefficient, 9 tests (25.7%) were in the acceptable range and 27 tests (74.3%) had a low discrimination coefficient. The mean SD scores of 6 tests (17.1%) were in the acceptable range. Considering a base of 80% as the limit for the questions with/without structural problems, 22(62.9%) test were below the baseline (inappropriate) and 13 tests (37.1%) were above the baseline (appropriate).
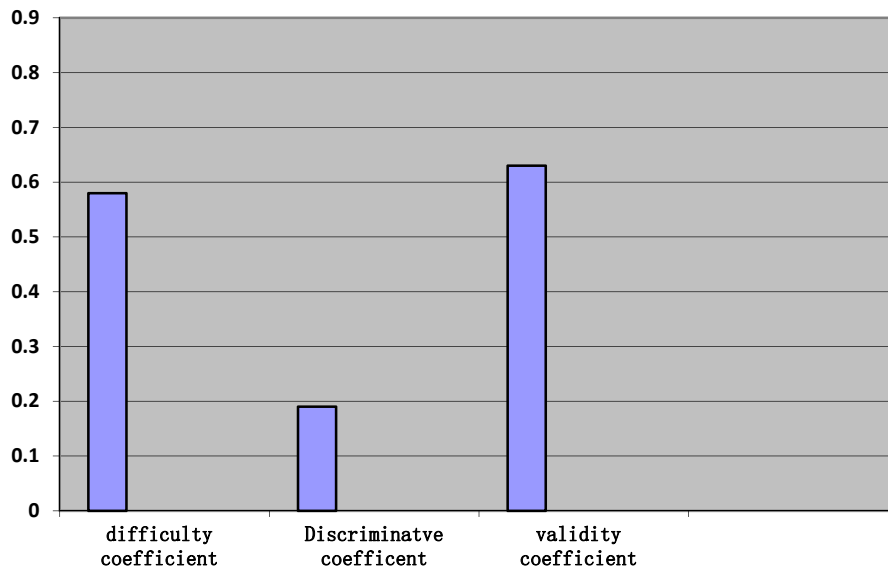

Figure 1. Quantitative Coefficients of the multiple choice questions (MCQ)

Taxonomy I ranged from 6% to 80% and the frequency of taxonomy I equal to 25% (acceptable limit) was 8.6%. Taxonomy II ranged from 5.3% to 69% and the frequency of taxonomy II equal to values of 45% and 52.5% was 5.7% (for both values). For taxonomy III in the range was measured as 0% to 47% and the highest frequency at the limit of 32.5% was 8.7% for taxonomy III. Amount of questions without structural problems ranged from 45% to 97.8% with the highest frequency of 17.1% at the limit of 85% and 11.4% at the limit of 75%.

With 95% confidence interval we claim that the releability of the whole exams ranged from 0.2 to 0.6, mean difficulty coefficient ranged from 0.57 to 0.8 and the discrimination coefficient ranged from -0.15 to 0.2. Average percentage of questions with taxonomy I was 36.7 (±18.69) with a significant difference among whole the exams after being examined by independent t-test (t = 3.73, P = 0.00). Average percentage of questions with taxonomy II was measured as 42.46 (±15.51) without any significant difference among the whole exams. The average percentage of questions with taxonomy III was recorded as 20.73 (±12.83) which showed a significant difference among the all exams after being tested by the independent t-test (t =- 4.27, P = 0.00), Figure 2. Average percentage of questions with no structural problems was 55.23 (±13.23) with a significant difference among the exams (t = 2.19, P = 0.041).

Chi square test showed that the test designed by male professors were more acceptable than those developed by females in the category of test validity (P = 0.02). The difference was apparent in the categories of difficulty coefficient, the mean standard deviation of the scores, and the discrimination coefficient but not statistically significant. The percentage of taxonomy I and II among the questions of the exams designed by male professors was more than those developed by females, but the difference was not statistically significant. However, male professors had assigned more questions to taxonomy III which was statistically significant (P = 0.059). Considering the mean percentage of questions without structural problems, no significant difference was observed between the tests designed by male and female professors.
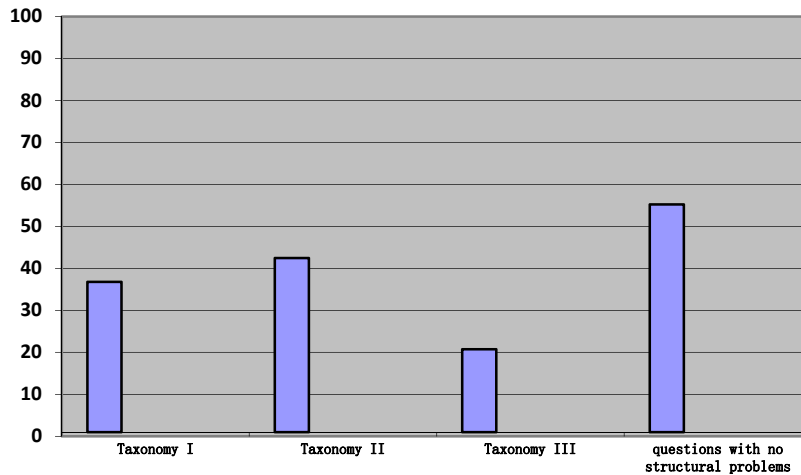
Figure 2. Qualitive percentage of the multiple choice questions (MCQ)

4. Discussion

Regarding the whole test validity and the test difficulty coefficient, the exam designed by the professors of the, was within the acceptable standard ranges. When compared with the results of studies conducted elsewhere, it is apparent that the multiple choice tests provided by the professors of the Kermanshah University of Medical Sciences - Faculty of Allied Science is superior to other university exams in reliability, difficulty and discrimination coefficients. A study in Mazandarans medical university suggests a minimum reliability coefficient and the difficulty coefficient being lower than the standards (Haghshenas, 2012). The results of another Medical School reports only 16.3% of the questions being appropriate in the difficulty and the discrimination coefficients (Kazemi, Ehsanpour, & Hassanzadeh, 2010). Even a study outside of Iran by Bachoff et al. (2000) in University of Mexico announced a difficulty coefficient of 0.34 and a discrimination coefficient of 0 (Backhoff, Larrazolo, & Rosas. (2000). However, in the Kuala Lumpur Medical School - Malaysia, the quantitative analysis of 12 multiple choice exams in the subjects of anatomy, physiology, biochemistry, genetics, statistics and behavioral sciences showed that the average difficulty coefficient of the questions was 64-89 percent (Mitra, Nagaraja, Ponnudurai, & Judson, 2009). The acceptable difficulty coefficient in the multiple choice questions of bronchoscopy training exam in Argentina is reported 0.65 ± 0.22 (Quadrelli, Davoudi, Galíndez, & Colt, 2009).

A discrimination coefficient lower than the standard range in the tests designed by the professors of the Faculty of Allied Science is also found in the study by Shaban (2006) (Shaban & Ramezani, 2007). However, in the exams developed in the Kuala Lumpur Medical School - Malaysia, 87% of the questions had an acceptable discrimination coefficient more than 0.2 and 55% had a very discrimination coefficient (greater than 0.39 (Al-Naggar, Musa, Al-Jashamy, & Isa, 2010). The discrimination coefficient in the MCQ of bronchoscopy training exam in Argentina was 0.52 ± 0.28 (Quadrelli, Davoudi, Galíndez, & Colt, 2009).

In the qualitative analysis of the questions regarding the indexes of taxonomy I, II and III, and also the percentage of questions without structural problems, the average percentage of questions with taxonomy I was 36.7%, taxonomy II, 42.46% and taxonomy III, 20.73%. When compared with the study by Shaban (2007) which showed 48.6% of questions to be in taxonomy I (Shaban & Ramezani, 2007). The better quality of questions regarding the taxonomic standards is indicated. Lack of adherence to the questions taxonomic standards, which indicate learning according to the educational goals classification (Bloom classification), has also been reported in other medical fields. A typical example is the low question taxonomy II and III percentages in the resident promotion exam which is reported in studies by some national medical universities (Kazemi, Ehsanpour, & Hassanzadeh, 2010). However, adherence to the taxonomic standards in tests designed by the professors of Kermanshah University of Medical Sciences in the resident promotion exam of the year 2013 has also been reported (Pourmirza Kalhori, Rezaie, Shojee Moghadam, Sepahi, & Memar Eftekhari, 2015; Cox, Irby, & Epstein, 2007). Average percentage of questions without structural problems in the exams designed by the professors within the Faculty of Allied Science was 75.23%. These findings compared with the qualitative analysis of the multiple choice tests in Mazandaran Medical School which showed only 46% of the questions to had no structural problems (Haghshenas et al., 2012) and in Mashhad Medical School which showed 34.9% of the questions to be free of structural problems (Sadatkhah, 2009) and could been better with education, indicates

a more adherence to the structural framework of multiple choice questions designation among the professors (Razazian & Pourmirza Kalhori, 2014) of Kermanshah University of Medical Sciences - Faculty of Allied Science. It seems structural problems in the residency promotion test exams little than in comparison with other exams. Most structural problems had reported were incongruity of distracters (34%), editing errors (26.5%) and double negative (14%), respectively (Razazian & Pourmirza Kalhori, 2014). In a study were reported by Pourmirza Kalhori et al. (2014) with aims to analyses MCQ exams of medical residents administered at Kermanshah University of Medical Sciences during 2008-2012 taxonomy II and III were decreased and taxonomy I was increased over the specified five years and recommend the specialist board members to consider improving MCQ quality of future exams. One of the suggestions of this study determining Correlation of quality and quantity index of multiple choice questions exams. Studies reported that reversed correlation was observed in difficulty level and discrimination index and design of simple and very hard tests induced decreasing discrimination index.

5. Conclusion

Based on the research findings, the test hypotheses is approved which indicates the adherence of the analysis of the multiple-choice tests designed by the professors of Kermanshah University of Medical Sciences Faculty of Allied Science in the year 2009 to the standards. The tests designed by the professors of Kermanshah University of Medical Sciences Faculty of Allied Science were in the acceptable standards of multiple choice exams regarding the whole exam validity coefficient and the difficulty coefficient. Despite the low Taxonomy I, II and III percentage of the questions and also the low rate of questions without structural problems, according to the basis of the standard defined in the study, by comparing with the studies mentioned in the discussion we can realized that one of the main problems by the multiple choice questions developers are these indicators. Strengthening the scientific knowledge of the professors, in order to upgrade their knowledge in multiple choice tests designation as a tool for evaluating students is beneficial.

Competing Interests Statement

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

Al-Naggar, R. A., Musa, R., Al-Jashamy, K., & Isa, Z. M. (2010). Knowledge, attitude and practice (KAP) towards sleep among medical students Of International Islamic University Malaysia (IIUM). *Chief Editor, 2*, 6.

Backhoff, E., Larrazolo, N., & Rosas, M. (2000). The level of difficulty and discrimination power of the Basic Knowledge and Skills Examination (EXHCOBA).Revista Electrónica de Investigación Educativa, 2(1). Retrieved from: http://redie.uabc.mx/vol2no1/contents-backhoff.html.

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment, 11*(1), 33-63. http://dx.doi.org/10.1207/s15326977ea1101_2

Burton, R. F. (2005). Multiple‐choice and true/false tests: myths and misapprehensions. *Assessment & Evaluation in Higher Education, 30*(1), 65-72. http://dx.doi.org/10.1080/0260293042003243904

Cox, M., Irby, D. M., & Epstein, R. M. (2007). Assessment in medical education. *New England Journal of Medicine, 356*(4), 387-396. http://dx.doi.org/10.1056/NEJMra054784

Gillam, S., Rodrigues, V., & Myles, P. (2015). Public health education in UK medical schools—towards consensus. *Journal of Public Health*, fdv069. http://dx.doi.org/10.1093/pubmed/fdv069

Haghshenas, M., Vahidshahi, K., Mahmudi, M., Shahbaznejad, L., Parvinnejad, N., & Emadi, A. (2012). Evaluation of multiple choice questions in the school of medicine, Mazandaran University of Medical Sciences, the first semester of 2007. *Journal of Strides Development Medical Education, 5*(2), 120-127.

Hingorjo, M. R., & Jaleel, F. (2012). Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *J Pak Med Assoc., 62*(2), 142-147.

Kazemi, A., Ehsanpour, S., & Hassanzadeh, A. (2010). Investigating the Academic Achievement Evaluation of

Specialized Theoretical Courses of Midwifery BS. *Iranian Journal of Medical Education, 9*(4), 346-355.

Khan, H. F., Danish, K. F., Awan, A. S., & Anwar, M. (2013). Identification of technical item flaws leads to improvement of the quality of single best Multiple Choice Questions. *Pakistan journal of medical sciences, 29*(3), 715. http://dx.doi.org/10.12669/pjms.293.2993

Khan, M., & Aljarallah, B. M. (2011). Evaluation of Modified Essay Questions (MEQ) and Multiple Choice Questions (MCQ) as a tool for Assessing the Cognitive Skills of Undergraduate Medical Students. *International journal of health sciences, 5*(1), 39-43.

Kibble, J. D., & Johnson, T. (2011). Are faculty predictions or item taxonomies useful for estimating the outcome of multiple-choice examinations? *Advances in physiology education, 35*(4), 396-401. http://dx.doi.org/10.1152/advan.00062.2011

Mitra, N., Nagaraja, H., Ponnudurai, G., & Judson, J. (2009). The levels of difficulty and discrimination indices in type a multiple choice questions of pre-clinical semester 1, multidisciplinary summative tests. *IeJSME, 3*(1), 2-7.

Newble, D. I., & Cannon, R. A. (2001). *A handbook for medical teachers*: Springer Science & Business Media.

Parmelee, D., Michaelsen, L. K., Cook, S., & Hudes, P. D. (2012). Team-based learning: A practical guide: AMEE Guide No. 65. *Medical teacher, 34*(5), e275-e287. http://dx.doi.org/10.3109/0142159x.2012.651179

Paulino, A. C., & Kurtz, E. (2008). American College of Radiology In-Training Examination for Residents in Radiation Oncology (2004–2007). *International Journal of Radiation Oncology* Biology* Physics, 70*(3), 666-670. http://dx.doi.org/10.1016/j.ijrobp.2007.09.049

Pourmirza Kalhori, R., Darabi, F., Roshanpour, F., & Rezaie, M. (2013). Analytic assessment of multiple choice questions (MCQ) of faculties' members' scholarship in Kermanshah University of Medical Sciences, 2010-2011. *J Clin Res Paramed Sci, 2*(1), 43-51.

Pourmirza Kalhori, R., Rezaie, M., Shojee Moghadam, A. R., Sepahi, V., & Memar Eftekhari, L. (2015). Correlation of quality and quantity index of multiple choice questions exams of residency promotion in Kermanshah University of medical sciences, 2013. *J Clin Res Paramed Sci, 4*(1), 71-78.

Pourmirza Kalhori, R., Roshanpour, F., Rezaei, M., & Naderipour, A. (2011). Academic member's knowledge improvement effect on the results of multiple choice questions in residency exams analysis (2009). *Journal of Kermanshah University of Medical Sciences (J Kermanshah Univ Med Sci), 15*(2). 112-118.

Quadrelli, S., Davoudi, M., Galíndez, F., & Colt, H. G. (2009). Reliability of a 25-item low-stakes multiple-choice assessment of bronchoscopic knowledge. *CHEST Journal, 135*(2), 315-321. http://dx.doi.org/10.1378/chest.08-0867

Razazian, N., & Pourmirza Kalhori, R. (2014). The Analysis of the Structural Problems of the Residency Promotion Test Items at Neurology Department, Kermanshah University of Medical Sciences in 2013. *Educational Research in Medical Sciences Journal, 3*(1), 44-45.

Rudner, L. M. (2009). Scoring and classifying examinees using measurement decision theory. *Practical Assessment, Research & Evaluation, 14*(8), 2.

Sadatkhah, T. M., Bahrini Tossi, M. H., Drakhshan, A., Khajeh Dalloee, A., & Golami, H. (2009). Analytic assessment of multiple-choice tests. *Journal of Medical Education, 2*(2). 87-91.

Shaban, M., & Ramezani, F. (2007). Effect of test item analysis of summative exams on quality of test designing. *Hayat, 13*(1), 5-15.

Shah, J. J., Woodward, J., & Smith, S. M. (2013). Applied Tests of Design Skills—Part II: Visual Thinking. *Journal of Mechanical Design, 135*(7), 071004. http://dx.doi.org/10.1115/1.4024228

Copyrights