A Comparative Study between the Conventional MCQ Scores and MCQ with the CBA Scores at the Standardized Clinical Knowledge Exam for Clinical Medical Students

Mahmood Ghadermarzi, PhD[1]; Shahram Yazdani, MD[2]; Arash Pooladi, MD[3*]; Mojdeh Bahram-Rezaei, MD[2]; Fakhrosadat Hosseini, MD[2]

[1] PhD of Biochemistry, Assistant Professor of Kurdistan University of medical sciences.
[2] Education Development Center (EDC), Shahid Beheshti University of Medical Sciences, Tehran, Iran
[3] Education Development Center (EDC), Shahid Beheshti University of Medical Sciences, Tehran, Iran

Abstract

Background and purpose: *Partial knowledge is one of the main factors to be considered when dealing with the improvement of the administration of Multiple Choice Questions (MCQ) in testing. Various strategies have been proposed for this factor in the traditional testing environment. Therefore, this study proposed a Confidence Based Assessment (CBA) as a pertinent solution and aims at comparing the effect of the CBA Scoring system with that of the conventional scoring systems (with and without negative score estimation as penalty) on the students' scores and estimating their partial knowledge on clinical studies.*

Methods: *This comparative study was conducted using a standardized clinical knowledge exam for 117 clinical students. After two-step training, both the conventional MCQ and CBA examination was given in a single session simultaneously. The exam included 100 questions and the volunteers were requested to complete a questionnaire regarding their attitude and satisfaction on their first experience of the CBA after exam. A new confidence based marking system was selected for the scoring, which was a hybrid of the UCL and MUK2010 systems. The MCQ-Assistant, SPSS and Microsoft office Excel software were used for scoring and data analysis.*

Results: *The mean age of the volunteers was 27.3±5.47, of whom 43.6% were men and 69.2% were senior medical students. Exam reliability was 0.977. The fit line of the MCQ scores without penalty estimation was R^2=0.9816 and Intercept=18.125 or approximately.2 deviation in the low scores. The MCQ scoring with penalty had a fit line approximately parallel to the 45-degree line but on or above it and the CBA scoring fit line was nearer to the 45-degree line, parallel to it and a little below it. These two sets of scores had a significant p value0.037. The response percentage to the CBA is higher (p value=0.0001). The discrimination power of the MCQ and the CBA for the upper and lower 1/3 of the students was not significantly different (p value=0.34). The students' satisfaction score was high and acceptable to the CBA system and expressed a positive perspective on this system for their examinations.*

Conclusions: *The CBA method can increase the competencies of the MCQ exams. It was found to have a greater fairness assessment, was an effective examination, an authentic testing method, with precise estimation and higher constructs validity than the conventional MCQ exam. The CBA simulate the reflection for deeper learning among the students.*

Keywords: *STUDENT ASSESSMENT, PARTIAL KNOWLEDGE, MCQ, CONFIDENCE-BASED ASSESSMENT, EXAM SCORING SYSTEM*

Journal of Medical Education Winter 2015; 14(1):6-12

Introduction

Multiple Choice Questions (MCQ) have been the most widely used assessment method and measurement tool for assessing the students' knowledge for many decades, because of their sufficiency in assessing a wide range of objectives, easy usage, feasibility for large scale exams with several examinees, easy scoring, ability in assessing the higher steps of the cognitive domain due to Bloom's taxonomy and high reliability. Compared with other testing formats the MCQ test is more easily affected by the guessing system which is the most important problem confronted by MCQ testing. This problem arises because of its all-or-none or dichotomous scoring system. In this scoring system an examinee will check the correct answer if she/he has complete knowledge regarding the Item; otherwise she/he will either omit the question or hazard a random guess from among the alternatives offered (1, 2). In fact, the examinee who does not know the correct answer will always make his/her selection based on partial knowledge (1, 3). However, while we estimate the guessing range by a negative score (as penalty) in the MCQ, it is impossible to distinguish correct answers based on complete knowledge from those arrived at from merely lucky guesses.

Various solutions have been proposed to overcome the issues described above. Marking Scheme for MCQ like order-of-preference scheme, answer-until-correct procedure and Confidence Based Assessment (CBA) or confidence-weighting assessment marking are all attempts to diminish the influence of blind guessing and to assess the partial knowledge of the examinees (4). The solutions mentioned above are all designed for the conventional Paper-and-Pencil (PandP) or Computer-Based Testing (CBT). In 1990, the CBA method known as Information Reference Testing (IRT) was designed by Dr. James Bruno and later renamed CBA (Confidence-Based Assessment). Recently, Dr. Gardner-Medwin from UCL University suggested changing the name to Certainty-Based instead of Confidence-Based Assessment or marking.

For each of the marking schemes or scoring systems (conventional MCQ, CBA, or …), the specific formula for scoring has been accepted as the most widely used method to reduce the effect of guessing and discourages the examinees from blind guessing and has been employed in this study. These formulae were derived from voluminous theoretical and empirical literature that had been developed by the issues such as the logic of mathematics, impact of test directions on testing scores (1, 2-5), test reliability and validity under formula scoring (5-9), and the tendencies to guess the different personality traits and different cultural backgrounds (2, 10). A majority of these problems lies in the dichotomous nature of the conventional MCQ test whereas in the CBA these problems are the least felt. Additionally, in the CBA test the reliability, validity, efficiency and the authenticity are better when compared with the MCQ. The construct validity of the test increases effectively in the CBA, if the scoring encourages the examinees to select all of the items in the test (by the omitting the negative score for the lowest level of confidence chosen by the examinee, for example). The CBA also encourages reflection, self-awareness and expression of the appropriate levels of confidence (11).

Due to the advantages of the CBA method and the useful information derived from it, we used this method for the assessment of the clinical knowledge of the clinical medical students and compared their final scores with the scores drawn from the conventional MCQ (with and without penalty). This comparison can show some of the incompetency of the MCQ in measuring the real knowledge level (and partial knowledge) of the examinees. This study emphasizes the importance and usefulness of the CBA in student assessment.

∗**Corresponding author:** Arash Pooladi, MD. Educational Development Center (EDC), Kurdistan University of Medical Sciences, Kurdistan, Iran.
Tel: (+98) 918 3736577
Email: a_pooladi@yahoo.com

Methods

This experiment is a comparative study conducted by an unpublished standardized clinical exam on 117 senior and junior clinical medical students in Iran. The exam includes 100 questions (items) selected from the Step-2 CK (Clinical Knowledge) of USMLE, by the Medical Board of the Shahid Beheshti and Kurdistan Universities of Medical Sciences. The standard time allotted for the exam was estimated at 150 minutes (the time for the confidence selection was estimated within it, as well). The number of choices for each of the questions in the exam is variable -83% of the items (questions) have five choices, 16% have six choices and only 1% has four choices. As this was their first experience in our CBA exam, the students were trained in two steps; first by the sheets as an exam guide for familiarity with the CBA concept, CBA scoring system, and how to fill in the blanks; next an explanation was given regarding the CBA and other necessary information through a 20-minute presentation. All the students found it easy to understand. The examinees were asked to select the questions that they would not have answered had the exam been a conventional MCQ with the usual penalty scores, using a small identification mark on the item in their questionnaires.

In the conventional MCQ exam, the scoring was estimated both with penalty and without it. The formula employed for the penalty estimation for the wrong answers is: (-1/n-1) where n=Number of choices for each question. (Four, five and six choices were given for the questions in this exam).

In the test, the examinees were requested to state, with each answer, their level of confidence (1, 2 or 3) in the correctness of their decision. Corresponding marks of 0.33, 0.66 and 1 were awarded if their selection was correct, while 0, -0.33, and -1.33 were the penalties levied (respectively) otherwise, (as the negative coefficient of the wrong confidence into the penalty formula given above: 1/n-1). This marking scheme is derived from the UCL (Medwin-Gardner) and combined with the MUK2010 (Kurdistan University of Medical Sciences) scheme which was calculated using the MCQ-Assistant Software. This software as well as the unique and innovative format was designed by Dr. Mahmood Ghadermarzi (from the MUK-Kurdistan University of Medical Sciences), and the OMR of the students' exam papers for mark reading with human control (for student response and their confidence level in each question), Scoring, item analysis and the analysis of the whole exam was performed by this software.

In Confidence Based Assessment (CBA) the student is asked to (1) select an answer and (2) the level of his/her confidence that he/she has answered correctly. The score of a CBA-type question is determined by the combination of two variables; the validity of the answer (correct / wrong) and the level of confidence. In table 1 the confidence based marking scheme is explained.

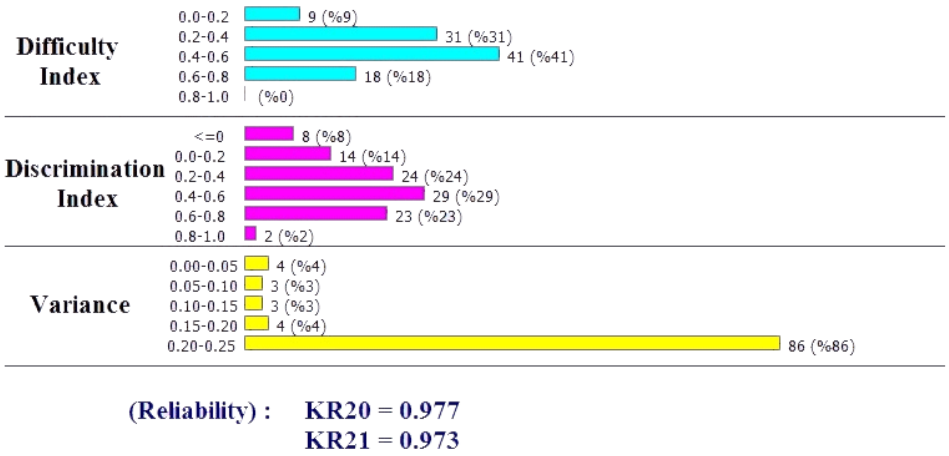
The data was analyzed by SPSS ver.14 software and Microsoft Office Excel 2007 was utilized for some estimations and Graph drawings.

Results

In this study we examined 117 volunteers of whom 43.6% were men, 69.2% (81 persons) were senior clinical medical students and the others were juniors. The mean age was 27.3±5.47 years.

Table 1. The confidence-based mark scheme

Confidence Level	C=1 (Low)	C=2 (Mid)	C=3 (High)	No Reply
Mark if Correct	o.33	0.66	1	0
Penalty if wrong (as the coefficient of negative score)	0	-0.33	-1.33	0



Graph 1. Clinical Knowledge Exam Indices in the overall item analysis (this image is one of the outputs of the MCQ-Assistant Software).

The indices of the exam analysis were estimated and the reliability of the exam was 0.977, as seen in graph 1, Difficulty Index, Discrimination and variance of items, arranged and dispread in an acceptable pattern.

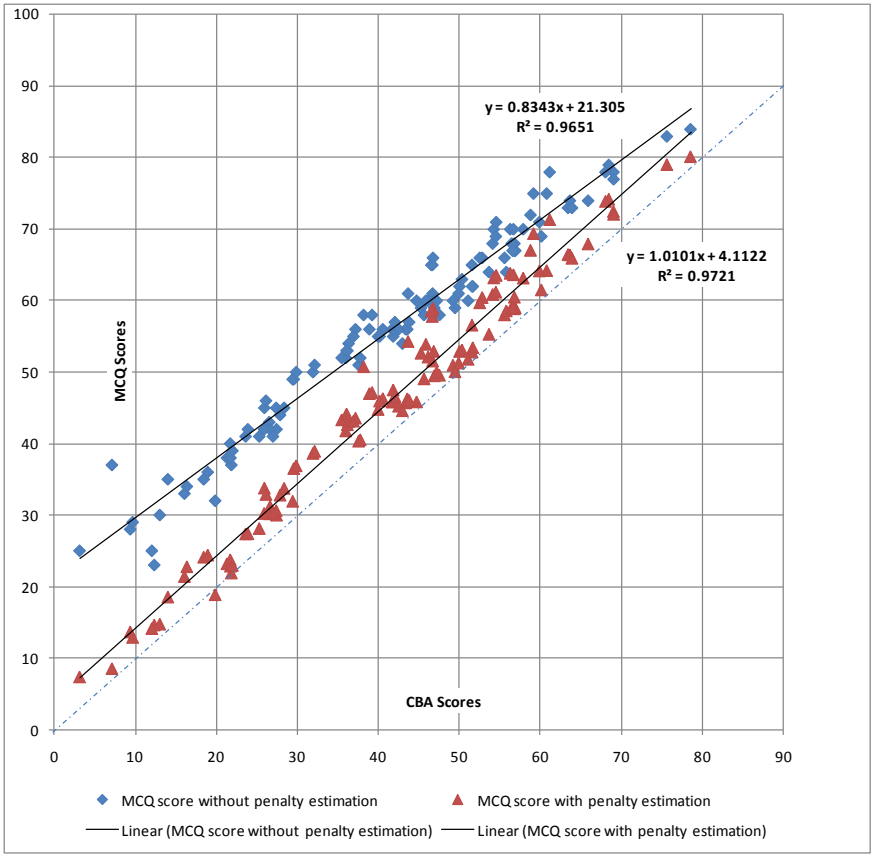
A comparison was done between the MCQ exam scoring results and that of the CBA, (as seen in the scatter plot of Graph 2). First, we found that the MCQ with penalty had a better determination of the students' scores than the MCQ without the penalty as the slope of the MCQ without the negative score (penalty) had a score of 20 or approximately 20% failure in the low scores (compared with the 45-degree line, MCQ without the penalty fit line with R-square (R^2): 0.9816 and Intercept: 18.125, has a 0.2 deviation in the low scores). The statistical difference between the two sets of scores is indicated by p-value: 0.0001 (df: 224.28, t: 4.913 and MD: 9.869). This graph is drawn based on the CBA scoring in the X axis.

Graph 3 when compared with the previous one (Graph 2), better matched the full estimation of the students' knowledge. Statistical analysis showed the p value: 0.037 (df: 233, t: 0.098 and MD: -4.53) for the MCQ (without penalty) and the CBA scores fit lines in Graph 3 (based on the MCQ with penalty scores in the X axis).

The average response percentage for the exam, in the MCQ type exam was 94.068±7.41 (therefore, the mean of the no-response questions was 8.034±4.98) and this percentage for the CBA system was 99.196±4.98. The statistical difference between these two means was significant (p-value: 0.0001, df: 203.036, t: -6.207 and MD: -5.128).

The discrimination power for the CBA scoring system and the conventional MCQ (with penalty), were determined by a comparison between the mean differences of the upper one-third (1/3) of the students based on their final exam scores, with the lower one-third ones. In the MCQ the mean difference was 36.43±1.69 (39.8 to 33.06) and in the CBA that is 37.27±1.77 (40.8 to 33.73), these ranges are different and wider in the CBA but have no statistical significance (p value= 0.34, df: 76, t:-0.959).

Student satisfaction earned by 7 open questions showed that almost all of them were unfamiliar with the CBA concept before this exam. The exam system (in the CBA scoring, coding system for question responses and determination of confidence level) was very simple and easy to learn for 93.2% of them (109 persons). They expressed that the other students had a positive perspective regarding the CBA clinical knowledge exam format as their first experiment. In fact,



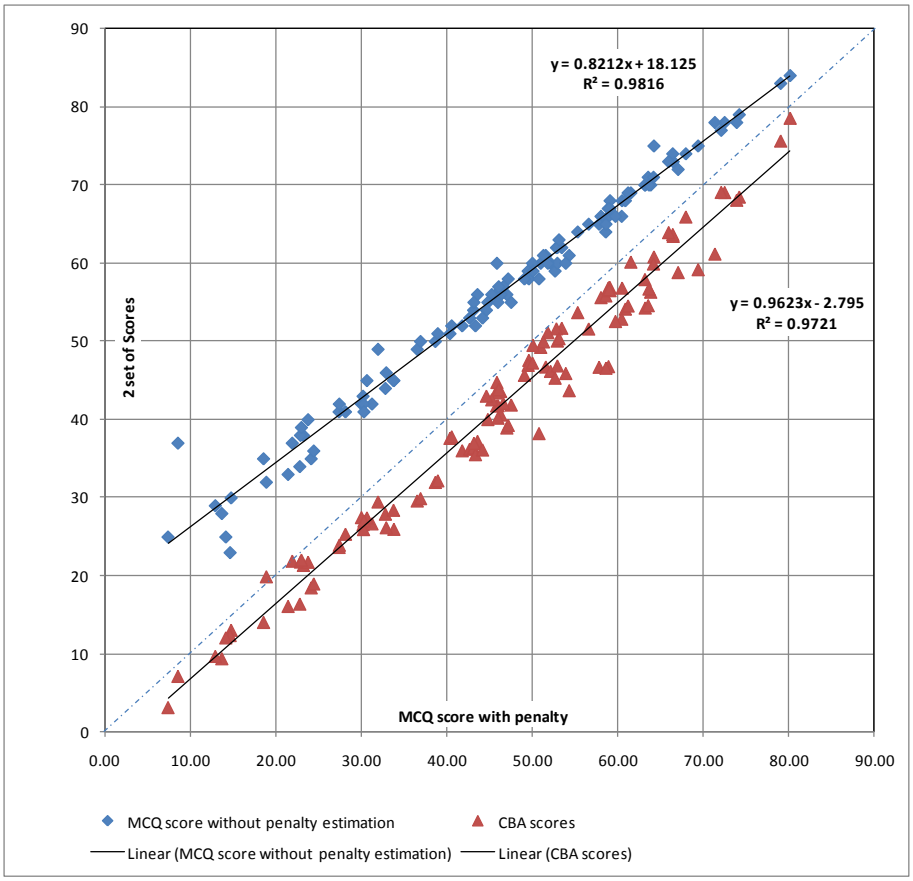
Graph 2. Comparison between the distribution of scores estimated by the MCQ without penalty and the MCQ with penalty, based on the CBA scores on the X axis.

90.6% of the students stated that this system for examinations had a positive effect on fairness of the test on students' assessment, 91.5% thought that this system could help them in deeper and effective learning. Finally, 82.9% of these volunteers wished that their prior exams had been conducted using the CBA system and 81.2% preferred that their future exams be set based on this system.

Discussion

This study is based on a unique and innovative confidence-based marking system for the MCQ exams and estimation of the students' partial knowledge. In fact, this system is a hybrid of two scoring methods, the UCL method and the MUK 2010 method (4, 11-13). By the UCL marking system, the penalties of each confidence level and by the MUK2010 system, the coefficient of the penalties are defined (as the number of choices in each questions estimates in the formula).

A properly designed scheme for the CBA ensures that in order to get the best marks the students must discriminate between the responses based on sound knowledge or understanding and those in which there is a significant risk of error. This motivation scheme implies that the confidence answers gain more marks if correct, but at the risk of significant penalty (negative scores) if wrong; low confidence benefits the student when there are responses for reservation, because the penalties are proportionally less or absent (11-14). The scores or mark scheme for the CBA is shown in Table 1. The CBA marking system is shown as justifiable and rational and as seen in Graphs 2 and 3, it can estimate the most proper and precise scores



Graph 3. Comparison between the distribution of scores estimated by the MCQ without penalty and the CBA scores based on the MCQ with penalty on the X axis.

in the exam for assessment of the students' knowledge.

The MCQ-Assistant software for the estimation of scores and analysis of the exam and usage of the standardized questions for this study are the strong points of the project.

About the exam quality, the overall item indices show that the exam has good characteristics, for example (as seen in Graph 1), skewed or shift of the items mean discrimination indices and variances to the upper range and bell shaped (normal) curve of the difficulty index, can confirm the test quality and validity of the results as we can read them in some references (16, 17).

The necessity for the penalties for the MCQ tests as shown in Graph 2 and the approximately 20% error for weak students (in the low scores) are results of the chance phenomenon (20% error for the test with 5 and 6 choices for each question, is rational).

Increase in the response percentage for one test, can increase the construct validity because the low omission rate of the questions in the exam is a construct for knowledge assessment. Increase in the discrimination and mean difference in the scores of the upper from the lower students could help the fairness of the assessment process.

Based on the results of satisfaction and the students' perspectives, the CBA method can stimulate the reflection learning, their self-awareness along with the training process, deeper learning and fairness of the assessment method. Additionally, by implementing the CBA the authenticity of the exam can be increased. These results are supported by some previous researchers (18-20).

Conclusion

The CBA method can increase the competencies of the MCQ exams to ensure greater fairness of assessment, effective examination, authentic testing and precise estimation and higher construct validity than the conventional MCQ exam. The CBA stimulate the reflection for deeper learning among the students.

Acknowledgments

Special thanks to the EDC members of Shahid Beheshti University of Medical Sciences and the EDC members of Kurdistan University of Medical Sciences.

References

1. Lord FM. Formula scoring and number-right scoring. Journal of Educational Measurement. 1975;12:7-22.
2. Rowley, G. L. and Traub, R. E. Formula scoring, number-right scoring, and test-taking strategy. Journal of Educational Measurement. 1977;14:15-22.
3. Jackson, R. A. Guessing and test performance. Educational and Psychological Measurement. 1955;15:74-9.
4. Yen YC, Ho RG, Chen LJ. A Polytomous Computerized-Adaptive Testing that Rewards Partial Knowledge. Frontiers in Artificial Intelligence and Applications; Vol. 151. Proceeding of the 2006 conference on Learning by Effective Utilization of Technologies: Facilitating Intercultural Understanding. 2006, P:629-36.
5. Diamond J, Evans W. The correction for guessing. Reviews of Education Research. 1973;43:181-91.
6. Bliss LB. A test of Lord's assumption regarding examinee guessing behavior on multiple-choice tests using elementary school students. Journal of Educational Measurement. 1980;17:147-52.
7. Alnabhan, M. An empirical investigation of effects of three methods of handling guessing and risk taking on the psychometric indices of a test. Social Behavior and Personality. 2002;30,645-52.
8. Burton RF. Misinformation, partial knowledge and guessing in true/false tests. Medical Education. 2002;36:805-11.
9. Burton, R. F. Multiple-choice and true/false: myths and misapprehensions. Assessment and Evaluation in Higher Education. 2005;30:65-72.
10.Gafni N, Melamed E. Differential tendencies to guess as a function of gender and lingual-cultural reference group. Studies in Educational Evaluation. 1994;20:309-19.
11.Gardner-Medwin AR, Gahan M. Formative and summative confidence-based assessment. In Proceedings of the 7th International Computer-Aided Assessment Conference, Loughborough University, UK. 2003;147-55.
12.Bryan C Clegg K. Innovative assessment in higher education. (Part 3, simulating learning-confidence-based marking: toward deeper learning and better exam, By Gardner Medwin A R). Routledge London. 2006;p:141-9.
13.Gardner-Medwin AR. Rational and irrational marking schemes. Journal of Physiology. 1999;515P: 48P
14.Issroff K, Gardner-Medwin AR. Evaluation of confidence assessment within optional coursework. In : Oliver, M. (Ed) Innovation in the Evaluation of Learning Technology, University of North London: London. 1998;pp:169-79.
15.Hammond EJ, McIndoe AK, Sansome AJ, Spargo PM. Multiple-choice examinations: adopting an evidence-based approach to exam technique. Anaesthesia. 1998;53:1105-8.
16.Coombs CH, Womer FB. The assessment of partial knowledge. Educational and Psychological Measurement. 1956;16:13-37.
17.Budescu D, Bar-Hillel M. To guess or not to guess: a decision-theoretic view of formula scoring. Journal of Educational Measurement. 1993;30:277-91.
18.Abu-Sayf FK., Diamond JJ. Effect of confidence level in multiple-choice test answers on reliability and validity of scores. Journal of Educational Research. 1976;70:62-3.
19.Hassmen P, Hunt DP. Human self-assessment in multiple-choice testing. Journal of Educational Measurement. 1994;31:149-60.
20.Cagnone S, Ricci R. Student ability assessment on two IRT models. Metodoloski zvezki. 2005;2:209-18.