Computer-based MCQ Assessment for Students in Dental Medicine – Advantages and Drawbacks

Cristina Gena Dascalu, Angela Mariana Enache, Radu Bogdan Mavru *, Georgeta Zegana

*University of Medicine and Pharmacy "Grigore T. Popa", Iasi, Romania
†Municipal Emergency Hospital Pascani, Romania

Abstract

The aims of the study was to compare the marks obtained at the final assessment, using MCQ tests and at the oral evaluation, for a preclinical and a clinical discipline, on a sample of 285 undergraduate students at dental medicine divided into two groups according to their year of study, in order to evaluate the advantages and the drawbacks of these performances. The statistical results showed that the grades’ distribution at MCQ tests follow the normal law, but their averages were significantly lower than the average marks obtained at the oral evaluation, regardless the year of study or gender.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the Scientific Committee of PSIWORLD 2014.

Keywords: MCQ tests, assessment methods, undergraduate students’ performance, medical university education.

1. Introduction

„Assessment of learning” represents the measure of knowledge acquirement in a certain scientific domain, while „assessment for learning” has the purpose to check the assimilated notions during the learning process (Furnham, Batey, and Martin, 2011). „Constructive alignment” is a modern concept in pedagogy, which assumes a unitary approach of knowledge objectives, content and final evaluation forms, so as the teacher’s image about the lecture to overlap the student’s image. The major objective of modern superior education curricula doesn’t only concern acquiring theoretical knowledge, concerning equally the developing of critical thinking, the students’ analysis, reasoning and decision skills and the formation of correct learning skills (Vaughan-Wrobel, O’Sullivan and Smith, 1997).

* Corresponding author. Tel.: +40-742-560864.
E-mail address: radumavru@gmail.com
In pedagogical theory there are three main categories of assessment: self-assessment, group assessment and peer assessment. The first two are useful during the learning process, while the last is necessary during the final evaluation stage. The non-standard forms of assessment, especially those computer-based, are very popular in the USA, in ascension in UK and more and more popular among students (Ventouras, Triantis, Tsiakas, and Stergiopoulos, 2010). The MCQ tests (Multiple Choice Questions) can be used in any type of assessment, in order to evaluate the student’s capacity to remember exact data, to interpret data or to analyze a proposed material. The MCQ tests must cover the six purposes of education process, specific for cognitive sciences and described in Bloom’s taxonomy (Higgins & Tatham, 2003). In this end, our research’s purpose was to evaluate the advantages and the drawbacks of MCQ tests (preferred by the students) as assessment forms at the end of specific modules from the public university educational system of dental medicine. Our study had the aim to compare the marks obtained at final assessment made using MCQ tests and at the oral evaluation, for a preclinical and a clinical discipline, on a sample of undergraduate students divided into two groups according to the year of study and separated by genders.

2. Method

2.1. Participants

The study was conducted at University of Medicine and Pharmacy "Grigore T. Popa", Iasi, Romania, during January - July 2013. The sample consisted from 285 undergraduate students (111 boys and 174 girls) with ages between 19-26 years old, examined at the end of some modules from the Dental Medicine Faculty curriculum. The participation was voluntarily, the students were informed about the aims and the nature of the research and the study was conducted in accordance with the Helsinki Declaration of 1975, as revised in 2000. The students were divided in two groups, according to the year of study: group A=153 students (52 boys and 101 girls) from the 1st year, examined at the end of a preclinical discipline; group B=132 students (59 boys and 73 girls) from the 5th year, examined at the end of a clinical discipline.

2.2. Instruments

Before the assessment, the students were asked about their preferences regarding the type of written evaluation. The students’ theoretical knowledge was evaluated through the final written MCQ test, and the practical knowledge through oral evaluation, in different days. The assessments’ results were quantified through marks from 1 to 10. The structure of the MCQ tests consisted of 45 items with five variants of answers and a single correct option, with a solving time of 60 minutes. 40% items had a difficulty level under the average, 40% items had an average difficulty and 20% item had a difficulty level above average. The MCQ tests were adequate with the theoretical content of the courses, being written according to certain rules: (1) clear and not ambiguous questions, with short enouncements, regarding a single subject; (2) the right answer randomly positioned, and not deductible from the other items; (3) short, homogenous and plausible distractors, without double negations or negative premises, containing the students most frequent errors and with no clues regarding to the right answer (Holsgrove, 1992; Higgins & Tatham, 2003).

2.3. Procedure

Each right answer was marked with 0.2 points, and into the base of calculation was included 1 point, in order to obtain the 10 grade for a perfect answer. The test checking was fully computerized, the papers being scanned and the results being automatically interpreted. The oral evaluation consisted from making a practical maneuver selected from the module’s curriculum and explaining the execution steps and methods; the allocated time was 40 minutes, in the presence of an assessment committee. Each student received two marks, one for the written test and the other for the oral evaluation. The final mark was calculated as the weighted mean between the following items: the mark given by group assistant (15%), the oral (35%) and the written assessment (50%) for the preclinical discipline; the mark given by group assistant (10%), the oral (40%) and the written assessment (50%) for the clinical discipline.

Statistical analysis was performed using the SPSS 16.0 package (SPSS Inc., Chicago, IL) for Windows. We calculated the frequency distributions and the parameters of descriptive statistics: mean, standard deviation (SD) and
standard error of mean (SEM). We compared the students’ marks globally and separately on genders, within each year of study, between the two types of evaluations. The data fitting with normal distribution was checked using the Kolmogorov-Smirnov One Sample test; in order to compare the marks at different types of assessments we used the Wilcoxon Signed Ranks test for paired samples, the Mann-Whitney test and the t-test for independent samples when the normal law was verified. For all the statistical tests we used the significance level p≤0.05 corresponding to a confidence interval of 95%.

3. Results

At the question “Which is the most relevant method to evaluate the theoretical knowledge?” over 50% of students answered that they preferred the MCQ tests (figure 1).

![Fig. 1. The students’ answers distribution regarding the written assessment methods: (a) 1st year of study; (b) 5th year of study](image)

The distribution of students’ marks at MCQ tests follows the Gauss curve, for the students in the 1st year, as well as for the students in the 5th year (Kolmogorov-Smirnov test, p=0.272), while the students’ results at the oral assessment do not follow the normal law in both years of study (Kolmogorov-Smirnov test, p=0.000) (figure 2).

![Fig. 2. The students’ marks distribution at the two types of assessments: (a) 1st year of study; (b) 5th year of study](image)

| Variables                           | MCQ tests | Oral assessment | Wilcoxon Signed Ranks test |
|-------------------------------------|-----------|----------------|---------------------------|
|                                     | Mean ± SD | SEM            | Mean ± SD                 | SEM          | Z value | p value |
| Preclinical discipline (n=153)      | 5.16 ± 1.058 | 0.086         | 8.53 ± 1.595              | 0.129        | 10.540  | 0.000*  |
| Boys (n=52)                         | 5.07 ± 1.211 | 0.168         | 8.53 ± 1.742              | 0.242        | 5.993   | 0.000*  |
| Girls (n=101)                       | 5.20 ± 0.974 | 0.097         | 8.52 ± 1.524              | 0.152        | 8.678   | 0.000*  |
| Clinical discipline (n=132)         | 6.31 ± 1.342 | 0.117         | 8.14 ± 1.327              | 0.115        | 9.113   | 0.000*  |
| Boys (n=59)                         | 6.21 ± 1.153 | 0.150         | 8.03 ± 1.348              | 0.175        | 5.880   | 0.000*  |
The average marks obtained at MCQ tests are lower at the preclinical discipline compared with the clinical discipline and at the oral evaluation, the girls obtaining slightly better results than the boys, in all cases. We found statistically significant differences between the two types of assessments for both preclinical and the clinical disciplines, reported on years of studies and genders (table 1).

Regardless the assessment type (MCQ tests or oral evaluation), the differences between the marks obtained by the students in the 1st and the 5th year of study are statistically significant, while, within each year of study, the differences between the marks obtained by girls and boys are not significant (table 2).

Table 2. The statistical comparison of students’ grades and of genders at the two types of assessments (*p<0.001)

| Compared groups                                  | MCQ tests | Oral assessment |
|--------------------------------------------------|-----------|-----------------|
| Groups 1st year (n=153) and 5th year (n=132) students | 7.981     | 7731.500        |
| Group 1st year students: boys (n=52) and girls (n=101) | 0.689     | 2497.500        |
| Group 5th year students: boys (n=59) and girls (n=73) | 0.764     | 1914.000        |

Our research compared the undergraduate students’ performance level in public university medical educational system, measured by marking the knowledge of two groups of students, globally and separated on genders, after two types of assessments, at the end of a preclinical and a clinical discipline, in order to evaluate the advantages and the drawbacks of MCQ tests (systematically preferred by the students).

Our study is particular through the specificity of the analyzed sample and the originality of the applied methods. Our results derive from certain aspects regarding the students’ preferences for specific written assessment methods, the distribution of obtained marks at MCQ test and their comparison with the marks obtained at the oral evaluation. Although most students preferred the MCQ test instead the essay for the written assessment, the marks they obtained were low, at the preclinical as well as the clinical discipline. To solve a single question, the student had an allocated time of 1.33 minutes, and the psychological and emotional factors were minimized, because there was no direct contact with the examiner. Our results confirm previous studies which showed that the students perceive the MCQ tests as a simple form of evaluation and practice a surface learning, obtaining consequently low results (Scouller, 1998; Furnham, et al., 2011). Other authors have studied other variables, outlining the student’s profile correlated with his preferences for certain assessment forms: the emotional students prefer the MCQ tests, while those less emotional choose other methods of assessment (Struyven, Dochy, and Janssens, 2005). The introverted prefer the MCQ tests while the extroverted prefer the oral assessments, as well as the conscientious (Furnham, Christopher, Garwood, and Martin, 2008). The Openness-to-Experience personalities agree the essays and the oral evaluation, rejecting instead the MCQ tests. There are no identified correlations between the intelligence level or the students’ university performances and the preferred assessment method (Chamorro-Premuzic, Furnham, Dissou, and Heaven, 2005).

The emotional stress level was increased during the oral evaluation, because of the direct questions and the examiner’s presence, but the students’ results were far better than those obtained at MCQ tests. Using the Kolmogorov-Smirnov normality test, we found that the marks’ distribution at MCQ tests followed the normal law, which signifies that the computerized evaluation was objective, while the marks distribution at oral assessment did not fulfill this property, which means a certain degree of subjectivism in examiner’s notation, revealed also in previous studies. Thus the following advantages of the MCQ assessment type were revealed, namely the increased objectivity, ease of filling and the speed of obtaining the results, which explains the students’ preference, fact that was also noted by Hammond, Mclndoe, Sansome, and Spargo (1998) and Brady (2005). Higgins and Tatham (2003) and Ventouras et al. (2010) point out the advantages of MCQ test and especially the disadvantages, citing the risk of dumbing down.

The comparative results of our study showed significant differences between the assessment methods on disciplines, years of study and genders. We found that the average results at MCQ test were significantly lower than the average results at the oral evaluation, regardless the year of study or gender, even if the girls obtained slightly
better results than the boys in all cases. Irrespective of the assessment method, the students in the 1st year obtained lower marks than the students in the 5th year. Our results contradict previous studies, which demonstrated statistically that the efficiency of MCQ tests is entirely equivalent with the efficiency of oral evaluations, which can be therefore successfully replaced (Ventouras, Triantis, Tsiakas and Stergiopoulos, 2011).

Our study results identified certain advantages of MCQ tests, also highlighted by the specialty literature: the possibility to check a large area of knowledge in a relatively short period of time (Brady, 2005); the testing and stimulation of students’ critical thinking (Leung, Mok, and Wong, 2008); the significantly shortened time for computerized tests evaluation, which simplifies the teachers’ work; the student’s performance is not influenced by the speed of his handwriting (Higgins & Tatham, 2003). MCQ tests still have some important disadvantages: they are quite restrictive and do not encourage the student’s skills to develop ideas organized into a coherent argumentation; the limited options and the predefined answers do not allow the evaluation of the student’s reasoning, also encouraging the surface learning; do not allow the teacher to evaluate the real level of the students’ knowledge; there is still the chance for the student to guess the right answers (Higgins & Tatham, 2003; Ventouras et al., 2011). That is why some authors recommend combining different methods of assessment for the students in medical university education, in order to train good practitioners (Hayes & McCrorie, 2010).

4. Conclusions

The present study revealed that all students were disfavored by the MCQ test compared with the oral assessment, obtaining weak results regardless the discipline, year of study or gender. Despite their advantages, the use of MCQ tests in the university medical education is not enough, if it is desired a formative evaluation for a process of analysis allowing the diagnosis, prognosis and the assessments’ adjusting, because the most important attributes of a successful career in this domain are the capacities of analysis, synthesis and quick applying of knowledge in solving new, complex and individualized practical problems. A further research should target what are the implications of using MCQ in the medical university education, given the fact that this method is applied empirically, without an analysis of structure, reliability, validity and efficiency of the tests applied.

References

Brady, A. (2005). Assessment of learning with multiple-choice questions. *Nurse Education in Practice*, 5, 238–242.
Chamorro-Premuzic, T., Furnham, A., Dissou, G., & Heaven, P. (2005). Personality and preference for academic assessment: A study with Australian University students. *Learning and Individual Differences*, 15, 247–256.
Furnham, A., Batey, M., & Martin, N. (2011). How would you like to be evaluated? The correlates of students’ preferences for assessment methods. *Personality and Individual Differences*, 50, 259-263.
Furnham, A., Christopher, A., Garwood, J., & Martin, N. (2008). Ability, demography, learning style and personality trait correlates of student preferences for assessment method. *Educational Psychology*, 28, 15-27.
Hammond, E. J., McIndoe, A. K., Sansome, A. J., & Spargo, P. M. (1998). Multiple-choice examination: adopting an evidence-based approach to exam technique. *Anaesthesia*, 53, 1105-1108.
Hayes, K., & McCrorie, P. (2010). The principles and best practice of question writing for postgraduate examinations. *Best Practice & Research Clinical Obstetrics and Gynaecology*, 24, 783-794.
Higgins, E., & Tatham, L. (2003). Exploring the potential of Multiple-Choice Questions in Assessment. *Learning and Teaching in Action*, 2(1), 1-12.
Holsgrove, G. (1992). Guide to post graduate exams: multiple-choice questions. *British Journal of Hospital Medicine*, 48(11), 757–761.
Leung, S. F., Mok, E., & Wong, D. (2008). The impact of assessment methods on the learning of nursing students. *Nurse Education Today*, 28, 711-719.
Scouller, K. (1998). The influence of assessment method on students’ learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35, 453-472.
Struyven, K., Dochy, F., & Janssens, S. (2005). Students’ perception about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education*, 30, 325-341.
Vaughan-Wrobel, B., O’Sullivan, P., & Smith, L. (1997). Evaluating critical thinking skills of baccalaureate nursing students. *Journal of Nursing Education*, 36(10), 485–488.
Ventouras, E., Triantis, D., Tsiakas, P., & Stergiopoulos, C. (2010). Comparison of examination methods based on multiple-choice questions and constructed-response questions using personal computers. *Computers & Education*, 54, 455-461.
Ventouras, E., Triantis, D., Tsiakas, P., & Stergiopoulos, C. (2011). Comparison of oral examination and electronic examination using paired multiple-choice questions. *Computers & Education, 56*, 616-624.