COMPARISON OF ORAL EXAMINATION AND EXAMINATION METHODS BASED ON MULTIPLE-CHOICE QUESTIONS USING PERSONAL COMPUTERS

Dimos Triantis, Charalampos Stergiopoulos and Panagiotis Tsiakas
E-learning Support Team, Technological Educational Institution (T.E.I.) of Athens, 28 Ag. Spyridonos st., Athens, Greece

Keywords: Computer-aided assessment, Evaluation methodologies, Automated grading, Post-secondary education, Evaluation of CAL systems.

Abstract: The aim of this work was to compare the use of multiple-choice questions (MCQs) as an examination method, to the one based on oral-response questions (ORQs). The MCQs have an advantage concerning objectivity in the grading process and speed in production of results. But they also introduce an error in the final formulation of the score. The error concerns the probability of answering a question by chance or based on an instinctive feeling. In the present study, both MCQ and ORQ tests were given to examinees, in the framework of a computer-based learning system. Avoiding the procedure of mixed scoring, e.g. both positive and negative markings, a set of pairs of MCQs was composed. The MCQs in each pair were similar, produced by the same topic. This similarity was not evident for an examinee without adequate knowledge on the particular topic. The examination based on these "paired" MCQs, by using a suitable scoring rule, when made to the same sample of students, on the same topics and with the same levels of difficulty, gave results that were statistically indistinguishable with the grades produced by an examination based on ORQs, while both the "paired" MCQ test results and the ORQ test results differed significantly from those obtained from a MCQ using positive-only scoring rule.

1 INTRODUCTION

Nowadays, information technology, computers and telecommunications networks are continuously advancing. Everyday life is changing by this progress which is also accompanied by the global explosion in knowledge. Education as an essential aspect of our life is also affected by these changes (Fox, 2002). This revolutionary progress can lead us to acknowledge that learning is substantially based on new technologies (Crossman, 1997; Daniel, 1996; Phillips, 1992).

The use of PCs has helped educators to invent new methods or to adjust older ones in the educational process. Various studies have reported that these methods based on new technologies exert positive influence on the quality of teaching and are quite effective. (Lehmann, Freedman, Massad & Dintzis, 1999; Goggin, Finkenberg & Morrow, 1997; Castellan, 1993). Therefore computer technology might constitute a useful tool for a successful teaching and learning environment (Johnston, 1997). Nevertheless, the role of new technologies is not to replace or even degrade the traditional forms of teaching, but to strengthen what already exists and finally improve the quality and the efficiency of the provided education (Dede, 2000). An examination method has to be reliable and valid. Extra care has to be taken in order to adjust all possible parameters that will lead to this result. This is why an intense discussion has taken place regarding this matter (Bennett, Rock & Wang, 1991; Bridgeman, 1991; Wainer, Wang & Thissen, 1994). MCQs have a significant advantage: Scoring is absolutely objective and may be automated by the use of specialized software. In order to ensure that the results would be realistic, reliable and comparative, it is essential to meet some basic requirements. A basic requirement on behalf of the teacher is that the MCQs must be correctly formulated. Well designed questions and choices require from a student to have specialized knowledge and decision making skills taking into account that a specified time might be pre-determined for answering the whole set of questions.

Triantis D., Stergiopoulos C. and Tsiakas P. (2009).
COMPARISON OF ORAL EXAMINATION AND EXAMINATION METHODS BASED ON MULTIPLE-CHOICE QUESTIONS USING PERSONAL COMPUTERS.
In *Proceedings of the First International Conference on Computer Supported Education*, pages 13-18
DOI: 10.5220/0001977300130018

MCQs provide also the possibility to the teacher to ascertain the degree of assimilation of knowledge on specific topics of the module. A disadvantage of MCQs is that the examinee is judged solely on the choice of the answer and not on the steps made for selecting the particular answer. Using conventional assessment methods of MCQs, it is not always possible to thoroughly investigate whether a topic, which a specific question addresses, has been fully understood. There is always a chance that a student might gain some points by sheer luck if a positive-scoring rule is used. For eliminating this disadvantage, various alternative scoring methods have been proposed and implemented in the T.E.I. of Athens. The main one is based on a set of mixed-scoring rules for marking, in which students gain points for correct answers and loose points for incorrect answers. In that case, students were less willing to answer questions compared to MC tests based on positive scoring rules (Bereby-Meyer, Meyer & Budescu, 2003).

On the other hand, using examinations such as the one of ORQs, the examiner has the possibility to check the way the student developed the subject under question. The disadvantage of this method is the fact that subjects that might be examined cannot always cover all the topics of the module. It can also include grading difficulties as, sometimes, are not fully objective.

During the last five years, at the Technological Educational Institution (T.E.I.) of Athens, a considerable effort has been made in order to acquire, manage and disseminate educational material in digital form. A Web-based Course Management System called "e-education" has been created, based on the e-class platform, developed by GUNET (Greek University Network, 2000). E-class was based on the Claroline system, which is an open source software package (Open Source eLearning and eWorking platform, 2001). The system is used for the dissemination of the digital educational material that was created and allows professors and lecturers to create and administer modular websites through a web browser. Having provided to students, through the above-mentioned system, a significant amount of multimedia training content, in conjunction with the teaching provided through lectures, during the last three years, various computer-based examination methods have been introduced, offering to students specially structured questionnaires, mainly of the MC type. The results of those examination methods have been extensively discussed in previous publications and relative conclusions have been extracted (Stergiopoulos, Tsiakas, Kaitsa & Triantis, 2006; Triantis, Stavrakas, Tsiakas, Stergiopoulos & Ninos, 2004; Stergiopoulos, Tsiakas, Kaitsa, Triantis, Fragoulis & Ninos, 2006).

The aim of the research presented in this work was the comparison of the ORQs and MCQs examination methods. Students were examined in a PC laboratory room by the aid of special software. Two MCQs assessment models were used on the same examination. The first one is based on a positive scoring rule (referred as PSR-MCQs), and the second one assesses sets of pairs of MCQs (referred as "paired" MCQs). MCQs in each pair concerned the same topic, but this similarity was not evident for a student who did not possess adequate knowledge on the topic addressed in the questions of the pair.

The ORQs examination method took place in a lecture room by three examiners which posed to each student a set o questions.

2 METHODS & PREPARATION

2.1 The Examined Course and the Sample of Students

The course that was selected for comparing the scores of the two examination methods was a general interest course entitled "Physics of Semiconductor Devices". According to the current curriculum of the Department of Electronics, constitutes one of the basic modules and is taught during the first semester of the course. Knowledge obtained is essential and fundamental for the subsequent study of analog and digital electronics. A class of 34 students participated in the examination. All of them had previously been given instructions for the examination and had available related study material.

2.2 Constructing the Multiple Choice Questions

"E-examination" is a stand alone application created by the T.E.I. of Athens. It is mainly a managing and editing tool which can help the teacher to build and deploy assessment tests in a suitable form so as to be displayed in a web browser. In this way, it is assured that each test is portable and cross-platform. The examinee has to answer a series of questions through a user-friendly interface.

From previous examinations a database of MCQs

had been created. This database contains a pool of 300 questions which covers all the topics of the module.
A set of $\{q_1, q_2, \ldots, q_n\}$ (n=40) MCQs was randomly selected from the database, having into consideration to cover each teaching unit proportionally. A weight was assigned to each question, depending on its level of difficulty.
Next, a set of 20 ORQs was created. ORQs were short subject development questions which had to be answered orally in front of three teachers. For every student five questions picked up randomly. Extra care was taken so that the MCQs were, overall, of equivalent level of difficulty with the corresponding ORQs. The final score (M1) is the average one of the score given from every teacher. M1 score was normalized to value m1, whose maximum was 100, i.e.:

$$m1 = \frac{M1}{\sum_{i=1}^{n} c_i} \cdot 100 \quad (1)$$

It must be noted that 50.0/100.0 was the minimum normalized score required for passing the examination. This enabled the comparison of the scores of the different examinations.

2.3 ORQ Examination Procedure and Scoring Methodology

Students were firstly examined orally in a lecture room. Every student went alone in the room. A set of five ORQs was posed to him/her by three teachers. It was expected from the student to answer the questions the best he/she could. Every teacher gave the student a mark. The final mark of the oral examination is the average score of the three grades. Immediately after this first examination the student was led to the PC lab. There, a personal computer was waiting for him/her in order to take the electronic test comprised of MCQs.

2.4 MCQ Examination Procedure and Scoring Methodology

During the second phase, MCQs were given to students. After the end of the pre-determined examination duration time, a report page was produced by the system for each student, on which were recorded the final score, as well as each question with the indication of the correct answer and whether it was correctly or wrongly answered. One copy was given to the student and one to the examiner, for processing the scores. The students were assessed by using two different methods as described in the next section.

2.4.1 MCQs Positive Grade Method

Based on the MCQs answers, the positive scoring rule consisted in giving positive grades only to correctly answered questions. No grade was given for unanswered or wrongly answered questions. The overall examination score, M2, was computed according to the following formula:

$$M2 = \sum_{i=1}^{n} (q_i \cdot c_i) \quad (2)$$

where: n=40, q_i=1 if answer q_i had been correctly answered, q_i=0 if answer q_i had been wrongly answered or omitted, and c_i is the weight factor of question q_i, which takes a value from 1 to 3. As can be seen from Equation 2, this method of scoring of MCQs does not impose a penalty to the student by imposition of negative marking for incorrect answers or unanswered questions.
M1 score was normalized to value m1, whose maximum was 100, i.e.:

$$m2 = \frac{M2}{\sum_{i=1}^{n} c_i} \cdot 100 \quad (3)$$

It must be noted that 50.0/100.0 was the minimum normalized score required for passing the examination.

2.4.2 Paired MCQs

The other examination method based on a refined version of MCQs, using pairs of such questions. This was done in order to investigate its objectivity, based on the scoring, in evaluating the knowledge acquisition of the students, in comparison to the two examination methods that were mentioned above.
The same set of the 40 questions were used. Actually, the system was able to assess the electronic test using both methods at the same time. So, at the end of the examination the produced report had the results of each method. in the previous examination method of MCQs were excluded. When these questions had been selected an additional factor had taken into consideration. This factor is that these questions could form two subsets of 20 questions. The first subset is $\{q_{a1}, q_{a2}, \ldots, q_{ak}\}$ (k=20). The second subset is $\{q_{b1}, q_{b2}, \ldots, q_{bk}\}$ (k=20).

Each question q_{bi} having a similarity to question q_{ai} (i=1,…,k), forms a pair of MCQs, according to the following rationale: a) both questions referred to the same topic and b) the knowledge of the correct answer for question q_{ai}, from a student, who had proceeded to a systematic study and is cognizant of the topic, implied the knowledge of the correct answer for q_{bi} and vice versa. Furthermore, each question in a pair had the same weight c_i in the final score. The presentation of the 2k=40 questions that the students had to answer in the PC screen was designed so that the questions were given with a random sequence, taking care that each question q_{bi} was presented after a lapse of at least 10 questions after the presentation of question q_{ai}. Questions were automatically given through the software system, with suitable programming.

Therefore, during the examination time where a group of 34 students was present, two categories of examination were given to them: a set of 2k=40 MCQs qa1, qa2, …..qak and qb1, qb2, …..qbk. For the paired MCQs category the score (M2) was computed as follows:

$$M3 = \sum_{i=1}^{20} \left(q_{ai} + q_{bi} \right) \cdot c_i \qquad (4)$$

Where:

$$q_{ai} + q_{bi} = \begin{cases} 5/2 & \text{if both } q_{ai} \text{ and } q_{bi} \text{ are correct} \\ 1/2 & \text{if either } q_{ai} \text{ or } q_{bi} \text{ is correct} \\ 0 & \text{if both } q_{ai} \text{ and } q_{bi} \text{ are wrong} \end{cases}$$

Therefore, to produce score M3, a bonus is given to the student if he/she answered correctly both questions of the MCQ pair (q_{ai} , q_{bi}) and a penalty if he/she answered correctly only one question of the pair. M3 was next normalized to value m3, with maximum value 100, according to the formula:

$$m3 = \frac{M3}{\sum_{i=1}^{20} 2.5 c_i} \cdot 100 \qquad (5)$$

3 RESULTS & DISCUSSION

Table 1 shows the overall results of the examination methods applied. A first remark that can be done is that the method of oral examination (ORQ) and the multiple choice method of paired questions (MCQ paired) produce very similar results. On the other hand, the method of multiple choice questions with positive grading shows a clear deviation from the other two methods.

Table 1: The overall results of the examination methods applied.

	ORQ	PSR MCQ (positive)	MCQ (paired)
Number of students	34	34	34
Succeeded (>5.0/10)	18	24	17
% Succeeded (>5.0/10)	53%	70.6%	50%
% Excellent score (>7.5/10)	17.64%	23.52%	17.64%
Average score of students who participated	50.5	58.6	49.8

From the results it is obvious that the evaluation of the students with standard MCQs gives greater success rates and scores than with ORQs. This bias is evident also by the regression line of ORQ to MCQ (Figure 1) and it might be probably related to the "sheer luck" factor, of correctly answering questions by chance, when no negative-marking penalty procedure is incorporated in the marking of the answers.

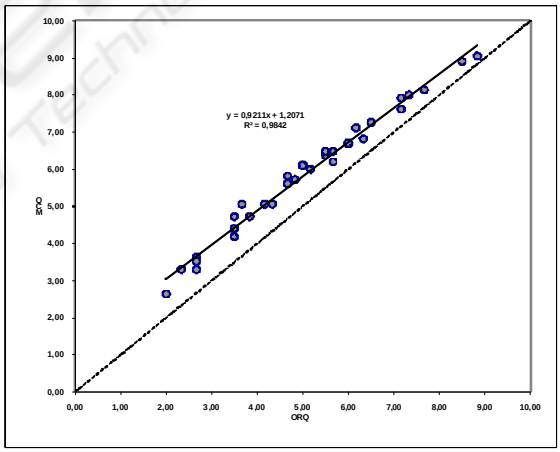

Figure 1: Regression line of normalized score ORQ to normalized score MCQ.

These discrepancies are reduced to insignificance or near insignificance when the set of students comprised only those students who got a normalized score greater than 50.0 or 70.0 respectively. Despite the fact that the maximum effort was taken so that examination categories with MCQs and ORQs are compatible concerning the content of the questions and their degree of difficulty, the two examination methods are not sufficiently equivalent.

The results also indicate that the paired MCQs

examination method with bonus/penalty adjustment (resulting in normalized score m3) is statistically equivalent to the ORQs examination method (resulting in normalized score m2), i.e. the traditional examination method used in most educational settings. Both methods differ significantly from the MCQs, which do not use a negative marking "penalty" procedure, i.e. the positive grading PSR-MCQs examination method (resulting in normalized score m1). The bias introduced by the "sheer luck" effect of PSR-MCQs, seem to be alleviated by the paired MCQs examination method with bonus/penalty adjustment, as indicated also by the regression line at Figure 2. This is achieved in the bonus/penalty paired MCQs examination method without explicit negative marking for incorrect answers, which might induce a "hampering" effect to the examinee, dissuading him/her from tackling a question for which he/she may possess an intermediate level of knowledge.

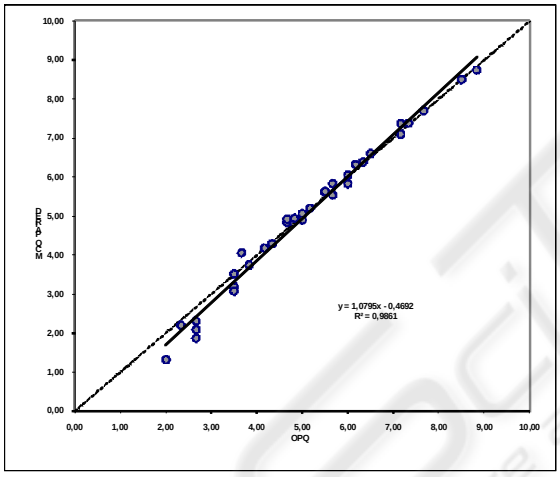

Figure 2: Regression line of normalized score OPQ to normalized score MCQ- paired.

4 CONCLUSIONS

The results of this study indicate that the examination method based on paired MCQs may constitute a reliable tool for the evaluation of students as long as the parameters regarding the validity of the examination are ensured. The positive grade bias introduced by PSR-MCQs can be alleviated without the use of negative marking for wrongly answered questions. A better knowledge check can be performed by the use of the combination of bonus and penalty in the pairs of MCQs. In this way the advantages of examinations based on MCQs can be fully exploited and the results can be comparable to those produced by a written or oral examination (Stergiopoulos, Tsiakas, Kaitsa & Triantis, 2006; Triantis, Stavrakas, Tsiakas, Stergiopoulos & Ninos, 2004; Stergiopoulos, Tsiakas, Kaitsa, Triantis, Fragoulis & Ninos, 2006). The two last mentioned examinations give the student the opportunity of free response which helps the teacher to ascertain level of knowledge assimilation (Bennett, Rock & Wang, 1991). The examiner has the possibility to check students' performance on almost the whole breadth of the topics covered by the material taught in the lectures offering also multiple advantages concerning the speed of results production and the transparency of the scores given. Furthermore, through suitable processing of the partial scores for each question, it is conjectured that a detailed investigation might be conducted concerning the weak points in the comprehension of the concepts that were presented in the teaching units of the course that was examined.

Therefore, useful conclusions could be drawn for the instructor, so that, among other possible remedying interventions, he/she could present in future lectures, in a more clear and thorough way, those topics where the MCQs examination indicated low success rates. This will provide the basis of future research of our group.

It is the opinion of the authors of the present study, that by making the manageable effort needed, paired MCQs might be suitably designed by examiners, so that MCQs examinations might properly quantify the knowledge and competences of the students and provide reliable assessment of their performance.

ACKNOWLEDGEMENTS

This work is co-funded 75% by E.U. and 25% by the Greek Government under the framework of the Education and Initial Vocational Training Program – "Reformation of Studies Programmes of Technological Educational Institution of Athens".

REFERENCES

Bennett, R. E., Rock, D. A,, & Wang, M. (1991). Equivalence of free-response and multiple-choice items. *Journal of Educational Measurement,* 28, 77-92.

Bereby-Meyer, Y., Meyer, J., Budescu, D. V. (2003).

Decision making under internal uncertainty: the case of multiple-choice tests with different scoring rules, *Acta Psychologica*, 112, 207–220.

Bridgeman, B. (1991). Essays and multiple-choice tests as predictors of college freshman GPA. *Research in Higher Educafion,* 32, 319-332.

Castellan, N. (1993). Evaluating information technology in teaching and learning behavior. Research Methods Instruments & Computers, 25, 233-237.

Crossman, D. (1997). The evolution of the World Wide Web as an emerging instructional technology tool. *In Badrul H. Khan (Ed.), Web-based instruction,* 19-23. N.J.: Educational Technology Publications.

Daniel, J. S. (1996). *Mega-universities and knowledge media: technology strategies for higher education*. London: Keegan Press.

Dede C., 2000, Emerging Technologies and Distributed Learning in Higher Education. *In: D. Hanna, ed., Higher Education in an Era of Digital Competition: Choices and Challenges, New York, 2000*, New York: Atwood, 71-92.

Fox R., 2002, Online technologies changing university practices, *In: A. Herrmann & M. M. Kulski, eds, Flexible Futures in Tertiary Teaching*, 2-4 February 2000, Curtin University of Technology, Perth, WA. Perth: Curtin University of Technology, 235-41.

Goggin, N.L., Finkenberg, M.E., & Morrow, J.R. (1997). Instructional technology in higher education teaching. *Quest*, *49(3)*, 280-290.

Greek University Network (2000). http://www.gunet.gr.

Johnston, I. (1997), The place of information technology in the teaching of physics majors, *AIP Conference Proceedings*, 399, 343-356.

Lehmann, H., Freedman, J., Massad, J., & Dintzis, R. (1999). An ethnographic, controlled study of the use of a computer-based histology atlas during a laboratory course. *Journal of American Medical Information Association*, *6(1)*, 38-52.

Open Source eLearning and eWorking platform (2001). http://www.claroline.net.

Phillips, R. L. (1992). Opportunities for multimedia in education. In S.Cunningham & R. J. Hubbold (Eds), *Interactive Learning through Visualization: The Impact of Computer Graphics in Education*, (pp. 25-35). Berlin: Springer- Verlag.

Stergiopoulos, C., Tsiakas, P., Kaitsa, M. & Triantis, D. (2006). Evaluating Electronic Examination Methods of Students of Electronics. Effectiveness and Comparison to the Paper-and-Pencil Method, *IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC 2006)* (Book 2, pp. 143-149).

Stergiopoulos, C., Tsiakas, P., Kaitsa, M., Triantis, D., Fragoulis, I. & Ninos, C. (2006). Methods of Electronic Examination Applied to Students of Electronics. Comparison of results with the conventional (paper-and-pencil) method, In *Proceedings of the 2nd International Conference on Web Information Systems and Technologies* (pp. 305-311).

Triantis, D., Stavrakas, I., Tsiakas, P., Stergiopoulos, C. & Ninos, D. (2004). A pilot Application of Electronic Examination Applied to Students of Electronic Engineering: Preliminary Results., *WSEAS transactions on advances in engineering education*, Vol. 1, 26-30.

Wainer, H., Wang, X.-B., & Thissen, D. (1994). How well can we equate test forms that are constructed by examinees? *Journal of Educational Measurement,* 31, 183-199.