Computers & Education 54 (2010) 455–461

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Comparison of examination methods based on multiple-choice questions and constructed-response questions using personal computers

Errikos Ventouras, Dimos Triantis, Panagiotis Tsiakas *, Charalampos Stergiopoulos

E-learning Support Team, Technological Educational Institution (T.E.I.) of Athens, 28 Ag. Spyridonos St., Athens, Greece

ARTICLE INFO

Article history:
Received 22 November 2008
Received in revised form 25 August 2009
Accepted 26 August 2009

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

ABSTRACT

The aim of the present research was to compare the use of multiple-choice questions (MCQs) as an examination method, to the examination based on constructed-response questions (CRQs). Despite that MCQs have an advantage concerning objectivity in the grading process and speed in production of results, they also introduce an error in the final formulation of the score. The error is traced to the probability of answering a question by chance or based on an instinctive feeling, which does not enable the ascertainment of the knowledge of the whole background included in the question. In the present study, both MCQ and CRQ 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 concerning the same topic, but this similarity was not evident for an examinee that did not possess adequate knowledge on the topic addressed in the questions of the pair. 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 CRQs, while both the "paired" MCQ test results and the CRQ test results differed significantly from those obtained from a MCQ test using positive-only scoring rule.

1. Introduction

The introduction of personal computers (PCs) in education constituted a landmark. The new technology promoted the implementation of innovative learning methods. Nowadays it may be possible to state that the prediction that learning will be substantially based on new technologies has been confirmed (Crossman, 1997; Daniel, 1996; Phillips, 1992). Computer-based learning, as well as distance learning through internet, constitutes main objectives of educational institutions.

Learning methods based on the use of PCs in the educational process have been reported to exert positive influence on the quality of teaching, without loss of effectiveness and they are also a useful study tool (Castellan,1993; Goggin, Finkenberg, & Morrow, 1997; Lehmann, Freedman, Massad, & Dintzis, 1999). Therefore computer technology might constitute a useful tool for a successful teaching and learning environment (Johnston, 1997). A wealth of references exist about the superiority of computer-aided learning in comparison to the traditional teaching methodology, concerning skill learning, such as critical thought and problem solving (Bowman, 1995; Safrit, Ennis, & Nagle, 1988).

Nowadays, a variety of online activities are offered concerning learning material, from the simplest, e.g. access to "static" course content in electronic files, to complete courses which a student may follow, including multiple forms of educational material such as text, audio, video, software and multimedia, allowing for a very wide adaptation to the computer-enhanced learning environment. In the framework of such integrated multimedia learning material are included students' self-assessment test packages, consisting of sets of questions. These questions may have the form of multiple-choice questions (MCQs) of statements that the student has to qualify as right or wrong as well as other forms. The development of such self-assessment tests provides the possibility, as well as the incitement for teachers to adopt examination procedures in classrooms equipped with PCs, i.e. PC laboratory rooms, offering to students a set of suitable questionnaires.

In the international literature, an intense discussion has taken place concerning the parameters which control the reliability and validity of examination methods (Bennett, Rock, & Wang, 1991; Bridgeman, 1991; Wainer, Wang, & Thissen, 1994). Previous research has provided

* Corresponding author. Tel./fax: +30 2105316525.
E-mail addresses: ericvent@teiath.gr (E. Ventouras), triantis@ee.teiath.gr (D. Triantis), pantsiak@ee.teiath.gr (P. Tsiakas), csterg@ee.teiath.gr (C. Stergiopoulos).

0360-1315/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compedu.2009.08.028

substantial indications that multiple-choice (MC) scores provide higher reliability and are as valid as scores from constructed-response (CR) tests (Lukhele, Thissen, & Wainer, 1994; Wainer & Thissen, 1993). Various application software exist which enable teachers to create tests, examinations and assignments, consisting of MCQs.

MCQs have the advantage that scoring is absolutely objective and may be automated, so that after the end of the examination procedure, the student may print the results, see his/her final score and his/her mistakes. Provided that the MCQs have been correctly formulated, right answering requires specialized knowledge, emphasis on detail and quick response and decision making, taking into account that a specified time might be pre-determined for answering the whole set of questions. Apart from the advantage of objectivity in checking knowledge acquisition and speed in results production, MCQs provide the possibility to the teacher to ascertain and control the degree of assimilation of knowledge, concerning the whole breadth of the subject material.

Disadvantages of MCQs are the limitation imposed on the kind of questions that might be posed, as well as the care needed by the examiner in formulating the MC items in comparison to CR items. The examinee is judged solely on the correctness of the answer he/she choose and not based on the path used for reaching the answer. Furthermore, it is not possible to investigate in depth whether the topic, which a specific question concerned, has been fully understood. Additionally, there exist the possibility that a student might collect some partial scores in the final score, by answering questions by chance, without possessing knowledge of the questioned material, if a positive scoring rule is used, e.g. students gain points for correct answers and have no losses or take reduced gains for omissions. For remedying this disadvantage, various alternative scoring methods have been proposed. The main method is having mixed-scoring rules for marking, in which students gain points for correct answers and loose points for incorrect answers. It has been shown that using such rules, participants were less willing to answer questions compared to MC tests based on positive scoring rules (Bereby-Meyer, Meyer, & Budescu, 2003). In other words, mixed-scoring rules 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.

On the other hand, using examination based on the more traditional method of CRQs, for example questions which are answered developing a set of subjects in the form of short essays, the examiner has the possibility to check more fully the knowledge of the student and especially the way that he/she developed the subject under question. The disadvantage of the CR method is focused on the fact that subjects that might be examined cannot cover a significant amount of the material taught during the courses and can also include grading difficulties.

During the last 5 years, at the Technological Educational Institution (TEI) of Athens, a considerable effort has been made in order to acquire, manage and disseminate educational material in digital form: text files, still images, audio and video files, research datasets and data acquired from experiments. 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. In particular, they are able to publish documents in various formats (Word, PowerPoint, PDF, HTML, etc.), to publish images, video and sound files, to create self-assessment tests, to administrate public or private discussion forums, to manage a list of useful links, to create student groups, to structure a modular agenda, to publish announcements in the web site and through e-mails, and to enable students to submit electronically their projects and assignments (Nafpaktitis, Triantis, Tsiakas, & Stergiopoulos, 2006; Tsiakas, Stergiopoulos, Kaitsa, & Triantis, 2005; Tsiakas, Stergiopoulos, Nafpaktitis, Triantis, & Stavrakas, 2007).

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 3 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; Stergiopoulos et al., 2006; Triantis, Stavrakas, Tsiakas, Stergiopoulos, & Ninos, 2004).

The aim of the research presented in this work was the comparison of the CRQs and MCQs examination methods. Students were examined in a PC laboratory room. The only modification to the CRQs examination method used in the past was that instead of taking the examination in a lecture room and writing answers on paper, students, being present in the PC laboratory room, typed their answers on PCs. Two MCQs examination method variants were used, one variant using MCQs graded with a positive scoring rule (referred as PSR-MCQs), and another variant using 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.

2. Methods and procedures

2.1. Comparison of CRQs and PSR-MCQs

According to the rationale exposed in the previous section, in order to compare the two examination methods, i.e. the one using MCQs with positive scoring rule (referred as PSR-MCQs) and the one using CRQs, in the examination period of June 2007, a course was selected in which the examination was implemented according to the following procedure.

2.1.1. 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 "Project Management", which belongs to the group of general interest courses, concerning the fields of Management, Economics, Legislation and Humanities, and is taught at many Departments of the School of Technological Applications of T.E.I. of Athens. Sixty-two students participated in the examination. All students had previously been given instructions for the examination and had available related study material.

2.1.2. Formulation of the questions

"E-examination" is a stand alone application developed at 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 tests can then be used either for examining students or for self-evaluation purposes. The examinee has to answer a series of questions through a user-friendly interface.

A database was constructed containing a large number ($N = 250$) of MCQs, covering the whole breadth of the subject material of the course, which included seven teaching units. By using "e-examination", a set of $\{q_1, q_2, \ldots, q_n\}$ ($n = 40$) MCQs was randomly selected from the database, taking care to cover each teaching unit proportionally. A weight was assigned to each question, depending on its level of difficulty.

Next, a set of five CRQs was created. CRQs were short essay subject development questions. Care was taken so that the MCQs were, overall, of equivalent level of difficulty with the corresponding CRQs. This enabled the comparison of the scores that would be achieved after the students would have given their answers in the two forms of examinations, i.e. the one using MCQs and the other using CRQs.

2.1.3. Examination procedure and scoring methodology

Students were examined in a PC laboratory room. In the first phase of the examination they were given the five CRQs. Answers were typed using the PC. After the end of the pre-determined examination duration time the answers were automatically formatted into pdf format files and were both e-mailed to the examiner and printed, so that each student got a copy of his/her answers.

In the second phase, MCQs were given to students. After the end of the pre-determined examination duration time, a results page was produced 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.

Based on the MCQs answers, the positive scoring rule consisted in giving positive grades (1) only to correctly answered questions. No grade was given for omissions or wrongly answered questions. The overall examination score, $M1$, was computed according to the following formula:

$$M1 = \sum_{i=1}^{n} (q_i \cdot c_i) \tag{1}$$

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 extended from one to three. As can be seen from formula (1), this method of scoring of MCQs does not impose a penalty to the student by imposition of negative marking for incorrect answers or omissions.

$M1$ score was normalized to value $m1$, whose maximum was 100, i.e.

$$m1 = \frac{M1}{\sum_{i=1}^{n} c_i} \cdot 100 \tag{2}$$

For the CRQs examination category, each of the five questions $j = 1, \ldots, 5$ was graded from $g_j = 0$ to $g_j = 20$. The overall examination score $m2$ was extracted, as the sum of the partial grades, and is by definition normalized to a maximum value of 100:

$$m2 = \sum_{j=1}^{5} g_j \tag{3}$$

In both examination categories 50.0/100.0 was the minimum normalized score required for passing the examination.

In order that the advantages of MCQs examination methods that have been stated in Section 1 are put to benefit in an actual examination process, indications should be provided for accepting MCQs examination methods as viable alternatives for the CRQs examination method, which might be assumed to be the "gold standard" method. Such an indication might be provided, if the scores obtained through an MCQs method are statistically indistinguishable from the scores obtained from the CRQs method. By adopting this line of thought, the hypothesis H_0 to be tested, in comparing the PSR-MCQs and the CRQs examination methods, could be stated as: "The mean of the distribution of scores $m1$ obtained using the PSR-MCQs examination method is equal to the mean of the distribution of scores $m2$ obtained using the CRQs examination method".

2.2. Comparison of CRQs and paired MCQs

In the next examination period (September 2007), concerning the same course, an examination method based on a refined version of MCQs, using pairs of MCQs, was implemented 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 exposed above.

From the initial database of 250 MCQs created for the examination period of June 2007, the 40 questions that were used in the examination period of June 2007 were excluded. Then, from the reduced database, with the same procedure described in Section 2.1.2, a set of questions was selected $\{q_{a1}, q_{a2}, \ldots, q_{ak}\}$ ($k = 30$). With those questions as reference, another set of new MCQs $\{q_{b1}, q_{b2}, \ldots, q_{bk}\}$ ($k = 30$) was composed, with each question q_{bi} possessing a similarity to question q_{ai} ($i = 1, \ldots, k$), constituting 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 $2k = 60$ 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. In Table 1 an example is given of such a pair of questions. The user interface for MCQs, which was common for the two examination periods, is presented in Fig. 1.

Therefore, in the examination period of September 2007, where 63 students were present, two categories of examination were given to them: a set of $2k = 60$ MCQs $q_{a1}, q_{a2}, \ldots, q_{ak}$ and $q_{b1}, q_{b2}, \ldots, q_{bk}$ and a set of five CRQs. Care was again taken so that the MCQs were, overall, of equivalent level of difficulty with the corresponding CRQs.

Table 1
Example of paired questions.

Question q_{a8}	Question q_{b8}
Which of the following does not characterize a «project»? ○ Upgrading, evolution. ○ Breaking the status quo. ○ Uniqueness. ○ Priorities readjustment.	*Which of the following constitute characteristics of a «project»?* ○ Stable environment. ○ Low risk. ○ Readjustment of responsibilities of the persons of the group. ○ Upgrading.

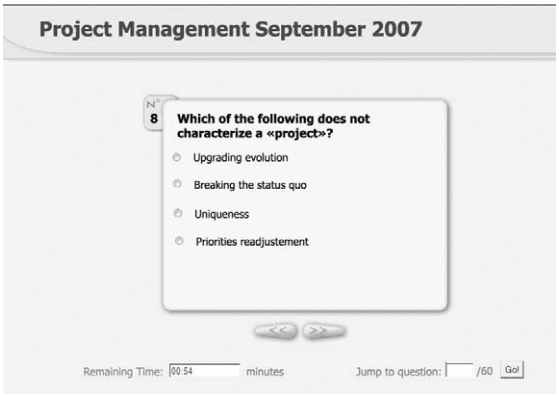

Fig. 1. User interface of the electronic examination for MCQs. The MCQ presented to the examinee is MCQ q_{a8} of Table 1.

For the MCQs category, two scores were computed as follows:
Score M1 was computed according to:

$$M1 = \sum_{i=1}^{30} (q_{ai} + q_{bi}) \cdot c_i \tag{4}$$

where

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

This method of scoring ignores any similarities existing between the questions of the pair and, as in the methodology exposed in Section 2.1.3, does not impose a penalty to the student by negative marking for incorrect answers. Therefore, through this scoring algorithm, the final score extracted corresponds to the MCQs examination method used in the previous examination period, i.e. the PSR-MCQs examination.

Score M3 was computed using formula (4) but the sum of the partial scores for the MCQs of each pair was computed as follows:

$$q_{ai} + q_{bi} = \begin{cases} 5/2 & \text{if both } q_{ai} \text{ and } q_{bi} \text{ are correct} \\ 1/2 & \text{if either } q_{ai} \text{ and } q_{bi} \text{ is correct} \\ 0 & \text{if both } q_{ai} \text{ and } q_{bi} \text{ are wrong} \end{cases} \tag{6}$$

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. $M1$ and $M3$ were next normalized to values $m1$ and $m3$, with maximum value 100, according to the formulae:

$$m1 = \frac{M1}{\sum_{i=1}^{30} 2c_i} \cdot 100 \tag{7}$$

and

$$m3 = \frac{M3}{\sum_{i=1}^{30} 2.5c_i} \cdot 100 \tag{8}$$

Finally, for the CRQs examination category, as in Section 2.1.3, each of the five questions $j = 1, \ldots, 5$ was graded from $g_j = 0$ to $g_j = 20$ and the overall examination score $m2$ was extracted, as the sum of the partial grades.

The comparison of the paired MCQs examination method to the CRQs examination method and the PSR-MCQs examination method aimed at providing indications for accepting MCQs examination methods as viable alternatives for the CRQs examination method. By adopting the same line of thought as in Section 2.1.3, the hypothesis H_0 to be tested in comparing the PSR-MCQs, the CRQs, and the paired MCQs examination methods could be stated as: "The mean of the distribution of scores $m1$, obtained using the PSR-MCQs examination method, the mean of the distribution of scores $m2$, obtained using the CRQs examination method, and the mean of the distribution of scores $m3$, obtained using the paired MCQs examination method, are equal". If hypothesis H_0 is rejected, i.e. the overall differences between

the three means are significant, then post-hoc pair-wise comparisons, with adjustment for multiple comparisons, should be used, in order to check the three "secondary" hypotheses, namely $H_{0(\text{PSR-MCQs to CRQs})}$ (i.e. "The mean of the distribution of scores $m1$, obtained using the PSR-MCQs examination method is equal to the mean of the distribution of scores m2, obtained using the CRQs examination method"), $H_{0(\text{PSR-MCQs to paired MCQs})}$ (i.e. "The mean of the distribution of scores $m1$, obtained using the PSR-MCQs examination method is equal to the mean of the distribution of scores $m3$, obtained using the paired MCQs examination method") and $H_{0(\text{CRQs to paired MCQs})}$ (i.e. "The mean of the distribution of scores $m2$, obtained using the CRQs examination method is equal to the mean of the distribution of scores m3, obtained using the paired MCQs examination method").

3. Results and discussion

In Table 2 the descriptive statistical values related to the two examination methods used in the June 2007 examination are presented, i.e. mean values and standard deviation (SD) of the normalized scores $m1$ and $m2$ taken from the two methods for the whole set of students examined, and for reduced sets of students, e.g. for those students who passed the examination (normalized score $\geqslant 50.0$), for those who failed the examination and for those who took a normalized score $\geqslant 70.0$. Furthermore, the difference of the mean values of the normalized scores $m1$ and $m2$, as well as the percentage of students belonging to each reduced set is given.

Paired samples t-test, for the whole set of students indicated significant difference ($p < 0.001$) between the two examination methods. Therefore, the tested hypothesis H_0 has to be rejected at $p = 0.001$ level of significance. From the results it is obvious that the evaluation of the students with standard MCQs gives greater success rates and scores than with CRQs. This bias is evident also by the regression line of $m1$ to $m2$ (Fig. 2) 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. These discrepancies are reduced to insignificance ($p = 0.164$) or near insignificance ($p = 0.047$) when the set of students comprised only those students who got a normalized score greater than 50.0 or 70.0

Table 2
Descriptive statistics of June 2007 examination.

Set of students	$m1$ (mean, SD)	$m2$ (mean, SD)	Mean value difference ($m1 - m2$)	Number (percentage) of students per examination method	
				MCQs examination method	CRQs examination method
Whole set ($n = 62$)	57.63 (20.31)	45.21 (21.87)	+12.42		
Set of students who passed successfully the examination (normalized score $\geqslant 50.0$)	69.90 (13.15)	65.64 (10.92)	+3.36	40 (64.5%)	28 (45.2%)
Set of students who failed the examination	35.32 (8.29)	28.38 (11.86)	+6.94	22 (35.5%)	34 (54.8%)
Set of students who got a normalized score $\geqslant 70.0$	81.20 (7.85)	75.83 (5.57)	+5.37	20 (32.2%)	12 (19.3%)

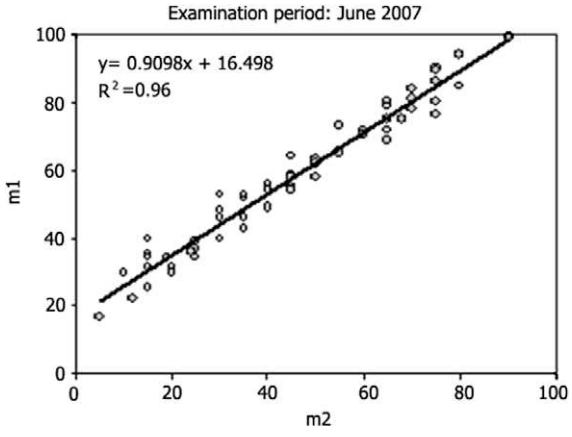

Fig. 2. Regression line of normalized score $m1$ to normalized score $m2$.

Table 3
Descriptive statistics of September 2007 examination.

Set of students	$m1$ (mean, SD)	$m2$ (mean, SD)	$m3$ (mean, SD)	Number (percentage) of students per examination method		
				MCQs score computed with formula (5)	CRQs	MCQs score computed with formula (6)
Whole set ($n = 63$)	58.25 (18.65)	46.24 (20.37)	45.95 (20.28)			
Set of students who passed successfully the examination (normalized score $\geqslant 50.0$)	70.05 (12.34)	64.89 (13.35)	64.46 (12.72)	39 (62%)	27 (43%)	28 (44%)
Set of students who failed the examination	39.08 (8.18)	32.25 (11.55)	31.14 (10.65)	24 (38%)	36 (57%)	35 (56%)
Set of students who got a normalized score $\geqslant 70.0$	81.39 (7.45)	78.55 (8.54)	77.08 (6.89)	18 (29%)	11 (17%)	12 (19%)

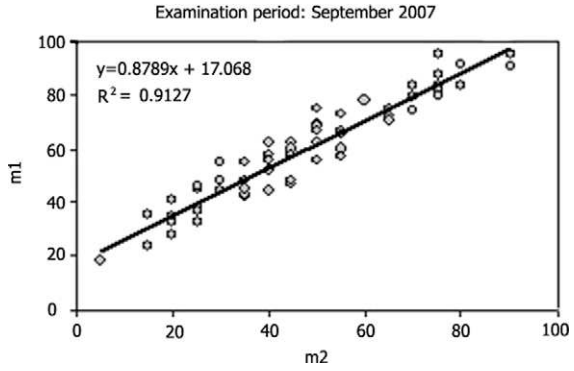

Fig. 3. Regression line of normalized score *m*1 to normalized score *m*2.

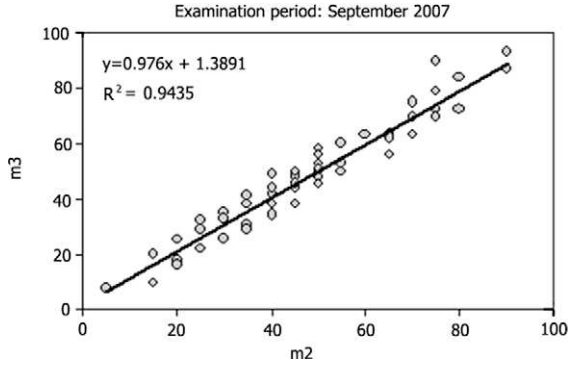

Fig. 4. Regression line of normalized score *m*3 to normalized score *m*2.

respectively. Despite the fact that the maximum effort was taken so that examination categories with MCQs and CRQs are compatible concerning the content of the questions and their degree of difficulty, the two examination methods are not sufficiently equivalent.

In Table 3 the descriptive statistical values related to the three examination methods used in the September 2007 examination are presented, i.e. mean values and standard deviation (SD) of the normalized scores *m*1, *m*2 and *m*3 taken from the three methods for the whole set of students examined, and for reduced sets of students, e.g. for those students who passed the examination (normalized score $\geqslant$ 50.0), for those who failed the examination and for those students who took a normalized score $\geqslant$ 70.0. Furthermore, the number and the percentage of students belonging to each reduced student set is given.

For the whole set of students who took the September 2007 examination, repeated-measures ANOVA with one within-subjects factor (method of examination, three levels) indicated that the within-subject effect was significant (F1.696,105.169 = 239.250, $p < 0.001$, degrees of freedom were corrected for non-sphericity according to the Greenhouse–Geisser procedure). Therefore, the hypothesis H_0 to be tested, that the overall differences between the means of the normalized scores *m*1, *m*2 and *m*3 are not significant, has to be rejected at $p = 0.001$ level of significance. ANOVA was followed by planned comparisons between each of the examination methods, assessed with post-hoc Bonferroni pair-wise comparisons at 0.05 level of significance. Significant differences existed between *m*1 and *m*2 ($p < 0.001$) and between *m*1 and *m*3 ($p < 0.001$). Therefore the secondary hypotheses $H_{0(\text{PSR-MCQs to CRQs})}$ and $H_{0(\text{PSR-MCQs to paired MCQs})}$ had to be rejected at $p = 0.001$ level of significance.

These results indicate that the paired MCQs examination method with bonus/penalty adjustment (resulting in normalized score *m*3) is statistically equivalent to the CRQs examination method (resulting in normalized score *m*2), 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 PSR-MCQs examination method (resulting in normalized score *m*1). 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 comparing the regression line of *m*1 to *m*2 (Fig. 3) with the regression line of *m*3 to *m*2 (Fig. 4). 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.

Additionally, *t*-test indicated that the examination method using CRQs presented remarkable similarity in results between the two examination periods ($p = 0.858$). The presumed equivalency of the PSR-MCQs examination method of the June 2007 period, with the score computation using formula (6) for the paired MCQs of the September 2007 period, is also ascertained by the *t*-test comparison ($p = 0.786$).

4. Conclusions

Based on the results of the present study, indications are provided that the examination method based on paired MCQs in a PC laboratory room may constitute a reliable tool for the evaluation of the performance of the students. The use of the combination of bonus and penalty in the pairs of MCQs, in an implicit way to the examinee, alleviates the positive grade bias introduced by PSR-MCQs, without the

need of negative markings for each wrongly answered question. In this way the advantages of examinations based on MCQs (Wainer & Thissen, 1993) can be fully exploited, concerning the speed of results production and the transparency of the scores given, while offering to the examiner the possibility to check students' performance on almost the whole breadth of the topics covered by the material taught in the lectures. 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.

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.

Bowman, A. (1995). Teaching ethics: Telling stories. *Nurse Education Today, 15*(1), 33–38.

Bridgeman, B. (1991). Essays and multiple-choice tests as predictors of college freshman GPA. *Research in Higher Education, 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* (pp. 19–23). NJ: Educational Technology Publications.

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

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.

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement, 31*, 234–250.

Nafpaktitis, D., Triantis, D., Tsiakas, P., & Stergiopoulos, C. (2006). Using new technologies for teaching power electronics and assessing students. In *Proceedings of the IEEE-ICELIE 2006 first international conference on e-learning in industrial electronics* (pp. 125–129).

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.

Safrit, M., Ennis, C., & Nagle, F. (1988). The use of problem-solving skills in computer-aided instruction: An evaluation. *Journal of Computer Assisted Learning, 4*, 227–232.

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. In *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, 1*, 26–30.

Tsiakas, P., Stergiopoulos, C., Kaitsa, M. & Triantis, D. (2005). New technologies applied in the educational process in the T.E.I. of Athens. The case of "e-education" platform and electronic examination of students. Results. In *Proceedings of the 2005 5th WSEAS international conference on distance learning and web engineering (DIWEB '05)* (pp. 238–242).

Tsiakas, P., Stergiopoulos, C., Nafpaktitis, D., Triantis, D., & Stavrakas, I. (2007). Computer as a tool in teaching, examining and assessing electronic engineering students. In *Proceedings of the IEEE-EUROCON 2007* (pp. 2490–2497).

Wainer, H., & Thissen, D. (1993). Combining multiple-choice and constructed-response test scores: Toward a Marxist theory of test construction. *Applied Measurement in Education, 6*, 103–118.

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.