

Paper-Based versus Computer-Based Testing in Engineering Education

Şeniz Özalp-Yaman
Atılım University, Faculty of Engineering,
Chemistry Group, 06836 Incek
Ankara, Turkey
seniz@atilim.edu.tr

Nergiz E. Çağiltay
Atılım University, Department of Software
Engineering, 06836 Incek
Ankara, Turkey
nergiz@atilim.edu.tr

Abstract-- Using computers for assessment can provide several benefits for educators and test-takers. However, in the literature, there is no consensus on the equivalence of paper-and-pencil and computer-based test environments. Accordingly, more evidences are needed especially for the engineering education. In this study, students' performance on different test modes was evaluated on 209 first year engineering students of a chemistry course. The results of this study showed that, there is no significant performance difference between paper-and-pencil and computer based tests. By comparing results with the previous studies, this study concludes that, personal characteristics of test takers, the features of computer-based testing systems and the test content are all possible confounding factors when comparing test modes and need to be considered by the implementers. The results of this study show that, once these factors are controlled, students' performance on computer-based tests and paper-and-pencil tests in chemistry courses for the engineering students will not vary. This finding is encouraging the educators to get benefits of computer-based tests without any affect on students' performance.

Index Terms— Computer based exams, Test-Mode Effect, Paper-and-pencil Based Exams, engineering education

I. INTRODUCTION

Exams are an important instrument of student assessment. As stated by Brown, Race and Bull [1] the style of assessment can have an important influence on student learning. It has been suggested that if an aspect of a course is not assessed, students will probably not learn it [2]. Hence, assessing students' performance is always an important issue for educational systems. Depending on the number of students, it is not easy to implement such techniques more often. In that sense, Computer Based Testing (CBT) systems can provide alternatives for implementing tests more often on different educational settings.

Even though there is a trend toward alternative assessment methods, CBT is definitely in ascendancy, especially in distance education, certification, and licensure [3]. According to Sawaki [4], in order to support construct validity of computerized tests such that the construct being measured is not being affected by the

mode of presentation, the equivalence of corresponding conventional and computerized test forms must be established from various direction. According to Ricketts and Wilks [5], issues related to student performance should be carefully considered when computer-based assessment is introduced. According to them, mode of presentation of assessment can significantly influence student performance. Questions about differential impact of paper-and-pencil (P&P) versus CBT on test performance are referred to as test mode effects.

The literature has contradictory results on the CBT vs. P&P tests. For example, there are some studies showing that there is no significant performance difference between P&P tests and CBT [6]-[14]. Among those studies, Ashton et al [14] have analyzed the medium effect (the screen dumps of the computerized version vs. computerized version of paper test) and rewording effect (original paper test vs. computerized version of paper test). They could not find any evidence of either medium effect or rewording effect.

However, results of some other studies showed that, students perform superior on the P&P test than the CBT [15], [16]. Contrary to these findings, in some cases students benefit most from CBT compare to P&P based tests [17]-[20].

By analyzing the findings in the literature, we can categorize main confounding factors when comparing test mode as: personal characteristics of test takers, features of computer-based testing systems and test content. The related literature shows that some personal characteristics of test takers may have impact on students' performance on different test modes. There are some studies showing that gender, socioeconomic status and computer experience affect performance on computer based testing [15], [18], [21]. For example, results obtained from 1,114 examinees that completed computer and P&P versions of the Graduate Record Examination indicated that the overall performance of male examinees on the computer version was better than would be expected from their paper scores, whereas female examinees performed better on the paper version than their scores on computer version [22]. Although gender does not interact with mode of test

administration for relatively low-stakes tests such as psychological assessments, gender differences in perceived self-efficacy regarding complex computer tasks are evident in other studies [22], [23]. It may be possible, however, to decrease gender differences by preparing female students for complex computer tasks. Consequently, female students with more experience in relevant computer tasks may feel more prepared for high-stakes, complex computer tasks. Specifically, female students with CBT experience most likely feel prepared for CBT [24]. Computer anxiety is considered one of the significant factors that would have negative effect on students' performance in computer-based testing [25]. Bugbee [3] pointed out that computer anxiety, combined with other things like test anxiety, computer experience, may influence the test taker. On the other hand, for some studies, computer experience and anxiety about using computers are not confounding factors for the test-mode effect [26]. When considering the mode of presentation, Noyes, Garland and Robbins [27] found that cognitive workload associated with P&P and CBT tasks can be another test-mode effect. They found that, the computer-based test should require more effort compare to the P&P tests.

Some features of a CBT system such as the reliability of the CBT system, abilities of the system, user interface design of the system, font sizes and font types can be another confounding factor. Supportive exam types such as essay type or multiple choice types as well as the evaluation procedure of the exam can all be considered in this category. For example, Bodmann, & Robinson, [11] did a research on user interface design of CBT systems such as various levels of flexibility to change and review answers. The results of this study show that, students completed the least-flexible mode faster than the other two modes [11]. In that context, Ashton et al [28] have used a tool called PASS-IT, which is capable of giving partial credits for mathematical calculations. In this study, they have shown that, testing platform which uses technology in a proper way, have minimized the effect of test mode on test takers' performance for the partial credit mathematical exams. Ricketts & Wilks [5] also showed that presentations which require scrolling are less acceptable than those in which questions are presented one at a time.

Test content should also be considered as another factor affecting test performance in different test-modes. For example, Russell [15] found that about 20% of the students who performed the math test on computer indicated that they had difficulty showing their work and/or needed scrap paper to work out their solutions. Another issue is the reading speed of the test takers which also can be related as test content. Generally, the literature review suggests that reading from computer screen is slower than that of on paper, which may positively or negatively affect test performance. For example, high rates of online reading speeds are positively correlated with good performance on the CBT TOEFL subset of reading

comprehension [29]. In that sense, if the test content requires long reading passages, then reading speed of the test takers need to be considered as well.

From the review of the literature, we understand that, both modes of tests (CBT and P&P) are some how different from each other and there is no consensus on the equivalence of P&P and CBT. Additionally, there are not many studies in the literature that focus on the engineering education students. On the other hand, many characteristics of engineering students and engineering education systems need to be considered for this specific domain. In order to better understand the factors effecting test mode for the engineering students, we need more evidences. Accordingly, this study aims to evaluate and compare engineering students' performance on P&P and CBT.

II. RESEARCH METHODOLOGY

This study was conducted within 209 students. Main research question of this study is,

Do engineering students perform better or worse on computer based test compare to paper and pencil test?

Students took three midterms and a final exam during the semester. In order to answer the research question, in the second midterm, the students were randomly divided into two groups according to their last names. The first group of the students (96 students) took the CBT version; whereas the second group of students (113 students) took the exam in P&P form.

The students did not have any practice on the computer-based exam interface before the midterm exam. After the exam, interviews were conducted with 7 students (2 female) who have taken CBT. Each interview took 45 minutes and they were held during the following week of the examination. Students were asked several questions about their feelings and opinions of the CBT that they had taken. In addition to this, an interview was conducted with the course instructor, to get her comments about both versions of the exams. Each interview session was recorded by a tape recorder and later transcribed by the researcher. To measure examinee characteristics such as gender, department, computer ownership, computer experience and CGPA, a questionnaire was used. On the test day, students first completed the questionnaire and then took their exam either on computer or on paper.

A. Course Description

In that university students are thought and examined in English. General Chemistry course (CHEM102) is a service course in our university, which is offered for all first year engineering students. The content includes introduction to atomic theory, chemical stoichiometry, thermochemistry, electronic structure, molecular structures, gases, properties of solutions, chemical kinetics, electrochemistry, and introduction to

thermodynamics. The course has two parts: theoretical and laboratory. The theoretical part covers 51 hours of instructions during the semester. The laboratory part covers 7 experiments in total. The instructor teaches the class in lecture format and assessment is made by 3 midterms, one final exam and laboratory assignments. Multiple choice or short answered questions are selected from each topic in order to cover whole chapter.

B. Midterm Exam

The second midterm exam covers thermochemistry, electronic structure, molecular structures and gases. There were 20 multiple choice types of questions in the exam. Students were asked to select the most appropriate answer among the given five different choices for each question. Most questions did not require mathematical calculations, however for a few questions students needed to make some mathematical calculations. Students were also given the periodical and the Molecular Geometries Based on VSEPR Theory tables (supportive materials) to be used to answer questions.

Both exams (P&P and CBT) were organized in the same way. However, in the computer version, there was only one question on each page, whereas in the P&P version of the test there were approximately five questions on each page. The supportive materials were provided as a web-link (Figure 1) in the computer version whereas they were provided as separate pages in the P&P version. Both exams were limited in time and the students were asked to finish it in 90 minutes. We have offered the second midterm in both versions by switching the groups in order to provide both groups a fared equality in the course. However, we did not consider the results of the second midterm in our study.

As shown in the Figure 1, the computer-based exam page is divided into three sub-windows. Students can always open a new window to reach the supportive materials by using the left window. In the right window, students read each question, type the answer for each question and by using the navigation buttons they can move around questions. By using the bottom window, students can finish the exam and submit their exam results to the server.

The question numbers are not shown on the screen. Font types used in both versions of the exams are the same. Both groups of students were provided some extra empty sheets of papers to make necessary calculations. Figure 2 shows the P&P version of a sample exam page.

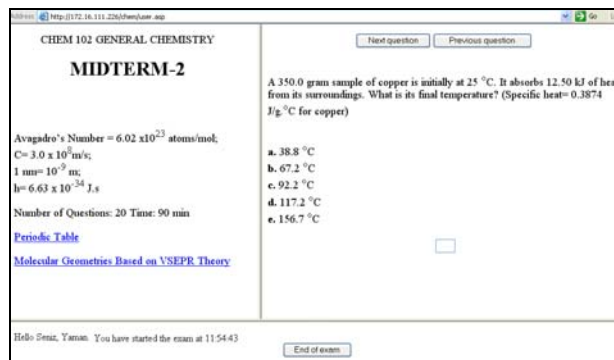


Figure 1. A Sample CBT Exam Page

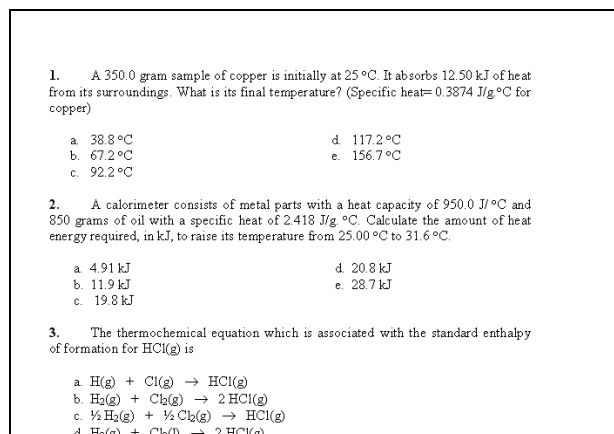


Figure 2. A Sample P&P Exam Page

C. Participants

The participants of this study were first year engineering students (from software engineering (SE), Manufacturing Engineering (MFGE), Mechatronics Engineering (MECE), Civil Engineering (CE), Industrial Engineering (IE), Electrical & Electronics Engineering (EE) and Computer Engineering (CENG)). Before taking this course, the previous semester, all of the students have completed a computer literacy course which is offered as two hours theoretical and two hours laboratory sessions in each week. The content of that course covers an overview of information technology, hardware and software components, CPU, peripherals, I/O devices, primary and secondary storage units, data communications, networks, and Internet. Word processing, spreadsheets, presentation software, and Internet applications are also introduced during the laboratory hours.

As shown in Table 1, 35 female and 174 male students participated in this study. Since in engineering departments the number of female students is usually lower according to the male students, this situation is also reflected in our sample as well which is very usual for the engineering departments.

Table 1. Students' Profile - Gender

Group	Female	Male	Total
CBT	16	80	96
P&P Test	19	94	113
Total	35	174	209

Table 2 summarizes students' distribution among the departments. Most of the students were from industrial engineering (IE) department. Among the students who have taken the CBT, only 10 of them were from computer engineering department. The rest were the students from other engineering departments.

Table 2. Students' Profile - Departments

Group	SE	MFGE	MECE	CE	IE	EE	CENG	Total
CBT	2	14	2	18	37	13	10	96
P&P	2	11	10	14	47	13	16	113
Total	4	25	12	32	84	26	26	209

III. RESULTS

In this section the results of the study presented in two parts: quantitative and qualitative.

A. Quantitative Data

The study was conducted at a private university. Compare to public universities, students' socio-economic status (SES) is higher in private universities of Turkey. As shown in Table 3, most of the students (203 students, 97%) have reported that, they have their own personal computers.

Table 3. Students' Profile – Computer Ownership

	Yes	No	Total
CBT	94	2	96
P&P Test	109	4	113
Total	203	6	209

In order to better understand students' computer experience, some questions were asked in the questionnaire. Students' responses on these questions summarized in Table 4. Majority of students like computers (79%), use computers very often (83%), use the Internet and web applications (85%) and access their e-mails very often (94%).

Table 5 summarizes the averages of Cumulative Grade Point Average (CGPA) for both groups. This shows that, the average CGPA of computer-based test group is slightly higher than P&P group. However in general, the female students' average CGPA is higher than that of males.

Table 4. Students' Profile – Computer Experience

Question	Students' Response
I use computers every day	173 (83%)
I frequently use the Internet and web applications	178 (85%)
I like computers	165 (79%)
I often access my e-mail	196 (94%)

Table 5. Students' Profile – CGPA

Group	Mean	Std. Deviation
CBT	1,34	0,86
P&P Test	1,13	0,75

An independent t test was conducted to evaluate if there is a statistically significant difference according to their CGPA between two groups. The result was not significant, $t(206)=1,86$, $p=0,064$. Students who have taken the exam on computer (M=1,34, SD=0,86) on the average have close CGPA with the students who have taken the exam on P&P (M=1,13, SD=0,75).

In the second midterm, 96 students took the exam in the CBT form, while 113 of them took the exam in P&P version. Table 6 summarizes the exam results.

Table 6. Exam Results

Test Mode	N	Mean	Std. Deviation
CBT	96	42,96	20,10
P&P Test	113	40,08	19,28

The average exam score of students who took the exam in CBT form is slightly higher than the P&P group. An independent-sample t test was conducted to evaluate if there is a significant difference between the exam results of these two groups. The result was not significant, $t(207) = 1,055$, $p= 0,77$. On the average, the students who took the exam on CBT form (M= 42,96, SD=20,10) had similar grades as the P&P group (M=40,08, SD=19,28).

A 2x2 ANOVA was also conducted to evaluate the effects of test mode (CBT or P&P) and gender on students' midterm scores. Means and standard deviations for the exam scores as a function of the two factors are presented in Table 7.

Table 7. Gender Effect

Gender	Test Mode	Mean	Std. Deviation
Female	CBT	53,38	18,54
	P&P Test	50,26	20,51
Male	CBT	40,88	19,85
	P&P Test	38,02	18,46

The results of the ANOVA indicated a non significant effect for test mode, $F(1, 209) = 0,69, p=0,40$. Results also indicates a significant effect for gender, $F(1,209) = 12,01, p<0,01$. However, the results indicated a non significant interaction between test mode and gender, $F(1,209) = 0,001, p=0,97$. Female students were more successful in the exam than the male students. However, the test mode (taking the exam on computer or P&P), did not effect this result.

B. Qualitative Data

The results of interviews are also supporting the quantitative results. During the interviews all students have reported that they use computers at least one hour in a day. Even two students have declared that they use computers more than 7 hours in a day. When we asked what they think about taking the exam on computer, three students said that, at the beginning they felt uncomfortable, however after ten minutes they got used to the system and they felt better. For example, one student said, "For the first 5 minutes I worried a little if I would accidentally press a wrong button. Because it was my first time to take the exam on computer. However, 5 minutes later I started to feel better". Another student also reported:

The P&P exams are what we have used to. However, taking the exam on computer is something very new to us. For this reason at the beginning I was in panic. After 10 minutes, I continued the exam as if I am taking it in P&P format.

Parallel to this, another student also declared that taking the exam on computer or paper does not affect his performance:

Taking the exam on computer or on P&P does not matter. They are the same. In both versions, you have to use paper to make some calculations. Once you know the concepts that the exam covers, they are the same. However, for those who are not familiar with computers it is normal that they feel a little bit uncomfortable. I liked the idea of taking the exam on computer.

When we asked them, what would happen if they had taken the exam on paper, all of them said that their grades would not change. For example according to one student, his exam grade would also not change. He said:

My grade would not change. The content is the same. It was easy for me. It is an exam, taking it on computer or on paper are the same.

When they were asked which format is better, taking the exam on computer or on paper, 6 of them declared that it does not matter, but two of them preferred computer version. For example one student said,

I think taking the exam on computer is easier and better. In fact they are the same; we only look at the monitor instead of the paper. For the next time I would prefer to take the exam on computer. Because reading the questions one by one on the screen increases my concentration on each question.

Another student prefers to take the exam on computer. He said, "I would like to take all of my exams on computer". One student said, "I would not say no for both of them. However to be sure, I would prefer the P&P version. You may face with some technical problems on the computer environment, such as losing the Internet connection etc."

Since students did not have any experience with the CBT system, at the beginning they were not very comfortable with it; however, after some time they felt better. They also said, this did not affect their scores. And finally, all students believe that their performance would not change if they have taken the exam on P&P form.

IV. DISCUSSIONS AND CONCLUSIONS

In this study, data were collected and analyzed to better understand the test mode effect on students' test performance. The subjects' socioeconomic status were high, most of them were computer owners and computer literate. The results of this study show that, students' performance on the CBT and P&P versions of the exam does not differ significantly. The qualitative data also supports this result. All students declared that, their exam grade would not change if they had taken the exam on paper.

In this study, there were some differences between the computerized and P&P versions of the exams (i.e. number of questions in each page and question numbers). However this difference did not affect the test results significantly. This result supports findings of Ashton [14] which they could not find any evidence of either medium or rewording effect on test mode.

Gender effect was also tested and we could not find any relation between gender and test mode.

The results of this study supports the results of several studies which have found no significant difference on students' scores in CBT and P&P testing environments, [6]-[14], and [28].

The result of this study is contradictory to some of the results found in the literature. For example, Lee and

Weerakoon, [16] and Russell [15] have found that, students perform better on P&P based tests than the CBT. However, they have also reported that, their subjects have little computer experience. For example, Lee and Weerakoon [16] reported that 27% of the students had moderate to high experience with email and 37% of students having moderate to high experience with the web. However, in our case, students have declared that they frequently use the Internet and web-based applications (85%) as well as they frequently access their emails (94%). Similarly, in Participants of Russell's study [15] subjects did not have a great deal of experience working with computers. We believe computer experience of test takers is a confound factors when comparing test modes.

The result of this study is also contradictory to the results of studies which have found that students' performance on CBT is higher than that of P&P [17]-[20]. However, Clariana and Wallace [19] have conducted their research on "Computer Fundamentals Course" which covers the fundamental concepts and use of a computer system. Similarly, Bocij and Greasley [18] conducted their research on elementary courses in Information Systems, Computer Science and Information technologies. We believe that, in courses teaching about computer systems, asking questions in computer could benefit students to better visualize and understand the question and this may positively affect their performance on the test.

On the other hand, Bugbee and Bernt [17] could not rule out substantial differences among the groups that have taken the exam on different exam modes. This may cause the performance difference between the test modes. Pomplun et al [20] conducted their study on a speeded reading comprehension placement test. They have also reported that, the main reason for the higher scores on CBT could be the clicking a mouse which allows an examinee to move quicker through the items than when forced to record answers with a pencil on a bubble sheet.

This study shows that, while considering the test-mode effect, characteristics of test takers, features of computer-based testing systems and the test content are possible confounding factors when comparing test modes. Once these parameters are controlled, on CBT, similar test performance can be reached with the P&P tests. This result is very promising to better get benefits of the CBT environments in classical and distance learning environments. Accordingly, once the CBT environments are decided to be set, the possible confounding factors such as personal characteristics of test takers, the features of computer-based testing systems and the test content need to be controlled.

REFERENCES

[1] Brown, G., Race, P. & Bull, J. (1999). Computer-assisted assessment in higher education. Kogan Page: London

[2] Ashcroft, K. & Palacio, D. (1996). Researching into assessment and evaluation in colleges and universities. London, UK: Kogan Page

[3] Bugbee, Jr. A. C. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education*, 28(3), 282-299.

[4] Sawaki, Y. (2001): "Comparability of conventional and computerized tests of reading in a second language". *Language learning & Technology*, 5 (2): 38 – 59.

[5] Ricketts, C. & Wilks, S.J. (2002). Improving Student Performance Through Computer-based Assessment: insights from recent research. *Assessment and Evaluation in Higher Education*, Vol. 27, No. 5, 2002.

[6] Schaeffer, G.A., Reese, C.M., Steffen, M., McKinley R.L., & Mills, C.N. (1993). Field test of computer-based GRE general test ETS Research report #93-07 (ERIC #ED385588).

[7] Finegan, J.E. & Allen, N.J. (1994) Computerised and written questionnaires: are they equivalent?, *Computers in Human Behaviour*, 10, pp. 483-496.

[8] Russell, M., & Haney, B. (1997). Testing Writing on Computers: An Experiment Comparing Students Performance on Test Conducted via Computer and via Paper-and-Pencil. *Education Policy Analysis Archives*, 5(3), 1-19.

[9] Mason, B.J., Patry, M., & Berstein, D.J. (2001). An Examination of the equivalence between non-adaptive computer-based and traditional testing, *Journal of educational Computing Research* 24 (1) 29-39.

[10] Weinberg, A. (2001). Comparaison de deux versions d'une test de classement: Version papier-crayon et version informatisee [Comparison of two versions of a placement test: Paper-pencil version and computer-based version]. *Canadian Modern Language Review*, 57, 607-627.

[11] Bodmann, S.M. & Robinson, D.H. (2004). Speed and Performance Differences Among computer-based and paper-pencil tests, *Journal of Educational Computing Research*, 31(1) 51-60, 2004

[12] Campton, P. (2004). A Comparative Analysis Of Online And Paper-Based Assessment Methods: A University Case Study. ACEC 2004, Conference Proceedings.

[13] Euzent, P.J., Putchinski, L., Moskal, P. & Martin, T.L. (2005). Further Evidence on the Effectiveness of WebCT Versus Paper Testing, 7th Annual WebCT User Conference.

[14] Ashton, H.S., Beevers, C.E., Korabinski, A.A., & Youngson, M.A. (2005). Investigating the medium effect in computer-aided assessment of school Chemistry and college Computing national examinations, *British Journal of Educational Technology* 36(5) 771-787, 2005.

[15] Russell, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives*, 7, 20.

[16] [16] Lee, G. & Weerakoon, P. (2001). The role of computer-aided assessment in health professional education: a comparison of student performance in computer-based and paper-and-pen multiple-choice tests, *Medical Teacher*, Vol. 23, No. 2, 152-157.

[17] Bugbee, Jr., A. C., & Bernt, F. M. (1990). Testing by computer: Findings in six years of use 1982-1988. *Journal of Research on Computing in Education*, 23, 87-101.

[18] Bocij, C., & Greasley, A. (1999). Can computer-based testing achieve quality and efficiency in assessment? *International Journal of Educational Technology*, 1, nl.

[19] Clariana, R. & Wallace, Patricia (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect, *British Journal of Educational technology*, 33(5), 593-602.

[20] Pomplun, M., Frey, S., & Becker, D. F. (2002). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement*, 62, 337-354.

- [21] Sutton, R.E. (1991). Equity and computers in the schools: A decade of research. *Review of Educational Research*, 61(4), 475-503.
- [22] Parshall, C. & Kromery, J.D. (1993). Computer testing versus paper-and-pencil testing: An analysis of examinee characteristics associated with mode effect. Paper presented at the Annual Meeting of the American Educational Research Association. Atlanta, GA. (12-16 April, 1993).
- [23] Busch, T. (1995). Gender differences in self-efficacy and attitudes towards computers. *Journal of Educational Computing Research*, 12(2), 147-158.
- [24] Lynch, C. D., Whitley, T. W., Emmerling, D. A. & Brinn, J. E. (2000). Variables That May Enhance Medical Students' Perceived Preparedness for computer-based testing. *The Journal of American Medical Informatics Association*, 7(5), 469-474.
- [25] Perkins, B. (1993). Differences between Computer Administered and Paper Administered Computer Anxiety and Performance Measures. Unpublished Research Report. ED 355905.
- [26] Wise, S. L., & Plake, B. S. (1989). Research on the effects of administering tests via computers. *Educational Measurement: Issues and Practice*, 8(3), 5-10.
- [27] Noyes, J., Garland, K., & Robbins, L. (2004). Paper-based versus computer-based assessment: Is workload another test mode effect? *British Journal of Educational Technology*, 35, 1, 111-113.
- [28] Ashton, H. S., Beevers, C.E., Korabinski, A. A. (2006). Incorporating partial credit in computer-aided assessment of Mathematics in secondary education, *British Journal of Educational Technology* 37(1) 93-119, 2006.
- [29] Al-Othman, (2003). The relationship between online reading rates and performance on proficiency tests". *The Reading Matrix*, 3(3).