Student achievement in paper, computer/web and mobile based assessment

Stavros A. Nikou
Information Systems
University of Macedonia
Egnatia 156, 540 06, Thessaloniki
+30-2310-891768
stavrosnikou@sch.gr

Anastasios A. Economides
Information Systems
University of Macedonia
Egnatia 156, 540 06, Thessaloniki
+30-2310-891768
economid@uom.gr

ABSTRACT
Assessment can be delivered through paper and pencil (Paper Based Testing-PBT), through computers (Computer Based Testing-CBT) or mobile devices (Mobile Based Testing-MBT). Conclusions from previous research are not consistent regarding the performance of test taker related to the test delivery mode. Different test delivery modes may lead to differentiated student performance. The goal of this study is to bring new evidence regarding the impact of PBT, CBT and MBT on students' performance. The findings of the study indicate that there are statistically significant differences in test scores of first year undergraduate students of Economics in the subject of ICT among MBT, CBT and PBT in favor of MBT.

Categories and Subject Descriptors
H.1.2 [**Information Systems**]: User/Machine Systems – *Human factors, Human information processing, Software psychology*; K.3 [**Computing Milieux**]: Computer Uses in Education – *Computer-assisted instruction (CAI), Distance Learning*

General Terms
Human Factors

Keywords
Mobile- based assessment, paper-based assessment, computer-based assessment, mobile test, mobile learning

1. INTRODUCTION
Assessment is a fundamental activity in the learning process because it does not only evaluate learners' knowledge, understanding, abilities and skills but also it can be used to evaluate the learning outcome itself, advancing through appropriate feedback mechanisms the learning procedure.

1.1 Computer/Web-based assessment
Beyond the traditional Paper-Based Test (PBT), assessment can be delivered in a standalone computer as Computer-Based Testing (CBT) or via the Internet as Web-Based Testing (WBT), which is the most usual scenario nowadays. CBT or WBT can be either linear (examinees receive the same number of questions in the same order) or adaptive (questions from a large item pool can be tailored and presented to the examinee according to his/her previous answer). Random choice from a large questions pool, innovative and sophisticated item formats, immediate feedback, automated scoring and reporting, advanced security are among some advantages of CBT over PBT. Many tests are administered through computers nowadays and in some cases computer technology is the preferred method of choice e.g. driving license examinations, language tests etc.

1.2 Mobile-based assessment
With the rapid growth of mobile technology, devices such as mobile phones, PDAs, palmtops, smartphones, tablets and netbooks, are becoming an important part of student life as communication, entertainment and multiple purpose information processing tools. Their ease of use, reduction in the use of paper, low cost, mobility, portability, interactivity, flexibility and ubiquity, are some of their characteristics that make them so popular among students. The integration of mobile devices in learning leads to a new learning mode called mobile learning. Mobile learning can be conducted "anywhere and anytime" in any authentic environment and can be context aware. Mobile handheld devices are increasingly being used in K-12 and higher education [8]. Integrating mobile technology in learning activities can be effective not only in improving students' learning attitudes, but also in their learning achievements and motivation ([12], [21]). Mobile devices with their pervasive and ubiquitous characteristics can also facilitate the assessment procedure, leading in an innovative assessment mode, called Mobile-Based Testing (MBT). Economides in [16] addressed critical factors that affect the quality of a mobile learning application where the mobile learner performs an educational activity supported by context aware adaptations. In a general framework for adaptive mobile learning, while the mobile learner performs an educational activity, an adaptation engine personalizes the educational activity according to the context. The context can be adapted to the learner's state, the educational activity's state, the infrastructure's state, and the environment's state ([15], [17]). Based on the mobility of the m-device, it is appropriate to implement either formative assessment or self-assessment in ubiquitous m-learning environments [20]. Due to their increasing popularity, mobile devices may become one of the primary modes of test delivery in the future. The comparability of test administration modes is an issue between assessment experts, researchers, practitioners and users [48]. The impact of test takers' characteristics, on his/her performance on CBT or MBT, should be considered by educators and test developers before replacing PBT with equivalent CBT or MBT versions. Are student's performance on written tests and computerized or "mobilized" versions identical? The "test mode" effect refers to the difference of a student's performance on equivalent assessments with different modes of delivery (PBT, CBT or MBT).

2. LITERATURE REVIEW

Conclusions from previous research are not consistent regarding the performance of test taker related to the test delivery mode and the interaction between the assessment modes and test taker variables. This inconsistency is somehow expected due to the fact that there have been so many studies to different groups of examinees with different designs and data collection techniques in a wide range of content areas and a variety of item formats. The issue of equivalency among different test delivery modes is not trivial. It was found that even simple change of color scheme had a significant effect on students' performance [26]. There are two main key factors with potential impact on students' performance on different test modes: 1) personal characteristics of test takers from one side, i.e. race, gender, cognitive processing, ability, different learning styles, computer familiarity, computer anxiety, computer attitude and 2) interface and technological issues from the other: interactivity, user interface, screen size, scrolling, modes of item presentation, multimedia and graphics etc. ([13], [27], [35]).

2.1 Computer/Web- vs Paper-based

The comparison of PBT vs CBT scores has been studied since the first attempts to introduce computers in assessment [29].

2.1.1 Paper-based outerperformed

Some early researchers have pointed out that CBT produced lower students' scores than PBT ([5], [33]). It was the early times of using computer technology in assessment. Back then students had more confidence with paper than with computers.

2.1.2 Equivalent performance

Performance differences due to computer unfamiliarity have been decreased over time because of the widespread adoption of computers in everyday life activities. Noyes & Garland [35] focused on equivalency issues between the two test delivery modes and argued that greater equivalence between CBT and PBT is being achieved today (especially in standardized and closed tasks e.g. multiple choice questions) than at the early times of computers. Research ([40], [30], [6], [2], [1]) found student performance to be comparable across test delivery modes. Equivalent performance with marginal differences between the two test modes have been found also in ([28], [25]). However, there are some studies still reporting lower CBT performance in cases that require text reading with scrolling ([49], [37], [3], [11]) or graphics and mathematics manipulations [24]. Even on student reading or mathematics achievement scores, administration mode had no statistically significant effect [48].

2.1.3 Computer/Web outerperformed

On the other hand, many studies, in a variety of settings, have revealed that there is a significant difference between the two testing modes in favor of CBT. Bugbee and Bernt [4] discussing the use of computer administered testing from 1982 to 1988, found student performance on a series of CBT to outperform paper based exams. Computerized versions of vocabulary tests produced higher scores (due to the higher response speed associated with use of a mouse to record responses in contrast to a pencil and answer sheet) than the paper-and-pencil form [38]. Studies in elementary education [9] and secondary education settings [14] as well as in undergraduate level ([10], [19], [50]) demonstrate that students performed better on the computer-based test than on the paper-based test. Clarianna & Wallace in [13] also reported better scores in CBT with the gender, competitiveness and computer familiarity not to be related to this performance difference. However the high achievers had a better performance on CBT. Overall, research findings are not conclusive but there seems to be a trend indicating that PBT and CBT are comparable ([36], [48]).

2.2 Mobile- vs Paper/ Computer/ Web-based

The interest in developing and using MBT in educational assessment in schools and educational institutions has been increased in recent years. A MBT may be a simple transfer of the paper format onto the screen of the mobile device. Furthermore, more sophisticated methods can be implemented with the use of multimedia and adaptation techniques. Many studies implement nowadays adaptive personalized approaches to mobile learning exploiting learner, location and other contextual information adaptations ([21], [44], [22]). However, there are not enough studies that evaluate the use of mobile devices for testing compared to CBT, WBT or PBT and inconclusive results have been reported regarding examinee performance.

3.1.1 Equivalent performance

No significant differences in the results obtained in [39] with the different versions of the test that were observed. Segal , Doolen & Porter [43] and Treadwell [46] showed handheld-based quiz to be more efficient, that is, students completed it in less time than they needed to complete the paper-and-pencil quiz while no differences in effectiveness (student's scores) were found between the two quiz types. Also, no significant difference between the achievement level of the students who took paper, web and mobile based assessment were found ([23], [42]).

3.1.2 Mobile–based outerperformed

On the other side, the scores of fifth grade students who used handheld computers against those who used paper and pencil for the same test were significantly higher [51]. In ([31],[34]) students who were assessed using mobile phone got higher marks in English literature than those who were assessed using paper and pencil. However, there was no statistically significant difference in the students' performance due to gender. The results from the previous literature review are summarized in table 1.

3. METHODOLOGY

The main purpose of this study is to evaluate the use of mobile devices in assessment in the settings of a Greek University and compare students' achievement among PBT, CBT/WBT and MBT.

3.1 Research Questions

This study attempts to answer the following questions:

1. Are the test scores of undergraduate students different among PBT, CBT and MBT?
2. Are the test scores of male undergraduate students different among PBT, CBT and MBT?
3. Are the test scores of female undergraduate students different among PBT, CBT and MBT?

Table 1. Summary of comparability results among PBT,CBT and MBT

Test Modes	Support evidence
PBT > CBT	Bunderson, Inouye, & Olsen, (1989); Mazzeo & Harvey, (1988); Way, Davis & Fitzpatrick, (2006); Pommerich, (2004); Bridgeman, Lennon, & Jackenthal (2003); Choi & Tinkler, (2002); Keng, McClarty & Davis (2006)
PBT = CBT	Noyes & Garland (2008); Russell & Haney (1997); Mason, Patry & Berstein (2001); Campton (2004); Bodmann & Robinson (2004); Akdemir & Oguz, (2008); Macedo-Rouet, Ney, Charles, & Lallich-Boidin, (2009); Kim & Huynh (2007); Wang et al. (2008); Paek (2005)
CBT > PBT	Bugbee & Bernt (1990); Pomplun, Frey, & Becker (2002); Chin & Donn (1991) ; Coniam (2006); Choi, et al (2003); Gretes & Green (2000); Wilson et al (2001); Clarianna & Wallace (2002)
MBT = PBT & MBT = CBT	Romero, Ventura & de Bra (2008); Segal , Doolen & Porter (2005); Treadwell (2006) ; Karadeniz (2010); Shroeders (2010)
MBT > PBT & MBT > CBT	Wu & Zhang, (2010); Masri (2012); Muhanna (2011)

3.2 Research Participants

The participants in this study were 203 first-year undergraduate students, 73 males (35%) and 130 females (49%), enrolled in an introductory informatics course, in the Department of Economic Sciences of a Greek University. The course was composed of two modules: Theory and Practice. The theoretic module introduced general concepts of ICT and the practical module introduced the use of Word Processing and Internet use. Student participation in the test became on a voluntarily basis. Two weeks in advance a preliminary questionnaire among students showed a high percentage of smartphone possession and a high willingness to use them in the forthcoming assessment. These results were in accordance with the research in [18] investigating students' usage, preferences and desires regarding the use of mobile devices in education. The participants were randomly assigned into three groups according to the delivery mode of the assessment: paper and pencil, computer/web and mobile- based.

3.3 Data Collection

The assessment had 30 multiple choice questions, 25 from the theory module and 5 from the practice module. Multiple choice quizzes are suitable for assessing a learner's factual knowledge [7] and lately gained large popularity due to their efficiency and objectivity [32]. Also, they can easily be transferred from the paper version to the computer or smartphone screen. The CBT group used the computers of the University Computing Center. The MBT group had to download and install the Android quiz application by scanning the appropriate QR code that was handed out to students before the examination. The use of the CBA and MBT system was very simple. The user first had to log into the system. Each page had the question, the four possible answers and the "next"/"OK" button. The student had only to choose the right answer and then he/she had to push the "next"/"OK" button. The text was in Greek and the assessment's duration was 30 min. The maximum score, if all questions were answered correctly was 30. The interface was kept as simple as possible to avoid possible destructions. Fig. 1 shows the assessment's interface through a sample question in CBT.

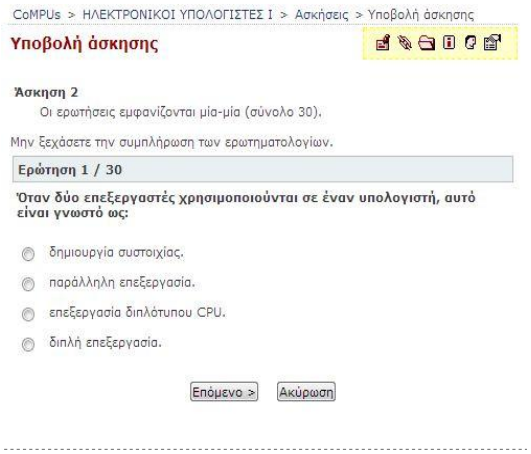

Figure 1. Sample question on CBT

Fig. 2 shows the assessment's interface through a sample question in MBT.

3.4 Data Analysis

Students' correct answers of all three modes of assessment and relevant data imported in the statistical package SPSS 20 for processing. One-way analysis of variance (ANOVA), with a significant level of 0.05, was used to test three hypotheses. ANOVA assumptions were satisfied. Scores in each testing group were normally distributed (Shapiro-Wilk normality test sig. > 0.05). Non-significant result in the Levene's test (sig.=0.285 > 0.05) indicated the homogeneity of variance.

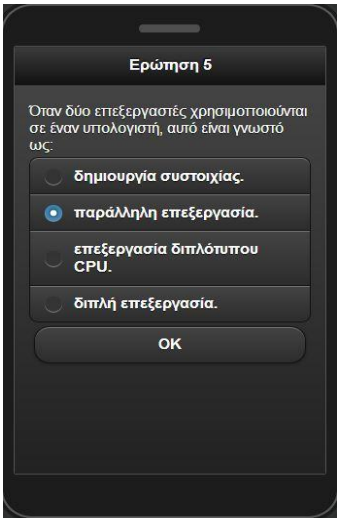

Figure 2. Sample question in MBT

Table 2 represents the distribution of students' scores in all three assessment modes. One-way analysis of variance rejects the first null hypothesis that test scores of undergraduate students were not different among PBT, CBT and MBT. (F=4.511, p=0.012 < 0.05). Table 3 represents one-way ANOVA results for students' scores in all three assessment modes. Tukey post-hoc analysis indicates that scores in MBT are significant higher (Mean Difference = 1,827 with sig.= 0.008 < 0.05) compared with the PBT equivalent (Table 4).

Table 5 represents the distribution of male students' scores in all three assessment modes. One-way analysis of variance fails to reject the first null hypothesis that test scores of male undergraduate students were not different among PBT, CBT and MBT. (F=1.255, p=0.221 > 0.05). Table 6 represents one-way ANOVA results for students' scores in all three assessment modes.

Table 7 represents the distribution of female students' scores in all three assessment modes. One-way analysis of variance fails to reject the first null hypothesis that test scores of female undergraduate students were not different among PBT, CBT and MBT. (F=3,400, p=0.036 < 0.05). Table 8 represents one-way ANOVA results for students' scores in all three assessment modes. Tukey post-hoc analysis indicates that scores in MBT for female students are significant higher (Mean Difference = 1,978 with sig.= 0.027 < 0.05) compared with the PBT equivalent (Table 8).

Table 2. the distribution of students' scores in all three assessment modes

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Paper&Pencil	99	18,21	3,895	,391	17,44	18,99	10	26
PC	53	18,85	2,964	,407	18,03	19,67	13	26
Mobile	51	20,04	3,310	,464	19,11	20,97	13	28
Total	203	18,84	3,590	,252	18,34	19,33	10	28

Table 3. One-way ANOVA comparing students' scores in all three assessment modes

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	112,376	2	56,188	4,511	,012
Within Groups	2491,259	200	12,456		
Total	2603,635	202			

Table 4. Tukey HSD Post-Hoc for students' scores in all three assessment modes

(I) TestMode	(J) TestMode	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Paper&Pencil	PC	-,637	,601	,540	-2,06	,78
	Mobile	-1,827*	,608	,008	-3,26	-,39
PC	Paper&Pencil	,637	,601	,540	-,78	2,06
	Mobile	-1,190	,692	,201	-2,82	,44
Mobile	Paper&Pencil	1,827*	,608	,008	,39	3,26
	PC	1,190	,692	,201	-,44	2,82

*. The mean difference is significant at the 0.05 level.

Table 5. the distribution of male students' scores in all three assessment modes

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Paper&Pencil	43	18,42	3,887	,593	17,22	19,61	10	26
PC	11	19,18	3,894	1,174	16,57	21,08	13	26
Mobile	19	20,05	3,440	,789	18,39	21,71	14	28
Total	73	18,96	3,791	,444	18,07	19,84	10	28

Table 6. One-way ANOVA comparing male students' scores in all three assessment modes

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	35,828	2	17,914	1,255	,291
Within Groups	999,049	70	14,272		
Total	1034,877	72			

Table 7. the distribution of female students' scores in all three assessment modes

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Paper&Pencil	56	18,05	3,929	,525	17,00	19,11	10	26
PC	42	18,76	2,721	,420	17,91	19,61	14	25
Mobile	32	20,03	3,287	,581	18,85	21,22	13	26
Total	130203	18,77	3,485	,306	18,16	19,37	10	26

Table 8. One-way ANOVA comparing female students' scores in all three assessment modes

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	79,650	2	39,825	3,400	,036
Within Groups	1487,427	127	11,712		
Total	1567,077	129			

Table 9. Tukey HSD Post-Hoc for students' scores in all three assessment modes

(I) TestMode	(J) TestMode	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Paper&Pencil	PC	-,708	,699	,569	-2,37	,95
	Mobile	-1,978*	,758	,027	-3,78	-,18
PC	Paper&Pencil	,708	,699	,569	-,95	2,37
	Mobile	-1,269	,803	,258	-3,17	,64
Mobile	Paper&Pencil	1,978*	,758	,027	,18	3,78
	PC	1,269	,803	,258	-,64	3,17

*. The mean difference is significant at the 0.05 level.

Figure 3 shows the mean scores for every group for male and female students

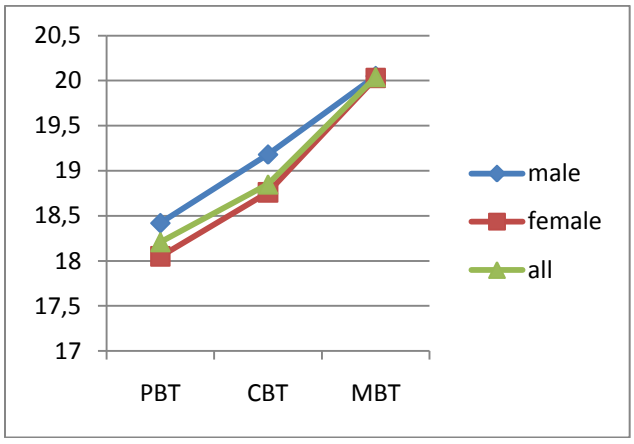

Figure 3. Means for PBT,CBT and MBT

DISCUSSIONS

The goal of this study is to bring new evidence as regards the impact of "computerized" or "mobilized" test delivery modes on students' performance. The motivation is to investigate the potential of replacing PBT and CBT with MBT in a University context. Students with mobile devices achieved a better score, while lower scores were accomplished in the paper & pencil group and in the computer group. Also, female students with mobile devices achieved a better score in MBT compared to CBT and PBT modes. Since MBT comparing to PBT and CBT produces better results, (or at least not worst), the shift to this innovative test delivery mode can be feasible and may be desirable.

However, there is the issue of comparability between the testing modes that needs to be resolved. CBT is more than a decade old. Standardization of testing procedures, lower costs, time savings, improved scoring accuracy, immediate feedback, adaptability are some benefits of CBT but still its comparability to PBT is controversial. On the other hand, the introduction of mobile devices such as PDAs, mobile phones, portable computers, into

the learning pedagogy can compliment e-learning by creating an additional channel of assessment [47]. Using mobile devices instead of computers may eliminate the need for a specialized computer classroom and can be used anywhere [20]. This shift to MBT looks interesting and promising also.

But before introducing these new test delivery modes in any assessment settings, it is crucial for educators and policy makers to examine the comparability of PBT, CBT and MBT. The incompatible findings of studies lead to the conclusion that the test mode effects of PBT, CBT and MBT will continue to be an issue.

According to Noyes and Garland in [35], equivalence of test modes, "relates to whether a task in paper form remains the same when transferred to a computer. Equivalence is going to be hard to achieve since two different presentation and response modes are being used." This is why it is easier to transfer closed tasks (i.e. multiple choice questions) from paper onto screens. Students tend to like more the multiple choice questions format. However, the multiple-choice examination does not challenge students to make an in-depth effort to study [45]. The mode of item presentation is a basic driving factor in the comparability of test modes. It is the cognitive workload that can be affected with different presentation formats. Another issue is the comfort of the examinee with the testing medium: students may be more familiar with scrolling and clicking a mouse or tapping on a touch sensitive screen instead of using the pencil as a primary writing tool. In this way they enjoy CBT or MBT and prefer them to traditional paper testing [49].

Another source of variation between "paper and screen" is the content area. Russell in [41] examined the mode effect on students' performance in three subject areas: science, math, and language arts. The students with computerized tests performed better in science, but no significant mode effects were found for language arts and math tests. Screen capacity limits and other destructors such as the difficulty to review a question, temporarily skip an item or to have an eye catching overview like a paper page may be factors that influence student's performance. Other sources of differentiation between "paper and screens" are examinee's gender as well as familiarity, anxiety and attitude related with computers or mobiles devices. MBT may cause lower anxiety levels to students with higher computer or mobile phone proficiency. Furthermore personality and psychological issues must be taken into account. MBT, under some circumstances, may stimulate students more in order to answer the questions with a higher level of concentration. The suitability of the test delivery methods needs to be investigated before any implementation. This is one future direction in our research.

MBT may be a promising alternative technique for the undergraduate students. The impact of testing mode needs to be considered though. Not all test delivery modes may be suitable for all students. Online test delivery may be helpful for students more comfortable to use computer or mobile technology. They probably should be able to choose the test method that fits to their style and preferences. CBT and MBT could complement conventional paper testing and not necessarily replace it.

REFERENCES

[1] Akdemir, O. and Aguz, A. 2008. Computer-based testing: An alternative for the assessment of Turkish undergraduate students. *Computers & Education,* 51,1198–1204.

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

[3] Bridgeman, B., Lennon, M. L., and Jackenthal, A. 2003. Effects of screen size, screen resolution, and display rate on computer-based test performance. *Applied Measurement in Education*, 16, 191-205.

[4] Bugbee, Jr., A. C. and 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.

[5] Bunderson, C. V., Inouye, D. K., and Olsen, J. B. 1989. The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement (*3rd ed.) (pp. 367–407). NY: American Council on Education – Macmillan.

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

[7] Chao, K.J., Hung I. C., and Chen, N.S. 2011. On the design of online synchronous assessments in a synchronous cyber classroom. *Journal of Computer Assisted Learning* , 28(4), 379-375.

[8] Cheung, W.S. and Hew, K.F. 2009. A review of research methodologies used in studies on mobile handheld devices in K-12 and higher education settings. *Australasian Journal of Educational Technology*, 25(2), 153-183.

[9] Chin, C. H. L., and Donn, J. S. 1991. Effects of computer-based tests on the achievement, anxiety, and attitudes of grade 10 science students. *Educational & Psychological Measurement*, 51(3), 735–745.

[10] Choi I., Kim K., and Boo, J. 2003. Comparability of a paper-based language test and a computer-based language test. *Language Testing*, 20 (3), 295-320.

[11] Choi, S. W., & Tinkler, T. 2002. Evaluating comparability of paper-and-pencil and computer based assessment in a K-12 setting. New Orleans, LA: Paper presented at the annual meeting of the *National Council on Measurement in Education.*

[12] Chu, H. C., Hwang, G. J., Tsai, C. C., & Tseng, Judy C. R. 2010. A two-tier test approach to developing location-aware mobile learning system for natural science course. *Computers & Education*, 55, 1618-1627.

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

[14] Coniam, D., 2006. Evaluating computer-based and paper-based versions of an English-language listening test. *ReCALL*, 18, 193-211.

[15] Economides, A.A. 2006. Adaptive mobile learning, In Proceedings *IEEE WMUTE - 4th International Workshop on Wireless, Mobile and Ubiquitous Technologies in Education*, pp. 26-28, Athens, Greece, November.

[16] Economides, A. A. 2008. Requirements of Mobile Learning applications, *International Journal of Innovation and Learning*, 5(5), 457-479(23).

[17] Economides, A. A. 2009. Adaptive context-aware pervasive and ubiquitous learning. *International Journal of Technology Enhanced Learning,* 1(3),169-192.

[18] Economides, A. A. and Grousopoulou, A. 2010. Mobiles in education: Students' usage, preferences and desires. *International Journal of Mobile Learning and Organisation*, 4(3), 235-252.

[19] Gretes, J.A. and Green. M. 2000. Improving Undergraduate Learning with Computer-Assisted Assessment. *Journal of Research on Computing in Education*, 33(1), 46-54.

[20] Huang , Y.M., Lin , Y.T., and Cheng, S.C. 2009.An adaptive testing system for supporting versatile adaptive testing system for supporting versatile educational assessment, *Computers & Education,* 52, 53-6.

[21] Hwang, G.J., and Chang. H., F. 2011. A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students, *Computers & Education*, 56, 1023–1031.

[22] Hwang, G.J., Wu, P.H., Zhuang, Y.Y., and Huang , Y.M. 2011. Effects of the inquiry-based mobile learning model on the cognitive load and learning achievement of students. *Interactive Learning Environments*, 1–17.

[23] Karadeniz, S. 2009. The impacts of paper, web and mobile based assessment on students' achievement and perception, *Scientific Research and Essays*, 4 (10), 984–991.

[24] Keng, L., McClarty, K. L., and Davis, L. L. 2006. Item-level comparative analysis of online and paper administrations of the Texas Assessment of Knowledge and Skills. Paper presented at the *Annual Meeting of the National Council on Measurement in Education*, San Francisco, CA.

[25] Kim, D.H., and Huynh, H. 2007. Comparability of Computer and Paper-and-Pencil Versions of Algebra and Biology Assessments. *Journal of Technology, Learning, and Assessment,* 6(4).

[26] Kveton, P., Jelinek, M., Voboril, D., and Klimusova, H. 2007. Computer-based tests: the impact of test design and problem of equivalency, *Computers in Human Behavior* , 23, 32–51.

[27] Leeson, H.V. 2006. The Mode Effect: A Literature Review of Human and Technological Issues in Computerized Testing, *International Journal of Testing*, 6(1), 1–24.

[28] Macedo-Rouet, M., Ney, M., Charles, S., and Lallich-Boidin, G. 2009. Students' performance and satisfaction with Web vs. paper-based practice quizzes and lecture notes, *Computers & Education* 53, 375–384.

[29] McFadden, A.C., Marsh II G.E., and Price B.J. 2001. *Computer Testing in Education, Computers in the Schools*, 18:2-3, 43-60.

[30] Mason, B.J., Patry, M., and 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.

[31] Masri, A.A. 2012. Using Mobile Phone For Assessing University Students in English Literature in Jordan. Paper presented in *2012 Orlando International Academic Conference*, Orlando, FL, USA.

[32] Mayotte, S. 2010. Online assessment of problem solving skills. *Computers in Human Behavior*, 26, 1253–1258.

[33] Mazzeo, J., and Harvey, A. I. 1988.The Equivalence of Scores from Automated and Conventional Educational and Psychological Tests. *College Board Report No. 88-8*, NY: College Entrance Examination Board.

[34] Muhanna, W. 2011. The Impact of Using Cell Phone Assessment on Jordanian University Students' Achievement in National Education, *European Journal of Social Sciences* , 20(1).

[35] Noyes, J. M. , and Garland, J. K. 2008. Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, 51(9), 1352–1375.

[36] Paek, P. 2005. Recent trends in comparability studies (*PEM Research Report 05-05*). Available from http://www.pearsonedmeasurement.com/downloads/research/RR_05_05.pdf.

[37] Pommerich, M. 2004. Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *Journal of Technology, Learning, and Assessment*, 2(6).

[38] Pomplun, M., Frey, S., and 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.

[39] Romero, C., Ventura, S., and De Bra, P. 2009. Using Mobile and Web-Based Computerized Tests to Evaluate University Students, *Computer Applications in Engineering Education*, 17(4), 435–447.

[40] Russell, M. and 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.

[41] Russell, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives*, 7, 20. Available at *http://epaa.asu.edu/epaa/v7n20.*

[42] Schroeders, U. and Wilhelm, O. 2010.Testing Reasoning Ability with Handheld Computers, Notebooks, and Paper and Pencil , *European Journal of Psychological Assessment*, 26(4) .

[43] Segall, N., Toni, L., Doolen, J., and Porter, D. 2005. A usability comparison of PDA-based quizzes and paper-and-pencil quizzes. *Computers & Education*, 45(4),417–432.

[44] Shih, J.-L., Chuang, C.-W., and Hwang, G.-J. 2010. An Inquiry-based Mobile Learning Approach to Enhancing Social Science Learning Effectiveness. *Educational Technology & Society*, 13 (4), 50–62.

[45] Struyven K., Dochy, F., and Janssens, S. 2005. Students' perceptions about evaluation and assessment in higher education: a review, *Assessment & Evaluation in Higher Education*, 30(4), 331–347.

[46] Treadwell, 2006. The usability of personal digital assistants (PDAs) for assessment of practical performance. *Medical Education*, 40 (9), 855–861.

[47] Triantafillou, E., Georgiadou, E., and Economides, A. A. 2008. CAT-MD: Computerized adaptive testing on mobile devices. *International Journal of Web-Based Learning and Teaching* Technologies (extended versions of the best papers presented at m-ICTE2006 Conference), 3(1), 13-20.

[48] Wang, S., Jiao, H., Young, M., Brooks, T., and Olson, J. 2008. Comparability of Computer-Based and Paper-and-Pencil Testing in K–12 Reading Assessments. *Educational and Psychological Measurement*, 68(1), 5-24.

[49] Way, W. D., Davis, L. L., and Fitzpatrick, S. 2006. Score comparability of online and paper administrations of Texas assessment of knowledge and skills. Paper presented at the *Annual Meeting of the National Council on Measurement in Education*. San Francisco, CA.

[50] Wilson, K., Boyd, C., Chen, L., and Jamal, S. 2011. Improving Student Performance in a First-Year Geography Course: Examining the Importance of Computer-Assisted Formative Assessment, *Computers & Education*, 57 (2), 1493-1500.

[51] Wu , J., and Zhang, Y. 2010. Examining potentialities of handheld technology in students' academic attainments *Educational Media International*, 47(1), 57–67.