# USING ELECTRONIC TESTS VERSUS PEN AND PAPER TESTS: THE EXPERIENCE OF DELTA UNIVERSITY

*By*

**Abdelrahman E. AlAdl**

*Associate Professor of TEFL*
*Delta University for Science and Technology*
*Egypt- Gamasa, International Road*

# Using Electronic Tests versus Pen and Paper Tests: The Experience of Delta University

**Abdelrahman E. AlAdl**
*Associate Professor of TEFL*
*Delta University for Science and Technology*
*Egypt- Gamasa, International Road*

## Abstract

*The aim of the current study is to investigate the impact of utilizing e-tests on Students' performance and attitude towards the online testing versus the traditional approach (pen and paper tests). 60 students from Delta University for Science and Technology, were participated in the study, while they were studying English 1 course in their preparatory year. They were divided randomly into two groups, (Experimental and control groups). The control group was tested by using pen and paper based exam, while the experimental group is assessed by using e test. The Scores of two groups were compared on achievement in English final test. Students' attitude towards the experience was measured by administering a questionnaire after finishing the tests. The main hypothesis of the current study was that there are differences between the scores of two groups on the post achievement tests and attitude scales in favor of the experimental group. The findings revealed that there were significant differences between the groups on the post achievement tests and attitude scales in favor of the experimental group. Moreover, e test version was more reliable in terms of internal and external validities and it decreased testing time and raised testing motivation of the students.*
*Key words: Assessment, e tests, attitude, performance*

## 1. Introduction

There are already different examination methods which are used in higher education institutions to appraise the academic progress of students. For example, paper-pencil-based examinations, assignments, presentations, and etc.,. Holifield, & Brown (2004) identified several techniques that are used within different institutions for measurement purposes; the most frequently performed are examinations (formative or summative). The speedy improvement of Information and Communication Technologies in teaching and assessment has transferred the archetype from paper-pen - based to computer-based system of examinations which are often termed as Computer Based Testing (CBT). (Uysal & Kuzu, 2009)

Using technology in education has become prevalent in the last years. There has been a growing interest in improving and utilizing CBT in

educational systems instead of traditional paper and pencil exams. There are several advantages of applying computer-based tests (CBT) over traditional paper-based tests (PBT), in terms of computer support for question development, reduced cost of test distribution and administration, reduced cost of distributing answers to graders, and possible automated support for grading. Additionally, new types of test items (e.g., based on audio, video, 3D engineering models, industrial tools, and interaction) can be enabled, thus giving a test with much higher validity and professional practices. (Kim, 2005)

CBT is the form of assessment in which the computer is an integral part of question papers' delivery, response storage, marking of response or reporting of results from a test (Whittington, Bull & Danson, 2000). Conole and Warburton (2005) defined it as 'the use of computers for evaluating students' learning'. Due to the inclusion of informational technology in education, it is necessary to re-consider and rethink, adapt or modify the conventional examination techniques. Electronic assessment tools had limited the load of teachers and assist to perform examinations decisively. It can be used to create more effectual learning by testing a series of knowledge, understanding and skills (Bull, 1999). Moreover, Bodmann and Robinson, (2004) reported that CBT present powerful methods for meeting the new challenges of designing and implementing assessments that go beyond the traditional practices and help in recording a wide range of cognitive knowledge and skills.

The findings Several research revealed that the majority of teachers have recently adapted e-tests rather than conventional methods of assessing students (Betlej, 2013). They attempt to make use of the designed mode of questions and the quick scoring technique. In the other hand, Some teachers were opposed to implement e-tests, as they were inclined to modify their assessment habits (Betlej, 2013). Introducing any new technology should engage all stakeholders in the arrangement process (Wibowo et al., 2016). So, the current study attempts to verify the best method of assessment for Delta University. Moreover, to provide teachers and students with both pedagogical and technical guidance for successive implementation of e tests.

## 2. Review of Literature

There were various studies that have been conducted using e tests to enhance students' achievement in assessment purposes. This topic has been treated in terms of activities, the design of e tests, advantages, and the attitudes of teachers and students.

For example, Calarina and Wallace (2002) attempted to verify numerous key factors in CBT versus PPT estimation. They were; computer acquaintance, content, gender and competitiveness. Results revealed that CBT delivery impacted positively on students' scores as compared to BBT. They showed also that Gender, competitiveness, and computer familiarity were not related to this performance variation. Moreover, Bodmann and Robinson (2004) carried out a study to compare performances and speed variations among (CBT) and (PBT). Procedures altered for the second tests, with the first group receive PPTs and second group CBTs with a period of two weeks. It was concluded that students accomplished the CBT faster than PBT with no difference in scores. On the other hand, Chua & Don (2013) carried out a study to investigate the effects of (CBT) on test performance and the attitudes of the students. Results indicated that there were significant testing effects on the test scores for the (PBT)and (CBT). The testing effect for (PBT) (d =-.83) was negative and also bigger than for the (CBT) (d = .57). In other words, (PBT) is associated with more serious testing effect problems than (CBT). There was no significant treatment effect of (CBT) on the test scores, and the effect of it on the test scores was actually due to the testing effect.

Chua (2012) conducted a study aimed at investigating the effectiveness of (CBT) on test performance, testing time and testing attitude by comparing it with (PBT). The results showed that the CBT technique is more reliable in terms of internal and external validity and it reduced testing time significantly. Also, it created a more self-efficacy, and positive attitude towards the testing process. Consequently, the purpose of Alzu'bi's (2015) study was to investigate how using e-tests affect on the Students' achievement and motivation in an English 101 Course. Two groups (experimental and control) of students, were used in this study. The control group was assessed by using pen and paper based exam and experimental group was assessed by using e-exam. The control group consisted of 58 students while the experimental group consisted of 58 students. The groups were compared on achievement of E101 and motivation toward learning English by e-exam. The study included 10 units. Scores on both achievement tests were collected at the final exam. The results revealed significant variation between the groups on the post achievement tests and motivation scales in favor of the control group.

The key findings of Wibowo's (2016) study showed that students and teachers in Australian universities are hopeful about the adoption of e-exams if their system is satisfactorily developed. They are attentive enough

regarding the benefits that could offered by technology in supporting learning and assessment. Moreover, considering e-test as an innovation in the assessment process in higher education.

At last, the study of Pizarro and Laborda (2017) examined test-takers' views on a computer-delivered speaking test in order to investigate the aspects they consider most relevant in technology-based oral assessment, and to explore the main advantages and disadvantages computer-based tests may offer as compared to face-to-face speaking tests. A small-scale open questionnaire was administered to 80 test-takers who took the Aptis speaking test at the Universidad de Alcalá in April 2016. Results revealed that examinees believe computer-based tests provide a valid measure of oral competence in English and are considered to be an adequate method for the assessment of speaking. Interestingly, the data suggest that the personal characteristics of test-takers seem to play a key role in deciding upon the most suitable and reliable delivery mode.

## 3. Context of the Problem

Regarding the scientific challenges and the information revolution of the 21st century, there is an extensive trend to use computers in all aspects of life. E-tests have created evident effects in the development of the educational process and have been gradually administered in the higher education stage worldwide (Wibowo et al., 2016). Consequently, e-tests are presently adopted by numerous private and public educational institutions in Egypt that they are effectual means for (a) providing instant feedback to students and (b) minimizing load on over-stretched academic staff (Jamila et al., 2012). E-tests have been found to be efficient in reducing the large amount of the workload on examination systems thus minimizing the load of teachers.

So, there has been an increasing interest for utilizing e-tests in Delta University colleges, Accordingly, the current study is developing an English computerized-test for English1 course to replace paper-based tests. An attempt also to confirm that e-test process is vital to support the university's assessment methods and to finally replace conventional paper-based tests. By applying e tests, the study could gain an understanding of a currently available e-tests system, moreover, exploring opportunities and challenges that occurred by implementing this system. Finally, evaluating students' attitude towards e-tests .

## 4. Research Questions

This study attempts to answer the following questions:

1. Are there any statistically significant differences at (0.05) between the achievement scores of students due to estimation strategy (paper and pencil, and e- test)?
2. What is the effect of e-test on students' attitudes towards it as a method of assessment?

## 5. Research Hypotheses

Based on the context of the problem, the following hypotheses could be stated:

1. There is a significant difference at (0.05) between the achievement scores of students who take the exams by using e- test versus the traditional paper and pencil form in English1 course.
2. There is a statistical significant difference in the students' attitude towards e tests.

## 6. Significance of the Study

The current study may contribute in introducing new ways of evaluation like e-tests for English courses to meet the new century demands and to explore the problems of using the traditional ways of evaluation. In addition, it may also explore the challenges and the opportunities of applying such technological method for examination, to meet the intensive demands of implementing e tests in all Egyptian institutions.

## 7. Methodology

### A) Participants

The population of the study was selected from students of first year in Delta University during the Fall semester of the academic year 2018/2019 and who are studying different specializations. The sample of the study consisted of 90 students who study English1 course, who were selected purposefully and distributed in two groups 45 students were evaluated using electronic assessment, and 45 students were evaluated using paper-pencil assessment).

### B) Procedures

This study was conducted to follow the equivalent pre /post-test two-group mode. The experiment included the participants in the experimental group who are exposed to the electronic test, while the control group are exposed to the paper and pencil-based assessment for the same period. A pre-test was given before the treatment of the estimation methods to all groups to ensure that they were equivalent and the same test was administered as a post-test after applying the evaluation methods to estimate which approach had more influence on the subjects than the other.

*C) Instruments*

1- English1 test

The aim is to compare the achievements of the two groups on the pre and posttests. The researcher designed a test based on the instructional material of English1course. The test contained 50 multiple choice items. The researcher developed the test in two ways; firstly by using paper-based assessment, secondly by using computerized the same test). The students in all groups are exposed to a pre-test (by using paper-based test) to verify their actual level before starting the experiment. After teaching both groups the content of the course, the same test was administered as a post-test at the end of the application to determine students' achievement (by using e-test with the experimental group and paper-based assessment with the control one). The time allocated between the pre-test and the post-test was 16 weeks (one semester); to minimize the effect of the pre-test on the findings of the implementation. (see Appendix)

To ensure face validity, the chosen test has been submitted to a jury of experts in the fields of TEFL. A pilot administration of the test has been carried out to determine the time allotted for responding to the test. It has been given to (20) freshmen students of Delta University during the period of the experiment. It is found that the time allotted for answering the test by students is 90 minutes.

To get the internal consistency of the test, Alpha Cronbach Formula has been used. The reliability coefficient is found to be 0.82. The subjects' performance was evaluated by two scorers for both pre and post tests. To ensure the reliability of the rating process, inter scorer reliability was calculated. It was 0.81 which appeared as an acceptable value of inter-rater reliability.

**2- Attitude questionnaire**

It was administered after the experiment. The questionnaire contains 12 questions, and it measures students' attitudes towards e- tests as an approach of assessment. A 5-point Likert scale is performed in this questionnaire and the students were asked to complete it. In order to examine the reliability of the questionnaire Cronbach's coefficient alpha was calculated. The results revealed that the reliability value was 0.75. which is acceptable for a measure with twelve statements. (Appendix B)

**8. Data Analysis and Results**

To answer the two questions of the study, means and standard deviation were used on pre and post tests of students' achievement of the experimental and control groups. Also, the method of (One- Way ANOVA)

analysis of variance was used to make a comparison between the control and the experimental groups.

To verify the equivalence of the groups, the mean scores and standard deviations were computed in the pre-test following the variable of the group (experimental and control). To show the statistical variations between the mean scores, the T test statistical procedure was computed as shown in Table 1.

**Table 1.**
**Means, standard deviation and t test of the Pre-test**

|  | Group | Mean | S. Deviation | T. Value | f | Sig. (2- tailed) |
|---|---|---|---|---|---|---|
| Pre test | Experimental | 29.23 | 9.293 | .021 | 14 | .983 |
|  | Control | 29.19 | 9.408 |  |  |  |

Table1 showed that the experimental group's mean score on pre-test was approximately similar to the control group's mean score on the same test (29.23; 29.19) correspondingly. T value was .021., which was not statistically significant at (.05). Therefore, this finding reveals that both groups are equivalent.

**Results regarding the first study question**

The first question was: "Are there any statistically significant differences at (0.05) between the achievement scores of students due to estimation strategy (paper and pencil, and e- test"? For answering this question, standard deviations, and adjusted means for the students' achievement were calculated following the variables of the study.

**Table 2.**
**Means and standard deviation of students' achievement**
**on the post-test.**

| Group | Means | S. Deviation | Adjusted mean | No. |
|---|---|---|---|---|
| Experimental | 31.27 | 9.624 | 31.283 | 45 |
| Control | 34.68 | 8.362 | 34.703 | 45 |
| Total | 32.97 | 9.137 | 32.98 | 90 |

Table 2 indicates an evident variation in the means, standard deviations and the adjusted mean of the students' achievements (experimental and control groups). To find out if there are statistically significant differences between the two groups' scores on the post-test, comparing with the pre- test scores, (One Way ANOVA) analysis was performed, as presented in table 3.

**Table 3.**
**The results of the performance of the two groups in the**
**pre and post tests**

|  | Group | N | Mean | SD | t-value | Df | Sig.(2-tailed) |
|---|---|---|---|---|---|---|---|
| Pretest | Experimental | 20 | 22.00 | 7.655 | .019 | 38 | .987 |
|  | Control | 20 | 22.05 | 9.876 |  | 38 |  |
| Posttest | Experimental | 20 | 38.75 | 6.056 | 5.103 | 38 | .000 |
|  | Control | 20 | 26.25 | 0.243 |  | 38 |  |

Table3 shows that there was a statistical significant difference between the two groups of the study at level .05 in the posttest for the sake of the experimental groups.

**Results related to the second question**

The second question is: What is the effect of e-test on students' attitudes towards it as a method of assessment? A questionnaire with Likert-scale was used to t test and to better understand the students' attitude towards LEP in learning English. The statistical analysis was used to determine the means and standard deviations for post data of students' motivation within the experimental group, as shown in table (4).

**Table 4. Students' Attitude Towards e-test**

| N. | Statement | M | SD | T | Sig. (2tailed) |
|---|---|---|---|---|---|
| 1 | I like using computer to answer E101 questions | 4.400 | .507 | 33.606 | 000 |
| 2 | I enjoy answering the English101 e-test | 4.333 | .617 | 27.191 | 000 |
| 3 | The time allotted for the test is sufficient | 4.266 | .703 | 23.482 | 000 |
| 4 | Answering English test is more significant for me than the score I obtain | 4.333 | .4787 | 34.395 | 000 |
| 5 | Answering English test gives me a sense of achievement | 4.466 | .639 | 27.033 | 000 |
| 6 | English test can guide me to obtain a good score | 4.533 | .516 | 34.000 | 000 |
| 7 | Using e-exam guides me to achieve my goals | 4.466 | .516 | 33.500 | 000 |
| 8 | Answering English101 e- test is not difficult | 4.666 | .617 | 29.283 | 000 |
| 9 | I put adequate effort into answering English test | 4.600 | .507 | 35.133 | 000 |
| 10 | I can learn better from taking an e-test | 4.266 | .457 | 36.101 | 000 |
| 11 | I expect to get better grades than other students in the English e-test | 4.466 | .516 | 33.500 | 000 |
| 12 | I can master the knowledge and skills by using an English e-test | 4.266 | .703 | 23.482 | 000 |

There were five choice ranging from strongly agree (=5), No idea (=3) to strongly disagree (=1) in each statement. * z, p> 0.05 (not significant,), p<0.05 (almost significant), p<0.01 (significant), p<0.001 (very significant).

Table4 reveals that there was an effect on students' attitudes towards using e-test in the assessment of English1 Course at level .05. This statistical difference is in favor of the experimental group. This result of attitude surveying revealed that students' enthusiasm for learning, increased by applying e-test approach. The attitude of EG was seen to change positively towards learning English by applying creative approaches for assessment.

## 9. Discussion

The results revealed that there was a significant difference at (0. 05) between the mean scores of the experimental group and that of the control group on the post-test in favor of the experimental group. In other words, the e-test affects positively on students' achievement rather than paper and pencil test. These findings agree with (Bodmann & Robinson,2004; Conole & Warburton, 2005; Chua (2012) and Chua & Don, 2013) studies. There are several reasons of these positive advantages of implementing e-tests as an approach of assessment, for example; the varieties of the interfaces, such as the use of graphics allows a dynamic presentation of the test content. Further, CBT allows other measures relating to cognitive and perceptual performance. In addition, it allows a more various sample to be located (Carlbring et al. 2007) because people only need access to a computer. It also lets people take part in testing from their homes; this group may not inevitably be available for testing in a laboratory setting due to mobility and, perhaps, disability issues. (3) Standardization of test atmosphere, that is, the test is administered in the same format and in the same way for an assigned time. Finally, Online scoring, which means presenting results in faster and more accurate way, in addition to reducing the human error. Information relating to test-taking behaviors, for example, how much time was spent on each item, can be readily collected (Liu et al. 2001). It is generally accepted that delivery and scoring of the test online leads to economic cost savings, especially for large samples.

In the contrary, the results of the current study disagreed with some studies. For example, Zandvliet and Farragher (1997) which found that a computer administered version of a test required more time than the written test. Further, there may be differences in the layout of the test depending on a respondent's particular browser software and settings. Moreover, the study

of Alzu'bi (2015) which claimed that there was no differences in achievement between the CBT and PBT groups.

Regarding students' attitudes towards using e-test, the means and standard deviation of students' scores showed significant variation. Thus, the students' attitudes towards e-test changed positively. They reported that the degree of test anxiety was minimized. Moreover, they showed their readiness to practice computerized test as they are familiar with technological platforms. In addition to, their passion to practice other types of tests which are similar to the international tests such as TOEFL, IELTS and TOEC. These results are supported by (Chua & Don; 2013; Wibowo, 2016 and Pizarro & Laborda, 2017)

Students, however, accounted certain challenges towards e-tests. Most were related to the usability issues. So, it was important to inquire them in the preparing stage if they have sufficient previous skills in using e-tests (Ryan et al., 2000). In conclusion, students showed positive attitudes toward the implementation of e-tests if the arrangements were considerately designed. They also are aware of the importance of technology for their future career. Recognizing students' attitude is an essential factor toward an integration of technology in the examination modes. It guides teachers in establishing future e-tests and it also offers a valuable construct in expecting students learning outcomes (Masrek, Abdul Aziz, & Johare, 2012; and Mehra & Omidian, 2011).

## 10. Pedagogical Implications and Conclusion

The current study presents the analysis and comparison of the results of conventional PBT and CBT modes of examination. The aim of the study was to find out if it is possible to replace conventional tests that students solve on papers by e tests. Based on the result of the study, it is concluded that CBT can substitute traditional PBT for student' knowledge estimation, but more attention should be paid to its composition. In the phase of questions designing, the existing question bank needs to be modified to the online testing design. The certain number of existing Essay type questions could be altered to the questions of objective mode (e.g. True or false, Connection or Multiple-Choice types). Regarding the test takers and teachers, it is crucial that they have sufficient training on e-tests. Furthermore, the inadequately designed e-tests may lead to negative thought and discourage teachers and students from using e-tests.

Students' responses were investigated to determine too easy questions (the questions with 100% correct answers), as well as questions for which there are no correct answers (the questions that are imprecisely designed and

therefore inconceivable to the students). It is also arranged to update the question bank with sets of several categories of questions for formative measurement. Students are able to solve a self-assessment test after completing each course unit. This will enable them to verify their acquaintance, in the process of preparing for the final tests. Both students and teachers reported several feedback and concerns, which could be used as valuable factors of reference for future applying of e-tests at Delta University. The incompetence of the presented system and other features that are related to the system could be enhanced. Future selection of e-tests mode should consider items missing in the current system. The application of e-tests is still new at Delta University, so, more work need to be done which requires a partnership among students, teaching, professional and technical staff, and decision makers at the institution.

## 1. Recommendations

Based on the discussed findings, the researcher recommended the following suggestions:

- The test developer should develop the computerized test than entering test items  from a traditional paper and pencil test into a computer.
- The university should solve  the  technical problems during applying e- examination and prepare enough modern labs.
- The use of software programs in language teaching should be explored further. Researchers should carry out further studies on the effectiveness of computer based method of teaching language skills and other components of the language.
- Teachers are recommended to vary their approaches and techniques of assessment and teaching, according to their students' needs and interests. They are also recommended to use the technological methods more intensively and more frequently to meet the recent demands and match students' interests in these fields.

## Reference

- Alzu'bi M. A. (2015) The Effect of Using Electronic Exams on Students' Achievement and Test Takers' Motivation in an English 101 Course. *Conference of the International Journal of Arts & Sciences,* Vol. 08(03): Pp 207–215.
- Betlej, P. (2013). E-examinations from student's perspective – The future of knowledge evaluation. *Cognition and Creativity Support Systems,* Vol. 9, Pp.12–21.

- Bodmann, S. M. and Robinson, D. H. (2004). Speed and Performance Differences among Computer-Based and Paper-Pencil Tests. *Journal of Educational Computing Research*, Vol. 31(1), Pp.51 – 60.
- Bull, J. (1999). Computer – Assisted Assessment: Impact on Higher Education Institutions. Educational Technology & Society, 2(3). Retrieved December 05, 2004 from http://ifets.ieee.org/periodicals.
- Carlbring, P., et al., (2007). Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Computers in Human Behavior*, Vol. 23, Pp.1421–1434.
- Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, Vol.33 (5), Pp. 593-602.
- Chua, P. & Don, Z. (2013). Effects of computer-based educational achievement test on test performance and test takers' motivation. *Computers in Human Behavior*, 29 Pp. (1889–1895).
- Chua (2012). Effects of computer-based testing on test performance and testing motivation. *Computers in Human Behavior*, Vol. 28, N.5, Pp (3-20).
- Conole, G., & Warburton, B. (2005). A Review of Computer Assisted Assessment. ALT-J, *Research in Learning Technology*, 13(1), 17-31.
- Jamila, M. et al. (2012). Computer-based vs paper-based examination: Perceptions of university teachers. *The Turkish Online Journal of Educational Technology*, 11, 371–381.
- Kim, C. (2005). Work in progress– Mobile computer based classroom assessment. Proceedings of 35th ASEE/IEEE Frontiers in Education Conference, October 19 – 22, Indianapolis.
- Liu, M., Papathanasiou, E., and Hao, Y.-W., 2001. Exploring the use of multimedia examination formats in undergraduate teaching: Results from the fielding testing. *Computers in Human Behavior*, Vol.17, Pp. 225–248.
- Masrek, M. N., Abdul Aziz, N. S., & Johare, R. (2012). The relationship between Internet attitude and Internet addition. *Asian Journal of Information Technology*, 11, 125–130.
- Mehra, V., & Omidian, F. (2011). Examining students' attitudes towards e-learning: A case from India. *Malaysian Journal of Educational Technology*, 11, 13–18.
- Pizarro, M. A. & Laborda J. G (2017) Analysing Test-Takers' Views on a Computer-Based Speaking Test. *Issues Teach.* Vol. 19, Pp. 23-38.

- Ryan, S., Scott, B., Freeman, F., & Patel, D. (2000). *The virtual university, the Internet and resource-based learning*. London, England: Kogan Page.
- Uysal, O. and Kuzu, A. (2009). A Thesis Proposal: Quality Standards of Online Higher Education in Turkey. Internationalization and the Role of University Networks. Proceedings of the 2009 EMUNI Conference on Higher Education and Research, Potorož, Slovenia, 25 – 26 September.
- Whittington, D., Bull, J., & Danson M. (2000). Web-Based Assessment: Two UK Initiatives. The Sixth Australian World Wide Web Conference, Rihga Colonial Club Resort, Cairns, 12-17 June 2000, Australia. Retrieved October 12, 2010 from http://www.ausweb.scu.edu.au/aw2k/papers/index.html.
- Wibowo S., et al. (2016) A Pilot Study of an Electronic Exam System at an Australian University, *Journal of Educational Technology Systems*, Vol. 45(1) 5–33.
- Zandvliet, D. and Farragher, P., 1997. A comparison of computer-administered and written tests. *Journal of Research on Computing in Education*, Vol. 29, pp. 423–438.