International Journal of Education & Literacy Studies
ISSN 2202-9478
Vol. 2 No. 4; October 2014
Australian International Academic Centre, Australia

Comparing Between Computer based Tests and Paper-and-Pencil based Tests

Marzieh Ghaderi
Department of English Language and Literature , Larestan Branch
Islamic Azad University, Larestan, Iran
E-mail: gh_mrzh@yahoo.com

Marzieh Mogholi
Department of English Language and Literature, Larestan
Branch, Islamic Azad University, Larestan, Iran
E-mail:mmoghali@yahoo.com

Afshin Soori (Corresponding author)
Department of English Language and Literature , Larestan Branch
Islamic Azad University, Larestan, Iran
E-mail: Afshin_soori@yahoo.com

Received: 25-07- 2014 Accepted: 19-09-2014 Published: 30-10-2014
doi:10.7575/aiac.ijels.v.2n.4p.36 URL: http://dx.doi.org/10.7575/aiac.ijels.v.2n.4p.36

Abstract

Testing subject has many subsets and connections. One important issue is how to assess or measure students or learners. What would be our tools, what would be our style, what would be our goal and so on. So in this paper the author attended to the style of testing in school and other educational settings. Since the purposes of educational system have been changed, there have been some inevitable changes in testing style. Therefore, in recent years, computer-based testing (CBT) has grown in popularity and will likely become the primary mode for delivering tests in the future. So different aspects of computer-based tests and paper-and-pencil based tests like motivation, anxiety and students' performance with different styles will be reviewed in this paper.

Keywords: paper and pencil (P&P) testing, computer-based testing (CBT), Motivation, Anxiety, Educational System

1. Introduction

Dillon published his critical review of the empirical literature on reading from paper vs. screen in 1992 and investigated the equivalence of computer- and paper. Many test experts are exploring the possibility of transitioning from a paper and pencil mode of test administration to a computer based test (CBT) mode.

Administration as an important factor in testing should be attended to in changing the mode of it, especially in creating questions as to whether or not the tests, and test scores derived from those tests, are equivalent, and whether there are effects of the testing mode.

Comparative studies are conducted to determine if such mode effects exist, and if scores derived from the same test in two different modes are comparable and can be used interchangeably.

In recent years, computer-based testing (CBT) has grown in popularity and will likely become the primary mode for delivering tests in the future. Computers revolutionized the world of training and development. Many investigators such as Fuhrer (1973) began researching on many points of mode which has enhanced training through computers. Many studies focused on the effects of using computers in the classroom for testing on various aspects of the learning environment such as student anxiety, teacher attitudes, student achievement and more. The advantages of computers are well known and apparent (Al-Amri, 2009). It seems that the integration of computers has increased the productivity and innovation in the area of testing. The standardization of test administration conditions is one of the benefits offered by Computer- Based Testing (CBT). No matter what the tests' population size is, CBT helps test developers to set the same test conditions for all participants. Al-Amri (2009) also believed that CBT improves all aspects of test security by storing questions and responses in encrypted databases and enables testers to create randomized questions and answers from vast question pools. Internet based assessment is one type of CBT, and the researchers in the field of psychology are doing the personality tests on internet. There are some studies on internet versus paper and pencil formats in psychological tests (Buchanan, 2000, 2001, 2002; Buchanan et al., 2005; Buchanan & Smith, 1999) Low cost and access to a greater number of subjects in much less time are some of the advantages of the on-line-administration. In addition, data entry requirements are largely eliminated, and the potential for errors in data entry is largely reduced

(Barak & English, 2002; Buchanan, 2006). Most of the authors (Chuah, Drasgow, & Roberts, 2006; Davis, 1999; Gosling, Vazire, Srivastava, & John, 2004) are positive about the relationship between the equivalence and similarity properties of the traditional paper and on-line formats of personality questionnaires.

There are some studie (e.g. Buchanan et al., 2005; Buchanan & Smith, 1999) on the effect of computer-administered testing on the learner. Many others, however, are focused on the differences between computerized tests and traditional paper and pencil tests without considering the effects on the learners adequately. Some studies are inconclusive. The 1988 statewide field test compared reading, writing, and computation scores for approximately 1,000 students.

Many studies (Chuah, Drasgow, & Roberts, 2006; Gosling, Vazire, Srivastava, & John, 2004) found significant differences between computer-administered testing and traditional paper and pencil testing. These studies and articles attributed achievement differences to several factors. Russell and Haney (1996) found significant differences in the performance of students on the National Assessment of Educational Progress computerized tests when compared to traditional paper and pencil tests. They compared 42 students tested on a computer-administered test with scores of 47 students tested on a traditional paper and pencil test. In addition to answering multiple-choice items, there were open-ended items requiring original responses. For scoring, raters only saw the computer products because all hand-written responses were entered into the computer verbatim after the test concluded. Larger mode effects were found on open-ended writing tasks than on multiple-choice tests. Additionally, analysis showed that students who wrote on the computer tended to organize their work into paragraphs and wrote responses nearly twice as long as the students who hand wrote their responses.

Karadeniz (2009) studied the impact of paper based, web based and mobile based assessment on students' achievement. A group of 38 students was experimented for 3 weeks. Significant differences were found between the scores achieved by the students in second week, but not in the first week. The authors perceived that students had positive attitude towards web based, mobile-based assessment due to ease of use, comprehensive, and instant feedback. Moreover, most favored tests were web-based and the least favored were paper-based.

In another experimental research, Bodmann and Robinson (2004) conducted an experimental study to compare speed and performances differences between computer-based (CBTs) and paper-pencil tests (PPTs). Both CBTs and PPTs contained 30 MCQs items with 35 minute of time limit. Approximately half the class (i.e. 28 students) took the first test on the computer and the rest preferred first test on paper. Procedures shifted for the second tests, with the first group receiving PPTs and second group received CBTs after two weeks. It was concluded that undergraduates completed the CBT faster than PBT with no difference in scores.

In fact, many research works have been conducted to evaluate the comparability of computer-based assessment and paper and pencil based assessment. Some studies revealed that there is a significant difference between the two testing modes on test scores (e.g. Scheuermann & Björnsson, 2009; Choi, Kim, & Boo, 2003), while other studies reported opposite or inconsistent results (e.g. Al-Amri, 2009; Boo, 1997). However, unlike the abundance of CBA research done with older or special needs students, there is a dearth of available research focusing on the issues of computer-based assessment with typically developing young children (Barnes, 2010).

2. Anxiety and Motivation comparing between CBT and P&PBT

Some studies (Buchanan et al., 2005; Gosling, Vazire, Srivastava, & John, 2004) examined the anxiety level of the test taker and the time required to complete the examination. Wise (1997) contemplated test administration from the perspective of the examinees during a computerized adaptive test and focused on issues surrounding the development of computerized-adaptive tests. The author examined how the ability to review answers affected examinees' anxiety and performance levels. Another issue interwoven with that of item review was the time limits on the test takers. He believed that because a computerized-adaptive test was usually shorter than a traditional paper and pencil test, computerized-adaptive test developers should be either be extremely liberal when establishing time limits or impose no time limits at all. Increased test anxiety, higher or lower examinee motivation, and equity in computer experience, he wrote, all have implications when considering inferences made from computerized-adaptive test scores and should be carefully weighed when developing computerized-adaptive tests.

Test anxiety was not the only emotion reported to affect test takers. Motivation levels also affected achievement. Feedback and knowledge of results were indicated as significant in several studies. Betz and Weiss (1976) tested 350 college students divided into high- and low-ability groups. Their reported motivation levels were found to be related to their ability level. Motivational differences were attributed to having immediate knowledge of test results, which resulted in greater standardization of the test-taking environment in a computer-adaptive test mode than in traditional paper and pencil testing environments.

3. CBT and P&PBT and its effect on Students' performance

The benefits of performance testing are well documented: Ayala, Shavelson, Lin and Shultz (2002) argue that science performance assessments can measure different types of knowledge including declarative, procedural and schematic knowledge. Ruiz-Primo and Shavelson (1996) say that performance tests produce high level of reasoning processes, since these tests are closely related to what students and scientists do in the lab. Haertel (1999) argues that these tests not only show how students learn, but also show the higher engagement of students in learning. Elliot (1995) argues that performance assessments provide evidence of what students know and are able to do. Quellmalz (1999) says that the evidence gathered during the performance provides insights to students' thinking, and at the same time introduces students to authentic real-world problems, which allows them to show how they can apply academic knowledge to

practical situations. For the reasons mentioned above performance assessment tests are alternatives to measure students' knowledge in electrical circuits.

4. Conclusion

Based on the above-mentioned discussions and reviews we can conclude that the difference of testing style in different situation has different conclusions and aspects. As the goals in testing have been changed, the styles should be changed to get better results.

References

Al-amri, S. (2009).*Computer-based testing vs. paper-based testing: establishing the comparability of reading tests through the evolution of a new comparability model in a Saudi EFL context.* Thesis submitted for the degree of Doctor of Philosophy in Linguistics.University of Essex.

Ayala, C. A., Shavelson, R. J., Yin, Y., & Schultz, S. (2002). Reasoning Dimensions Underlying Science Achievement: The case of Performance Assessment. Educational Assessment, 8(2), 101-122.

Barak, A., & English, N. (2002). Prospects and limitations of psychological testing on the Internet. Journal of Technology in Human Services, 19, 65–89.

Betz, N. E., & Weiss, D. J. (1976). Psychological effects of immediate knowledge of results and adaptive ability testing: Research Report 76-4. (Office of Naval Research, Publication Number ED129863). Arlington, VA. Office of Naval Research.

Boo, J. (1997).*Computerized versus paper-and-pencil assessment of educational development: score comparability and examinee preferences.*Unpublished dissertation, University ofIowa.

Buchanan, T. (2000). Potential of the Internet for personality research. In M. H. Birnbaum (Ed.), Psychological experiments on the Internet (pp. 121–265). San Diego, CA: Academic Press.

Buchanan, T. (2001). Online personality assessment. In U. Reips & M. Bosnjak (Eds.), Dimensions of Internet science (pp. 57–74). Lengerich, Germany: Pabst Science.

Buchanan, T. (2002). Online assessment: Desirable or dangerous? Professional Psychology: Research and Practice, 33(2), 148–154.

Buchanan, T. (2006). Personality testing on the Internet: What we know, and what we do not. Oxford Handbook of Internet Psychology.

Buchanan, T., & Smith, J. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. British Journal of Psychology, 90, 125–144.

Choi, I-C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing,* 20, 295-320.

Chuah, S. C., Drasgow, F., & Roberts, B. W. (2006). Personality assessment: Does the medium matter? No. Journal of Research in Personality, 40, 359–376.

Davis, R. N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. Behavior Research Methods, Instruments and Computers, 31(4), 572–577.

Dillon, A., 1992. Reading from paper versus screens: A critical review of the empirical literature. Ergonomics, 35, 1297–1326.

Elliott, Stephen (1995). Creating Meaningful Performance Assessments. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland

Fuhrer, S. (1973). A comparison of a computer-assisted testing procedure and standardized testing as predictors of success in community college technical mathematics (Doctoral dissertation, New York University, 1973). Dissertation Abstracts International, 34 (6), 3086.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. American Psychologist, 59(2), 93–104.

Haertel, Edward (1999).Performance Assessment and Education Reform. PHI DELTA KAPPAN, mayo 1999, 662-666.

Karadeniz, S. (2009). The impacts of paper, web and mobile based assessment on students' achievement and perceptions. *Scientific Research and Essay,* 4(10), 984 – 991. Retrieved May 15, 2011 from www.academicjournals.org

Legg, S. M., & Buhr, D. C. (1990, April). Investigating differences in mean score on adaptive and paper and pencil versions of the college level academic skills reading test. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, MA.

Quellmalz, Edys, Patricia Schank & Thomas Hinojosa & Christine Padilla (1999). Performance assessment links in science. Practical Assessment, Research & Evaluation, 6(10).

Ruiz-Primo, Maria; Shavelson (1996). Rethoric and Reality in Science Performance Assessment: And Update. Journal of Research in Science Teaching, 33(10), 1045-1063.

Russell, M., & Haney, W. (1996). Testing writing on computers: Results of a pilot study to compare student writing test performance via computer or via paper and pencil. Paper presented at the Mid-Atlantic Alliance for Computers and Writing Conference, Chestnut Hill, MA.

Scheuermann, F., & Bjornsson, J. (2009). *The transition to computer-based assessment*. Luxembourg: Office for Official Publications of the European Communities.

Wise, S. L. (1997, March). Examinee issues in CAT. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.