Computer Adaptive Test (CAT): Advantages and Limitations

Marzieh Rezaie 1*, Mohammad Golshan 1

1. Department of English, Maybod Branch, Islamic Azad University, Maybod, Iran.
* Corresponding Author’s Email: Marziehrezaie83@yahoo.com

Abstract – Today, the use of computers and electronic devices has become widespread all around the world. Moreover, in language testing field, technology has been an instrument which led to expansion and innovation in language testing. For example, computer adaptive test (CAT) is a kind of computer-based test which adapts to the ability levels' of the participants. Due to the effects of computer technology in the field of language assessment and testing, this study attempts to investigate different aspects of Computer Adaptive Test (CAT). The results of this study show that the use of CATs in language testing may have its own advantages and disadvantages. Therefore, authorities should be aware of the positive and negative aspects of a computer adaptive test when they want to administer a test in adaptive format.

Keywords: computer, adaptive, test, testing, technology

I. INTRODUCTION

The use of computers and electronic devices has become widespread all around the world; specifically, computers and on-line processes were increasingly used for evaluating the language proficiency of English learners (Fleming & Hiple, 2004). These improvements in computer technologies have affected many parts of educational settings such as learning, testing and assessment (Bennett, 2002; Pommerich, 2004). As a result, students can learn their lessons with the aid of computer (CALL); and computerized tests such as CBE (Computer Based Education), CBT (Computer Based Test), CBELT (Computer Based English Language Test), CAA (Computer Aided/Assisted Assessment), CAT (meaning either Computer Aided/Assisted or Computer Adaptive Test), CALT (Computer Adaptive Language Testing) can be used for testing purposes.

As it is mentioned above, the emergence of computers, Information and Communication Technology (ICT) has affected different areas of educational setting. We can see that the realm of computerized tests is growing and revolutionizing language assessment and testing. Therefore, language proficiency standard tests such as TOEFL and IELTS can be administered in computerized formats. Due to the effects of computer technology in the field of language assessment and testing, this study attempts to investigate different aspects of Computer Adaptive Test (CAT).
II. COMPUTER BASED TESTS

Computer based tests are tests in which computers are used in the testing process. Computer based testing can have a number of advantages. For example, the use of computer in testing field leads to productivity and innovation in this area (Al-Amri, 2009). The standardization of test administration conditions is another advantage of computer-based testing (CBT) (Al-Amri, 2009). Besides, CBT assists test developers to provide the same test conditions for all examinees, regardless of the tests’ population size (Al-Amri, 2009). Moreover, CBT can give immediate feedback to students and to testers, and examinees can take test at any time (Mojarrad et al., 2013). Additionally, different performance data such as latency information can be collected through computer based testing (Olsen et al., 1989).

However, computer based testing may also have some drawbacks. For example, examinees need computer literacy in order to eliminate the mode effect on computer-based testing (Alderson, 2000). Additionally, administering computer based tests requires an equipped computer lab (Rudner, 1998). Furthermore, some of the students may get anxious when tests are presented on a computer (Young et al., 1996). Besides, open ended questions are not usually presented in computerized formats because these kinds of questions are usually scored by human (Brown, 2003). Moreover, human interaction doesn't exist in these tests (Brown, 2003).

Due to the important role of computer in language testing, different scholars were motivated to explore comparability of paper-based tests and computer-based tests. For example, Hosseini et al. (2014) studied on comparability of test results of computer based tests (CBT) and paper and pencil tests (PPT) among English language learners in Iran. 106 Iranian English students which have been selected randomly from Azad University of Tehran participated in this study. All of the participants were second year students. One group was administered two parallel multiple-choice tests, one in paper based format and the other in computerized format. These tests were selected from General English Book. At first, the paper-based format of achievement test was administered to the participants; after two weeks, the computerized equivalent test was given to them. Moreover, a questionnaire which was chosen from Computer Attitude Scale (CAS) made by Loyd and Gressard (1984) and validated by Berberoglu and Calikoglu (1992) was given to the participants to explore their familiarity and attitude towards computer tests. The results of this investigation indicated that participants had better performance in paper-based tests than computer-based test. Additionally, the findings of some other research showed that examinees performed better in paper-based tests compared to computer-based tests (Coniam, 2006; Cumming et al., 2006; Salimi et al., 2011). Besides, Mazzeo et al. (1991) reported that mathematics and English scores of their examinees were greater in paper-based tests in comparison to computer-based tests.

However, De Angelis (2000) conducted a study on the students who took a dental hygiene course. The findings of De Angelis (2000) showed that students' midterm scores in computer-based test were greater than paper-based test.
Moreover, the results of several studies showed that in equivalent test administration conditions, there was not a significant difference between the performance of examinees in CBT and PPT except that examinees prefer computer-based test to paper-based test mode Bachman, 2000; Jamieson, 2005; Chapelle, 2007; Douglas & Hegelheimer, 2007). The findings of Mason, Patry, and Bernstein (2001) also revealed that there was not any significant difference between the results of computer-based and paper-based tests.

Nordin et al. (2010) also studied on the effectiveness of online and conventional mode of Malaysian English Competency Test (MECT) among undergraduates of University Kebangsaan Malaysia (UKM). In this study, the results showed that students' marks in paper test of Malaysian English Competency Test (MECT) and Electronic Malaysian English Competency Test (E-MECT) were not significantly different. Moreover, user friendliness, interface design, interactivity and time management play are features which can affect the success of an online test (Nordin et al., 2010). Additionally, Boo (2012) explored the comparability of scores obtained from computer and paper-and-pencil versions of the Iowa Tests of Educational Development and the examinees' attitudes about these two modes. The results of this study revealed that the scores obtained from computer and paper-and-pencil tests were comparable in terms of scaling (means and standard deviations), internal consistency, and criterion- and construct-related validity. However, examinees preferred the computerized tests to paper and pencil tests.

Besides, Mojarrad et al. (2013) compared paper and pencil reading comprehension assessment to computer-based assessment. 66 male English as a Foreign Language (EFL) learners aged 8 to 12 years from Iran participated in this study. Two different reading comprehension tests with the same level of difficulties on paper and computer screen were administered to the examinee. An attitude questionnaire was given to them to investigate their attitudes towards computerized testing. The findings indicated that examinees' reading comprehension scores across testing modes were not significantly different. Furthermore, most of the students prefer to take the test on computer.

### III. COMPUTER ADAPTIVE TEST (CAT)

Language performance can be assessed through different procedures. Adaptive testing is one of them. Computer adaptive testing is also called “tailored testing” (Madsen, 1991). Noijons (1994, p. 38) defines adaptive testing as "an integrated procedure in which language performance is elicited and assessed with the help of a computer, consisting of three integrated procedures including:

1. Generating the test
2. Interaction with candidate
3. Evaluation of response".
Furthermore, computer adaptive test (CAT) is a kind of computer-based test which adapts to the ability levels' of the participants. In computer adaptive tests, participants' responses to earlier items determine which subsequent items should be administered to them. If an examinee answers the item correctly, the next item will be more complicated. On the other hand, if the examinee answers the item wrongly, the subsequent item will be easier. This process will continue until the examinee's ability level is estimated. Generally, in computer adaptive testing, an item bank is used. This item bank includes a large number of test items which have different difficulty levels and the test items will be selected from it (Larson, 1989). The key point is that the quality of each item should be verified. Item discrimination and item difficulty are two important variables which should be calculated in this regard. Item response theory (IRT) is a suggested method which is used for analysis of computer adaptive testing (Baker, 1983).

IV. THE ORIGIN OF COMPUTER ADAPTIVE TEST (CAT)

Generally, computerized testing originated in the early 1970s (Drasgow, 2002; Wainer, 1990). The emergence of new technologies resulted in development and implementation of computerized testing in large-scale testing programs such as licensure, certification, admissions, and psychological tests (Kim & Huynh, 2007). Specifically, the first Computer Adaptive Test (CAT) was created by Larson and Madsen (1985) at Brigham Young University, in the USA. Besides, After Larson and Madsen (1985), several scholars (e.g., Kaya-Carton, Carton & Dandonoli, 1991; Burston & Monville-Burston, 1995; Brown & Iwashita, 1996; Young, Shermis, Bruten & Perkins, 1996) were motivated to construct and develop more computer adapted tests throughout the 1990s. As a result, some standardized tests were administered in computer-adaptive testing format. In 1998, the Test of English as a Foreign Language (TOEFL) began to use computer-adaptive testing format (Mojarrad et al., 2013). Besides, the Graduate Record Examinations (GRE) has been taken in computer-adaptive format for several years (Mojarrad et al., 2013). Today, we can see different testing programs like Graduate Record Exam, Graduate Management Admission Test, Scholastic Aptitude Test, Microsoft’s qualifications, etc. have used adaptive testing as their testing method (Giouroglou & Economides, 2004).

V. ADVANTAGES OF CATS

Scholars have mentioned a number of advantages regarding the use CATs; for example, test management is flexible, scores are immediately available, and it may motivate examinees (Linacre, 2000; Rudner, 1998) because items are appropriate for their own level and their test anxiety may be reduced (Mulkern, 1998). Besides, efficiency is the main advantage of CATs (Weiss, 1990; Straetmans & Eggen, 1998). CATs are cost saving compared to conventional methods. In conventional methods, different tests should be given to different groups of students and it is very time-consuming to prepare different tests. Marking is also very time-consuming.
The use of CATs decreases the amount of time needed for test preparation and marking and it can increase consistency of the results (Callear and King 1997). The administration of computer adaptive tests is more time saving compared to paper and pencil tests. Moreover, the number of items which are used in computer-adaptive tests and the needed time for answering those items are less than traditional paper and pencil tests (Madsen, 1991; Wainer, 1990). Additionally, the use of CATs can be very beneficial, where a large number of learners should be placed into different classes immediately (Weiss, 1990); because through the flexi-level strategy, examinees do not need to answer a large number of questions which are too difficult or too easy for them. In fact, in computer adaptive tests, the examinees can be given different tests which are appropriate for their own specific level (Larson and Madsen, 1985). Because each test is adapted to each examinee level, more information can be gathered from computer adaptive tests compared to traditional tests (Young et al., 1996). Additionally, one of the advantages of CATs, is "greater precision of measurement", which causes more accurate mastery classification"(Weiss, 1990, p. 454). This greater precision of measurement is due to using items that are at the maximum discrimination level. Besides, the score of each examinee was determined based on both the percentage of questions which were answered correctly and the difficulty level of these questions. As a result, if two examinees answer the equal percentage of questions correctly, the one who answers more difficult questions gets higher score (Economides and Roupas, 2007).

VI. DRAWBACKS OF CATs

Like other type of tests computer adaptive tests have their own special limitations. Therefore, test administrators should be aware of these constraints when they want to use them. Researchers have mentioned some of these drawbacks. For example, test administration procedures of CATs are different compared to paper and pencil tests, and this can be problematic for some examinees (Rudner, 1998). Besides, CATs are based on the Item Response Theory model (IRT) which cannot be used for all item types. Therefore, CATs are not applicable to open-ended questions and items which cannot be calibrated easily (Rudner, 1998). Additionally, another important drawback of CATs is that an examinee is not allowed to go back and change answers because the next items are selected based on the previous answered items (Rudner, 1998). Moreover, item calibration is an important factor which affects the success of a CAT. If the items are not appropriately calibrated on the difficulty/ability scale, the test will be neither valid nor reliable. Additionally, at each difficulty level, several items should be used to make repeated measures at that level possible. Consequently, a large bank of items should be created (Larson, 1987). Furthermore, the determination of performance cutoff points for grading, placement, advancement, remediation, and so on is one of the limitations of computer adaptive tests. Different departments may have different parameters regarding cutoff scores (Larson, 1987). Besides, other scholars believe that Unidimensionality is a drawback of CATs (Madsen 1991, Henning 1991, Canale 1986, Tung 1986; as cited in Starr-Eggle). Generally, computer adaptive tests are based on the item-response-theory (IRT) or latent-trait models. The assumption
which must be met in order to use these models is unidimensionality. It means that all test items must measure one trait (Larson, 1987). Some researchers claim that latent-trait models are not suitable for tests which measure more than one trait (Madsen 1991, Henning 1991, Canale 1986, Tung 1986; as cited in Starr-Eggle). Moreover, developing CAT is time consuming (Madsen 1991, Henning 1991, Canale 1986, Tung 1986; as cited in Starr-Eggle). Furthermore, security can be one of the limitations of CATs. Thus, in CATs, an item pool should be used in order to test the examinees' knowledge. In this process, some items may be chosen more frequently than other items and some items may be memorized and passed on to the other examinees (Wainer and Eignor; 2000).

VII. CONCLUSION

To sum up, technological advancements have moved very rapidly since the last century. Computers have become the most useful facilitator in achieving the majority of our goals. Moreover, in language testing field, technology has been an instrument which led to expansion and innovation in language testing. Today, computer adaptive tests (CATs) make the testing process in line with the needs of e-generation of second language learners by making the testing process innovative, flexible, individualized, efficient and fast.

Hopefully, the use of CATs can have different implications. For example, in academic institutions, it may take a long time to assess a large number of students, and it may be difficult to provide accurate and fast results to the examinees; CATs can help academic institution in this regard. Through the use of CATs, the results of the assessment can be available very quickly, the need for human resources can be reduced, and assessment can be more efficient. Besides, the use of CATs can also have implications for teachers and students. Generally, successful teachers and students should be aware of the new advancements in teaching and testing field and upgrade their skills. Computer literate teachers and students can use computer technology and be innovative. Specifically, through use of CATs, students can assess themselves at any time and at their own pace; consequently, they will become autonomous. Moreover, through use of CATs, scoring and assessment will not be a time consuming process and so on.

However, the use of technology in language testing may have its own caveats. Therefore, authorities should be aware of the positive and negative aspects of a computer adaptive test when they want to administer a test in adaptive format. The important point regarding the use of computer adaptive tests is that to what extent the scores obtained from a computer adaptive test are comparable to the scores obtained from a paper-and-pencil test or a computer-based test (CBT).

In this regard, Wang and Kolen (2001) believe that it is challenging to establish score comparability between CAT and paper test versions because CATs are given online rather than on paper and the other issue is the adaptive nature of CAT tests versus the non-adaptive format of paper-and-pencil administrations. However, in some CAT programs, comparability between
CAT and paper versions of a test has been established through use of score adjustments like those used in test equating (Segall, 1993; Eignor, 1993).

Furthermore, Schaeffer et al. (1995) studied on comparability of scores obtained from linear computer-based (CBT) and computer adaptive (CAT) versions of the three GRE General Test measures. The results of this study showed that the scores of verbal and quantitative CATs were comparable to their CBT counterparts. However, the scores of analytical CAT were not comparable to the analytical CBT scores.

As mentioned above, several scholars studied on comparability of computer adaptive tests and paper-and-pencil tests or computer based test. However, this is an area which requires further investigation.

REFERENCES


