Computer-based Language Test: Characteristics and Processes

Minglie HE

Foreign Languanges Department of the Engineering and Technical,College of Chengdu University of Technology, Leshan 614000, China

hmljob@163.com

Key words: CBLT; characteristics; processes

Abstract: Computer-based Language Test (CBLT) has developed over fifty years and become an integrated part of language test. This paper, based on the analysis of the characteristics of CBLT, explores the three processes of CBLT: construction of the item bank, testing and assessing. It can be foreseen that the modernization of CBLT is of great significance for improving the quality of language teaching and testing.

1. Introduction

Since the 1990s, the development of information technology has had a significant impact on the content and ways of language communication and computers have become an important media of it. For instance, young people have become accustomed to writing mails, applying for a job, checking information and even making friends by internet. Meanwhile, the development of computers and internet has also brought new ideas, ways and challenges for language test. Especially in recent 20 years, the theoretical and practical research of CBLT (Computer-based Language Test) has gone into a higher stage. Roughly speaking, the processes of language test can be divided into such three phases as construction of item bank, testing and assessing. Therefore, this paper will explore how computers can be used in language test on the basis of analysis on its properties.

2. The Characteristics of CBLT

Sometimes CBLT is regarded as "Tailored Test", because computers can produce different kinds of tests according to different kinds of candidates in the shortest time in order to exactly assess the real capabilities of each candidate. Generally speaking, the main characteristics of CBLT are as following:

2.1 Rapidity

It is common sense that there are many multiple choices in language tests, for instance, listening, reading comprehension, vocabulary and grammar etc. Hence, traditionally it will take a lot of time for teachers to make judgments about this kind of test. However, with the development of information technology, candidates can write their answers on the card for such types of test and thus teachers will spend shorter time to grade the papers by logging the answers into computer. What's more, if the examination is carried out on computers, it is convenient to judge the candidates' language competence in accordance with their test response and behavior and hence automatically produce the test paper and offer the feedback information.

2.2 Fairness

CBLT can be divided into double types: standalone language test and web-based language test. The former is usually carried out in a mass-media classroom in which the testing questions are shared from the teacher terminal to the candidate terminal and the testing time is arranged by a certain education department. The latter is much more convenient for candidates to freely select the test time and place. The key point is that candidates will not be disturbed in this situation. In the process of large-scale test, the candidates can get the test directions and content simultaneously;

therefore the regional differences can be reduced and the unfair factors can be eliminated to a certain degree.

2.3 Authenticity

The application of computers in language tests brings big changes for examination papers. With the development of multimedia technology, images, audio and video segments etc. can be widely used in CBLT; thus, the test situation will become more authentic, interesting and vivid. In some tests, encouraging words are even given. Therefore, such kind of authentic feelings help to relieve the tension and anxiety of candidates. What's more, it adds a lot of warm feelings to the cold test. The questionnaires of Stevensons and Cross (1991) shows that many candidates like this kind of language test.

2.4 Reliability

CBLT based on Item Response Theory (IRT) uses a lot of subjective questions; therefore it has higher validity. At the same time, its reliability is also higher, because CBLT enables the individualization of test to be enhanced and test items and speed can be adjusted according to the ability of each candidate, who can choose the examination time according to his or her own time and competitive state, so as to make the test results to reflect the true ability of candidates as much as possible. In addition, when advanced electronic scoring system replaces manual scoring, the grades of the test will become more accurate and reliable.

3. The Three Stages of CBLT

3.1 Construction of Item Bank

Studies on the item bank of CBLT have attracted increasing attention in the world. For instance, related research papers often appear in some international magazines, such as Language Testing. As the significant issue in this field, the construction of language item bank is not only the important way to realize the standard and scientific test but also the important way to realize the departure of teaching and examination. In recent years, it is also a new trend to establish a large-scale language item bank and to realize computerized adaptive test, so most of the large-scale standardized language tests with international influence have their own database. For example, CET-4 and CET-6 in China have established their own item banks. Although the research on the theory and mode of the database is mature both at home and abroad, it is a long way to establish a tenable and practical item bank because the item bank of language test is different from data bank and it is not just a simple set of questions in which each problem must be solved in a certain theoretic model.

At present, the construction of the database theory mainly includes: the stratification theory of discipline system and education goal, the theory of test function and types, the classical test theory (CTT), IRT, the parameter equivalence theory (PET) and the test generation theory, in which the IRT and PET based on CTT are the main theoretical bases used for the construction of item bank. IRT plays a significant role in providing technical parameters, scientific selection, revision and retrieval of the test. PET can guarantee the dynamic nature of the test because it can make the test parameters of different systems turn into a common measuring system and make the parameters of the item bank turn into a new measuring system which is suitable for a new situation after the change of the candidates' language competence. Thus, the related theories of tests and questions enable the item bank to produce qualified papers.

3.2 Testing

Generally, language test includes objective and subjective tasks for candidates. For the former, the test is set by multiple choices in which there is usually only one correct answer for the task;

hence, candidates can make the choice by lightly clicking the mouse. This kind of test is adopted by dichotomously-scored answer, that is to say, the candidate will get two results: getting scores or losing scores. Usually it will influence the validity of the test. In order to effectively control the test validity, education experts have designed simple answers or open-ended questions, which need candidates to give specific answers according to the reading or listening comprehension. For this kind of tasks, polytomously-scored system will be adopted. For instance, if the score scope of each question is 0 to 3 points, appropriate grades should be given according to the specific performance of the candidate. For this kind of subjective questions, using computer software to score is a very big challenge, but the development of his technology now has had great progress. Weigle (2002) introduced three scoring system by computers: Latent Semantic Analysis; Project Essay Grade and E-rater. He pointed out that there are both advantages and disadvantages for them, but each system is constantly improving.

At present, researchers of various systems are trying to find out the most reasonable features to make the scoring system more perfect. Taking E-rater of TOEFL as example, its characteristic to score is based on the following analysis: the analysis of syntax, discourse, theme and vocabulary. Meanwhile, it is taken into account to measure the length of writing in accordance with the words. Although Weigle (2002) holds that this trend is controversial in writing teachers, it has been reported that there is a high correlation between the automatic scoring system and the manual scoring system in wiring (Burstein & Chodorow, 1999). With the improvement of scoring systems, the question types of examination are more and more comprehensive. For example, some foreign universities or testing institutions have begun to use IBT for Test of English Proficiency. Thus the entire examination process from the registration to the performance report is very quick and simple.

3.3 Assessing

The technology of automatic scoring by computer is very mature for objective questions because they have only one answer. However, the score of subjective questions is more difficult because of the subjectivity, openness and divergence of the answers. Automatic scoring is a precursor of subjective questions scoring. As early as in 1966, Ellis Batten Page of the Duke University has invented PEG (Project Essay Grader) scoring system for writing. It mainly extracts the variables of effective text features for multiple regression analysis and ultimately forms the composition scoring model by the analysis of their surface features.

Since 1990s, IEA, E-Rater, INTERMETRIC and other automatic scoring systems have appeared. IEA mainly calculates the semantic similarity between the testing text and the training text by the use of latent semantic analysis (LSA) technology, so as to get the score of the text to be tested. LSA begins with the information retrieval and calculates the semantic relative frequency through the statistic and mathematical analysis as well as the lexical entry analysis in the text. E-Rater has the independent syntactic module, the text module and the theme analysis module and its structure validity is good; in addition, it can be used to comprehensively assess the language and form of compositions. Therefore, it is a mature automatic scoring system. In China, some scholars also developed the automatic composition scoring system which is suitable for English learners in China. This kind of research usually takes advantages of the merits of foreign automatic scoring system, takes into account the characteristics of English learners in China and assesses the quality of compositions from a multi-dimensional angle. It proves that this model can be used to effectively assess compositions of EFL learners in China.

4. Conclusions

CBLT is a new direction of modern educational technology to change the testing means and it is also used in many internationally influential English tests. It can be foreseen that in the future computers can not only be used to assess writing and translation but also be used to assess oral communication and interpretation. Hence, the modernization of CBLT is of great significance for improving the quality of language teaching and testing.

References:

[1] Burstein, J. & Chodorow, M. Automated Essay Scoring Nonnative English Speakers [C]. In Proceeding of the ACL99 Workshop on Computer-mediated Language Assessment and Evaluation of Natural Language Processing, 1999, June.

[2] Stevenson J &Gross S. Use Of A Computerized Adaptive Testing Model For ESOL/ Bilingual Entry/Edit Decision Making [M]. New York: Newbury House, 1991.

[3] Weigle, S.C. Assessing writng[M]. New York: Cambridge University Press, 2002.