Available online at www.sciencedirect.com

SciVerse ScienceDirect

AASRI Procedia 3 (2012) 101 – 106

AASRI Procedia

www.elsevier.com/locate/procedia

2012 AASRI Conference on Modelling, Identification and Control

Exploration and Reflection of the Design Models of Computerized Testing

Chanjuan Liu[1, a*], Jiaxia Li[2,b]

Department of Basic Courses, Shandong University of Science and Technology, Taian, China 271019

Abstract

With the development and popularization of the internet technology, Computerized or Computer-Based Testing （CBT） which help to improve the efficiency of evaluation and then to assess objectively and accurately， will gradually replace the traditional paper-based exam in many ways. This article focuses on four basic design factors：test format ， presentation of test items， the ability estimation and performance report. In practical applications, design of the examination is to optimize the design of these four basic components, to maximize the advantages of the computer and to meet requirements of tests.

Keywords：Computerized Testing. Test library. Adaptive testing .Design patterns

* Chanjuan Liu, Jiaxia Li
Tel 0086+538 3076263 Fax0086+538 3076263
echochanjuan@sina.com,lijiaxia@126.com

doi:10.1016/j.aasri.2012.11.018

1. Introduction

1.1. The development， advantages and characteristics of Computer-Based Testing

Hambleton (2004), the test methods and computer technology develop so fast that it is not easy to predict the future development of educational and psychological testing. But there is no doubt that in the next 20 years, more examinations will be carried out by computers. Application of computer technology has prompted the changes of examination in many ways, for example: Test design develops from the linearity (all candidates take the same tests) to the nonlinearity (the level of the test item adaptive to candidate's ability); from the traditional multiple choice questions to the constructive type ; Because the automatic scoring technology for subjective questions are in rapid development ,cognitive ability as the core of the examination achieve a higher level; Evaluation of the examination introduced procedural parameters, such as: time in response, the answer sequence and modification time with more flexible organization.

Computer-Based Testing research began in the 1970s, matured and then was really applied to the examination practice in the 1990s. Several universities in Europe in 1996 jointly developed European diagnostic language evaluation system-DIALANG which covers 14 European languages, being free of charge and open to Chinese learners. TOEFL_CBT 1998 developed into the TOEFL –

IBT in2005 .GMAT and GRE in 1998 and 1999 began to implement a computerized exam.

In 2004, Guangdong province launched the computerized English test for College Entrance Examination. In 2009, CET Band -4 and Band -6 were taken in some colleges and universities, as well as English listening and speaking test reform occurred in Guangdong Province. In 2011, an annual exam of more than 60 national higher education adults to apply for graduate examination were given on line.

2. The review and present situation of CBT technology

Four advantages of CBT are: Use of innovative test items, adaptive testing process, ability estimation and diagnostic information of test performance. It is necessary to re-recognize and give full play to the advantages and features of the CBT to solve the possible problems of the conventional paper and pencil tests, such as: improving the reliability and validity to reduce the risk of examinations, and increasing the organizational efficiency

2.1. Application of innovative test items

One of the advantages of the CBT is that it can improve the reliability and validity of the exam through the application of innovative test items, which encompass many innovations from five dimensions:

1) Item format, refer to the type of test reaction obtained from the candidate. The two main formats are selected response and constructed response; The multiple-choice questions are one of the most common questions in the exam format .Multiple-choice questions require candidates to select the best answer from of 2-5 ones in pen and paper exams. Questions in this format in the computerized exam can increase the alternatives or to provide a more direct test method to reduce the degree of speculation, such as the candidates click to read sentence in a short passage or part of complex graphics. Another example is in the test of writing skills, candidates are required to face one article containing some of the syntax errors but mislabeled. Candidates consider where necessary to modify and point to wrong location with a mobile key point, then the computer display the options to choose from or to give up. There are other multiple-choice formats which are used in the traditional pen and paper exam and easily transplanted to the computer. Candidates are allowed to repeatedly select to answer and final score is based on total number of selections if timely feedback is correct.

2) Response action refers to the means of candidate s to provide answers. The most common reaction is that candidates use a pencil in the corresponding position to make the selection or coat with an oval in the

computerized exam. It mainly includes the keyboard and mouse click. The examinee can use the mouse to click on a position in the text or graphic article; Highlight text or graphics； Drag text digital or graph to the specified location or order, etc. In addition to the keyboard and mouse, there are other input devices such as touch screen, fluorescence pen, track ball and joystick microphone such as using the microphone to collect oral response etc.

3) The introduction of media, refers to the non-text components added to the test, including images, motion pictures ,audio and video; The most important benefits to introduce media for measuring is perhaps extending the coverage of test content and cognitive skills ; increasing the non-text media and improving the situation of task and validity of examination to reduce the dependency on reading skills

4) Interactivity level: It means to what extent do test items respond to to input of the examinees .Most of the computerized examination papers are non-interactive, such as candidates clicking on the mouse to make the selection to complete the problem solving process. But there are also some of the limited interactions applied to computerized examination papers with relatively small highly interactive tests. There is an interactive combination of two or more steps of branch functions, e.g. a workshop is in conflict, the candidate's task is to resolve it. First step in this application is display of video images of the workshop conflict, and then the multiple-choice questions are related to how to resolve conflicts. Once candidates have selected an option, the second video scene related is on display .This assessment used a two-stage interactive branch level, which can be applied in computerized speaking test.

5) Scoring methods are how to quantify the examinees' responses. There have been some ready automatic scoring software like PEG, Erater intelligent engineer, assessor ,inquizit, standards of which are quite different from each other. some consider simply the shallow characteristics while others apply theory of advanced computational linguistics and methods. They are all similar to people and can act the role as markers and examiners

These five dimensions are independent. However, in most circumstances, the innovation test can only reflect one or two dimensions.

2.2. Other advantages include:

1 More flexible time.
2 The implementation of standardized organizational process.
3 Automatic rating processes.
4 Immediate feedbacks on exam results.
5 A more detailed feed back.
6 Diagnostic analyses provided for test behavior of the candidates.
7 First-hand research data collected for the research.
8 Cost-saving.
9 A more humanized environment.

CBT provide special technical support for people with disabilities, for example: to provide large-print or hearing aid, and provide other input devices for the visually-impaired candidates with writing difficulties. From the above discussion, we have seen many potential advantages of computerized exams, but design and organization of computerized test which can truly reflect these will encounter challenges from many aspects.

Like paper and pencil tests, the most important factor to consider in the design concerns reliability and validity issues. The following are unique factors affecting the reliability and validity of computerized examinations.

3. Design Issues of the CBT

3.1. The basic constitutes of test mode.

From paper and pencil exam to a CBT, the first consideration is the TEST model. Any computer based testing includes item bank ,answer, ability estimate, judgment, and score reporting In addition, the design must in accordance with the following purposes:1)How to select questions from the bank?2)How students answer questions?3)How the computer assess students' answers?4)How many of the questions students need to answer before computerized exam end?5)How to report results after the examination is over?

The development of the computerized exam basically revolves around these six integral parts and the theoretical basis of the computerized exam is mainly derived from measurement theory. Classical True Score Model, CTS) (Item Response Theory, IRT) are the two more popular measurement theories,

3.2. The examination type

Different types of computerized exams can have various forms of organization and implementation ,the commonly used include computerized fixed tests (CFT), automate test of assembly,(ATA),computerized adaptive Test, (CAT)and computerized classification Test,(CCT) The following is focused on one of them.

3.2.1 Computerized adaptive test

Computerized fixed tests and automatic spell volumes test described above can not meet the individual needs of those examinations. However, computerized adaptive testing (referred to as CAT) select the appropriate questions for different candidates. The selection of the test items is decided by candidates' response, which is the way CAT operates. Namely, when each candidates answer a question, the computer re-estimate its ability, as well as choose the one suitable for it until the measurement accuracy reach a predefined level. In computerized test, the number of test items, test order may be different due to different candidates and, but each assessment are based on the item response theory to estimate potential in the same scale.

There are several common design patterns of computerized adaptive testing: The two stage test, Error - controlled tests, variable –branching tests and step ladder tests.

Chart 1summarized the similarities and differences of the six modes.

CAT model contrast

Modes	Start-up project	Project selection rules	Ability estimation methods	Rules to terminate
The two stage test	Ability estimate of every candidate in the established practice testing model	Maximum test information function	Average difficulty scoring method	Test information function
Step ladder tests	Random item in the middle area	To move up and down According to the item response in each area	Maximum likelihood value calculation	Right number of item a certain of domain
Variable – branching tests	Item of Medium difficulty	Impairment According to the constants	---	Fixed quantity

Error - controlled tests1	---	The difficulty of the item matching the examinees ability	---	Standard error of the measurement

The implementation of two-step tests is in two stages

Firstly, all candidates participate in routine test and then secondly the optimizing test .To begin with ,optimizing test estimates of the ability of the candidates in the first stage and computer select the most optimized test items for each candidate, rules of which are to provide maximum test information with the least items. The maximum amount of information is also used as the termination principle; the average difficulty score method is the way to estimate ability (Weiss, 1974).

Degree of difficulty show s each candidate's ability to answer all of the questions correctly .In the error control test mode, due to the transformation of the initial project and the project selection in the actual operation of this model, there are three variations of which basic steps are exactly the same. Variable branching start from a moderately difficult project selection, the general principle is that candidates answered all the items if the answer is correct, otherwise go from the easier to the more difficult projects. Maximum likelihood estimation methods are used to estimate students' abilities. .The above-described are four basic model of machine variants, using different methods to accomplish essential parts in the CAT.

4. Conclusion

4.1. Errors and misconception

It is important to have a correct understanding that computerized testing is a trend, which is to achieve the individualization and (diagnostic), provide the adaptive examination of intelligence, to meet the learners life-long education demand in recent years, many examinations began to demonstrate feasibility However, in this process, many misconception exist .

1 Computerized exam is taken for paper and pencil exams computerized.

2 Computer Technology determines the development of the computerized exam.

3 Taking test set as test bank.

4 The computerized exam can simplify the work of the Examination. The test problem

In the design of the computerized exam, we must fully consider what factors may influence test behavior, especially the fairness of the examination, for example, whether every candidate are given equal opportunities for training and the same learning resources. Other problems include emotional attitude, proficiency in computer operating user interface software design, restrictions of the examination task, and the psychological state of the candidates in the exam.

4.2.1Organization is one of the important concerns in the development of computerized examination.

The following may some of the major encountered problems in the examination:

1Examination venues: choice of the examination venues include: number of seats and computers, security, test data storage and transmission. Some exam may also be limited by the software which can be used the examination process. For the same exam, all examination venues must be consistent in soft, hardware.

2The candidates must be able to know relevant information of the examination room in advance.

3Security of test items and examination process, such as examination fraud, hacker attacks. Input costs, the candidates are required to bear the cost.

4The examination time of each test and frequency of test organized. Annual frequency of computerized exam depends on

(1)The number of candidates

(2) Number of computers

(3)Time use and other issues also need to be seriously considered.

(4)Choice of examiners requires a rigorous process;

Difference between them and the organization of the traditional pen and paper exam must be able to manage, or even be able to solve emergency problems.

References

[1] Hamston,R.K. Theory methods and practices in the testing for the 21st century. [J], 2004,16(4):696-701

[2] Folk VG. Smith RL Models for Delivery of CBTs 2002

[3] Luecht, RM and Clauser, BE. Test models for complex computer based testing. [A] In C.N. Mils M.T. Potenza, J.J. Framer and W.C. Ward, Computer-based Testing: Building the foundation for future assessments [C]. Hillsdale, N.J: Lawrence Erlbaum Assicates, 2002

[4] Weiss DJ and Kingsbury, G.G. Application of computerized adaptive testing to educational problems. [J] Journal of Educational Measurements, 1984, 4; 361-375

[5] ZENGYONG- qiang,Some Basic Considerations into the Computer -based Testing，[J]Computer Assisted Foreign language Education 2010(1)；（in Chinese)

[6] ZENGYONG- qiang,Design Models of Computer Adaptive Test [J] Foreign Language Teaching and Research 1992（2）；19-22（in Chinese).