Arab Academy for Science, Technology & Maritime

College of Computing And Information Technology

Information Systems Department

**Protocol for Master Degree in**

**Information Systems**

**A Web Based Adaptive Testing System For Web Programming: A Case Study in AAST**

## Introduction

The use of computers and electronic devices in general for online processes has become increasingly massive. Such progress in computer technologies has been reflected to a great extend in many parts of educational and pedagogical methodologies; i.e. learning, testing and assessment.

The term "e-learning" has been introduced in 1999 at a CBT systems seminar. Other terms surfaced in search for an accurate description, such as “online learning” and “virtual learning”. Yet, the fundamental principles and values underlying e-learning have been clearly identified and distinctly recognized.

Although the term "e-learning" has only been used for the first time in 1999, distance courses were offered; long before the internet, to provide students with education on particular subjects or skills. Evidences, however, suggest that early forms of e-learning existed as far back as the 19th century.

E-learning aims at facilitating learning and improving performance by creating, using, and managing appropriate technological processes and resources; helping people thus to learn in ways that are easier, faster, reliable, and less expensive.

In an attempt to reduce the cost of exams, minimize human efforts, speed up scoring, eliminate human errors and increase accuracy many educational organizations are trying to improve examination settings. Accordingly, many examination organizations have shifted from paper-based examinations to computer-administered examinations.

Computer-based testing (CBT) is a feasible alternative to paper-and-pencil testing. However, the transition to CBT is neither easy nor inexpensive5. Many design engineering, test development, operations/logistics, and psychometric changes are required to develop a successful operational program9. Initially, CBT researches focused solely on the theoretical matters; i.e. improving adequacy levels and test score reliability, reducing time, diminishing human interference and accordingly decreasing cost. Nowadays, CBT is a general term that embraces a large variety of assessment types, purposes, test delivery designs, content balancing models, item selection designs, and psychometrics; thus CAT has been introduced.

Computerized adaptive testing (CAT) is a sophisticated method of delivering examinations, and has nearly 40 years of technical research supporting it.1 Computerized Adaptive Testing (CAT) is a computer-based test that follows a certain predefined approach, according to which questions are adapted and customized. Generally, adaptive tests are composed of a set of questions that are randomly or specifically chosen from a bigger set of questions, i.e. question bank, items are selected to correspond to the skills and capabilities of the test taker.

Many of the earliest computer-based tests moved abruptly to the computerized adaptive testing (CAT) mode, making use of faster computers and network technologies and aiming for more accuracy, less cost, minimal human involvement and shorter tests. In fact, there is no single CBT model that is ideal for all educational tests, rather, all models have their strengths and weaknesses, and some are better suited to the characteristics of a particular testing program than others.

## Problem & Motivation

Advantages of CBT can include shorter, quicker tests, flexible testing schedules, increased test security, better balancing of test content areas for all ability levels, quicker test item updating, quicker reporting, and a better test-taking experience for the test-taker.3

Models for delivering computer-based tests vary in their characteristics and elements. The aspects of variance include; the degree and nature of test adaptation; size and flexible units of test administration; user interface issues; automated test assembly and test form quality controls; and security risks.

A fundamental aspect that distinguishes among different CAT models is the degree of adaptation. An adaptive test tailors the difficulty of the test items to the ability or proficiency of each examinee. The specific goal in a purely adaptive test is to maximize the test reliability for every examinee, regardless of his or her score. Tailored or adaptive testing therefore leads to certain measurement efficiencies where a particular level of reliability can be achieved with fewer items and in shorter time. Accordingly score precision can be improved relative to a nonadaptive fixed-length test.

Unfortunately, CAT deployment doesn’t endure without concerns10. Certainly, the most significant concern detected is item overexposure5. The item selection algorithm applied in CAT application may cause particular items in the item pool to be over-selected; i.e. items providing the most discriminating information about the examinee’s ability may be presented frequently to various participants to the extent that they become familiar to examinees prior to testing, thus reducing test security and reliability.

Moreover, in a case of test items being over-selected and exposed, additional item development would be crucial, increasing accordingly costs of CAT maintenance, while at the same time large proportion of the item pool remains unused10. To overcome and eliminate the overexposure problem along all its consequences, the item selection method has to choose discriminating items while considering pool utilization.

Item exposure control as well as the related test security issues have been a major concern in the development and implementation of CAT. CAT item's selection is based primarily on achieving the maximum level of precision in estimating an examinee’s competence. Nevertheless, certain pool items tend to be used more often than others, making item exposure rates entirely biased. If the high exposure of some of the items in the MST procedure is of a concern, then the use of more blocks in the first stage is one way to reduce the high exposure rates of these items8.

E-Learning in general, requires more discipline by learners and more effort by professors and advisers than traditional courses. Same issue applies to CAT, which requires extensive examiner interaction with the learners' answers and results, rather than the learners themselves, to come up with appropriate guidelines that would help students in their studies.

Efficient and effective assessment of a learner's proficiency is a high priority for e-Learning environment that would allow e-Learning system to adjust its teaching and curriculum accordingly. It is a crucial criterion to achieve additivity 4. Providing a balanced exposure rate, assessing learner's answers precisely and providing a detailed feedback about points of strength and weaknesses would allow the examinee to gain a profound image of the learner's understanding of the curriculum.

## Objective

CAT may have conflicting goals of maximizing test efficiency and limiting the overexposure of individual items, while measuring the identical composite of traits across examinees through administration of items with the same content properties (Davey and Parshall - 1995)5. Different methods have been proposed and applied in an attempt to achieve content balancing while retaining test efficiency and keeping the exposure rate of items to a minimum. Several simulation studies took place to evaluate the performance of different content balancing methods under different conditions of length and item exposure.

In a previous study by Chi-Keung Leung, Hua-Hua Chang, and Kit-Tai Hau (2003), three content balancing methods have been examined, including the constrained CAT (CCAT), the modified multinomial model (MMM), and the modified constrained CAT (MCCAT). In this study, a research has been conducted to compare the three content-balancing methods using three different item selection designs: multi-stage a-stratified design (ASTR), a-stratified with b-blocking design (BASTR), and content-stratified, a-stratified with b-blocking design (CBASTR).

The research results revealed that the three examined content balancing methods resulted in identical resemblances between examinees estimated and real abilities. Yet, deviations were detected regarding item pool utilization, with the MMM method generally yielding best performance amid all three stratification designs.

In this study, the modified multinomial model (MMM) examined in the previous research performed by Chi-Keung Leung, Hua-Hua Chang, and Kit-Tai Hau, will be applied in a web based software application, in an attempt to produce a CAT application using the content balanced model with best performance and least exposure issues.

## Contribution

In CAT, tailored tests are submitted to examinees; every single item is selected at a time on the basis of the currently available estimate of the examinee’s ability (Lord, 1980; Weiss, 1982). The most significant advantage of CAT is enabling more efficient and precise trait estimation (Owen, 1975; Wainer, 1990). The key issue in CAT is the mean by which the best test items are adaptively selected from the item pool, which relies primarily on the former item's response. The traditional item selection algorithms rely on local item information, which means that an item is selected if it has the maximum Fisher information at the current ability estimate based on the responses to previously administered items11. However, this information criterion would cause skewed item exposure (Davey & Parshall, 1995; Sympson & Hetter, 1985; van der Linden, 1998). Items with large value of discrimination parameter may be overly exposed while some others are never used11. Yet, over-exposure would certainly risk item security, while under-utilizing those inactive items in the pool, reducing thus the cost effectiveness of developing and maintaining. Therefore, item exposure control and the improving pool efficiency are critical issues in CAT designs.

In an attempt to improve CBT in general and CAT algorithms specifically, a web based software application is to be developed. The proposed content balancing model, Modified Multinomial Model (MMM) has been previously examined and designated as one of the best models that succeeded to grant results resembling real examinees' abilities and the best model with least item exposure issue.

The Modified Multinomial Model (MMM): Primarily, a cumulative distribution is formed based on the target percentages of the content areas that sum to 1.0 and follow a multinomial distribution. Subsequently, a random number from the uniform distribution U(0,1) is used to determine the corresponding content area in the cumulative distribution where the next optimal item will be selected. Thus, whenever a target percentage is reached, a new multinomial distribution needs to be formed by adjusting the rest percentages of the remaining content areas11.

The exposure rate of an item is defined as the ratio of the number of times the item is administered to examinees over the total number of examinees taking the test11. Items with high exposure rates are in a greater risk of being known to prospective examinees, which accordingly would lead to test security and validity problems.

On the other hand, a large number of items with low exposure rates is an indicator that the item pool is not well utilized, which challenges directly the cost effectiveness of the item pool and the appropriateness of the item selection method.

## Methodology

In an attempt to improve CBT in general and CAT algorithms specifically, a web based software application is to be developed. The proposed content balancing model, Modified Multinomial Model (MMM) has been previously examined and designated as one of the best models that succeeded to grant results resembling real examinees' abilities and the best model with least item exposure issue.

Initially, a profound analysis of the MMM model and its development history will take place. Understanding the model and how it evolved will facilitate the attempt to apply it more efficiently.

A design of the proposed model will follow, which will establish a robust base for the following phases. The proposed design should emphasize its significant characteristics.

Based on the proposed design, a prototype will be created, which should be an exact emulation of the design. The prototype is a physical mean to assess and evaluate the proposed design and to trace and resolve any possible existing pitfalls.

After prototype creation and evaluation, the implementation will take place, based on the performed analysis and the final design. The implementation phase, however, will include several consecutive steps, i.e. Case Study, Database Design, Class Diagram, Coding and Testing. These steps are sequential and iterative.

The suggested model would be implemented in a web-based form, allowing thus remote access to learners, examinees and administrators, without the need to be physically working on the server, which mimics most CAT application currently in place. To create the web application, the web programming language, PHP, will be used, along with the database engine, MySQL. An Apache server will be set to host the whole framework.

The final phase, would be writing the thesis, that should include a thorough description of each and every performed step. It should also include all findings and any related information about the application.

## References

1. A Framework for the Development of Computerized Adaptive Tests, Nathan A. Thompson & David J. Weiss, Practical Assessment, Research & Evaluation - Volume 16, Number 1, January 2011.
2. Computerized Adaptive Testing: Theory and Practice,  Wim J. van der Linden & Cees A. W. Glas, Kluwer Academic Publisher - 2002
3. Computer-Adaptive Testing: A Methodology Whose Time Has Come, John Michael Linacre, Seoul, South Korea: Komesa Press-2000
4. E-Learning Paradigms and Applications: Agent-based Approach, Mirjana Ivanović, ‎Lakhmi C. Jain - 2013
5. A Review of Models for Computer-Based Testing, Richard M. Luecht and Stephen G. Sireci, College Board 2012
6. Computer Based Testing-Evaluation of Question Classification for Computer Adaptive Testing
7. [http://echo.edres.org:8080](http://echo.edres.org:8080/irt/), Lawrence M. Rudner, 2001
8. Comparison of Multi-Stage Tests with Computerized Adaptive and Paper and Pencil Tests, Ourania Rotou et al., American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME), 2003
9. Drasgow, F., Luecht, R. M., & Bennett, R. (2006). Technology and Testing. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 471–515). Washington, DC: American Council on Education/Praeger Publishers.
10. Computerized Adaptive Testing: A Comparison of Three Content Balancing Methods, Chi-Keung Leung, Hua-Hua Chang, and Kit-Tai Hau, 2003
11. An Examination of Item Selection Rules by Stratified CAT Designs Integrated with Content Balancing Methods, Chi-Keung Leung, Hua-Hua Chang, and Kit-Tai Hau, 2014