Triantafillou, E., Georgiadou, E., & Economides, A. (2006). CAT-MD:Computer adaptive test on mobile devices. In Méndez-Vilas,A. Solano Martín, J.A. Mesa González and J. Mesa González (eds.), "Current Developments in Technology-Assisted Education (2006), Vol. II: Technological Science Education, Collaborative Learning, Knowledge Management", Proceedings IV International Conference on Multimedia and ICTs in Education (m-ICTE2006), pp. 773-777. Seville, Spain, Formatex, Badajoz, Spain (2006).

CAT-MD: Computer Adaptive Test on Mobile Devices

Evangelos Triantafillou*, Elissavet Georgiadou, Anastasios A. Economides

University of Macedonia, Egnatias 156, Thessaloniki 54006, GREECE

The use of computerized-adaptive testing (CAT) has expanded rapidly the last years mainly due to the advancements in communication and information technology. Availability of advanced mobile technologies provides several benefits to e-learning and testing by creating an additional channel of access with mobile devices such as PDAs and mobile phones. This paper describes the design issues that were considered for the development and the implementation of a CAT on mobile devices, the CAT-MD (Computerized Adaptive Test on Mobile Devices).

Keywords Computerized Adaptive Test; Mobile Learning

1. Computerized Adaptive Test

In computerized adaptive test (CAT), a special case of computer-based testing, each examinee takes a unique test that is tailored to his/her ability level. As an alternative of giving each examinee the same fixed test, CAT item selection adapts to the ability level of individual examinees and after each response the ability estimate is updated and the next item is selected to have optimal properties at the new estimate [13]. The CAT presents first an item of moderate difficulty in order to initially assess each individual's level. During the test, each answer is scored immediately and if the examinee answers correctly then the test statistically estimates her/his ability as higher and then presents an item that matches this higher ability. The opposite occurs if the item is answered incorrectly. The computer continuously re-evaluates the ability of the examinee until the accuracy of the estimate reaches a statistically acceptable level or when some limit is reached; such as a maximum number of test items. The score is determined from the level of the difficulty, and as a result, while all examinees may answer the same percentage of questions correctly the high ability ones will get a better score as they answer correctly more difficult items.

Regardless of some disadvantages reported in the literature –for example, high cost of development, item calibration, item exposure [5, 3], the effect of a flawed item [1], or the use of CAT for summative assessment [8] – CAT has several advantages. Testing on demand can be facilitated so as an examinee can take the test whenever and wherever s/he is ready. Multiple media can be used to create innovative item formats and more realistic testing environments. Other possible advantages are flexibility of test management; immediate availability of scores; increased test security; increased motivation etc. However, the main advantage of CAT over any other computerized based test is efficiency. Since fewer items are needed to achieve a statistically acceptable level of accuracy, significantly less time is needed to administer a CAT compared to a fixed length Computerized Based Test [11, 9].

2. Mobile Learning

* Corresponding Author : email : vtrianta@csd.auth.gr, Phone: +30 2310420690

In the recent years the use of different mobile products such as mobile phones and Personal Digital Assistant (PDA) devices has increased rapidly. Moreover, much attention has been paid to mobile computing within Information Technology industry. Availability of advanced mobile technologies, such as high bandwidth infrastructure, wireless technologies, and handheld devices, has started to extend e-learning towards mobile learning [12]. Mobile learning (m-learning) intersects mobile computing with e-learning; it combines individualized (or personal) learning with anytime and anywhere learning. The advantages of m-learning include: flexibility, low cost, small size, ease of use and timely application [7].

The introduction of mobiles devices into the learning pedagogy can compliment e-learning by creating an additional channel of assessment with mobile devices such as PDAs, mobile phones, portable computers. Due to their convenient size and reasonable computing power, mobile devices have emerged as a potential platform for computer-based testing. Although, mobile computing has become an important and interesting research issue, little research has been done on the implementation of CAT using mobile devices and this is the focus of our research. The current study is an attempt to examine the design and development issues, which may be important in the implementation of a CAT using mobile devises such as mobile phones and PDAs. As a case study an educational assessment prototype was developed, called CAT-MD (Computerized Adaptive Test on Mobile Devices), to support the assessment procedure of the subject "Physics" which is typically offered to second grade students in senior high school in Greece.

3. System Architecture

The prototype CAT-MD uses the Item Response Theory (IRT) as an underlying psychometric theory, which is the base for many adaptive assessment systems and depicts the relationship between examinees and items through mathematical models [10, 6, 14]. Psychometric theory is the psychological theory or technique of mental measurement, which is the base for understanding general testing theory and methods. The central element of IRT is mathematical functions that calculate the probability of a specific examinee answering a particular item correctly. IRT is used to estimate the student's knowledge level, in order to determine the next item to be posed, and to decide when to finish the test.

In IRT-based CAT as each student answers a question, his or her response is evaluated as being either correct or incorrect. The process of displaying questions, evaluating responses and selecting the next question to be administered based on the student's latest estimated ability is repeated until a stopping rule has been reached or a certain number of questions has been administered, whichever happens first. There are four main components needed for developing IRT-based CAT: the item pool, the item selection procedure, the ability estimation and the stopping rule [4]. The following sections describe these components of the CAT-MD system.

3.1. Item pool

The most important element of a CAT is the item pool that is a collection of test items that includes a full range of levels of proficiency, from which varying sets of items are presented to the examinees. The item parameters included in the pool are dependent upon the Item Response Theory (IRT) model selected to model the data and to measure the examinees' ability levels. In IRT-based CATs, the difficulty of an item describes where the item functions along the ability scale. For example, an easy item functions among the low-ability examinees and a hard item functions among the high-ability examinees; thus, difficulty is a location index.

An ideal item pool needs many items, best spread evenly over the possible range of difficulty. In our approach CAT-MD includes a database that contains 80 items related to the chapter "Electricity" from the "Physics" subject. For every item, the item pool includes the item's text, details on the correct answer and the difficulty level. The difficulty level varies form "very easy" to "very hard" and the values used cover the range between -2 and +2.

3.2. Item selection

In IRT theory, the item selection procedure is the process of selecting an item from the item pool to be administered to the examinee. A reasonable assumption is that each examinee responding to a test item possesses some amount of the underlying ability. Thus, one can consider each examinee to have a numerical value, a score that places him or her somewhere on the ability scale. This ability score will be denoted by the Greek letter theta, *θ*. At each ability level, there will be a certain probability that an examinee with that ability will give a correct answer to the item. This probability will be denoted by *P(θ).* In the case of a typical test item, this probability will be small for examinees of low ability and large for examinees of high ability [2].

The main aspect of IRT is the Item Characteristic Curve (ICC) [2]. ICC is an exponential function, which gives the probability of answering a question of certain difficulty level correctly by a learner with certain skill level. ICC is a cumulative distribution function with a discrete probability. The models most commonly used as ICC functions are the family of logistics models of one (1PL), two (2PL) and three parameters (3PL). The 1-parameter logistic (1PL), or Rasch model is the simplest IRT model. The Danish mathematician Georg Rasch first published the 1-parameter logistic model in 1960s and as its name implies, it assumes that only a single item parameter is required to represent the item response process. This item parameter is termed difficulty and the equation for this model is given by:

$$P(\theta) = \frac{1}{1 + e^{-1(\theta - b)}} \qquad \textbf{(1)}$$

where, *e* is the constant 2.718*, b* is the difficulty parameter and *θ* is an ability level.

The Item Information Function (IIF) is also considered as an important value in the IRT's item selection process. It gives information about the item to be presented to the leaner in an adaptive assessment. For selecting a question appropriate to the learner, IIF for all the questions in the assessment should be calculated and the question with highest value of IIF is presented to the learner. This provides more information about the learner's ability and is given by the equation:

$$I_i(\theta) = P_i(\theta)(1 - P_i(\theta)) \qquad \textbf{(2)}$$

where $P_i(\theta)$ is the probability of a correct response to item *i* conditioned on ability *θ* [2, 10].

3.3. Ability estimation

After each item is administered and scored, an interim estimate of examinees' ability (*θ*) is calculated and used by the item selection procedure to select the next item. The most commonly used estimation procedure is maximum likelihood estimation (MLE) [10]. Similar to the item parameter estimation, this procedure is an iterative process. It begins with some a priori value for the ability of the examinee. In CAT-MD, it begins with *θ*=1. The estimation calculation approach is the modification of the Newton-Raphson iterative method for solving equations method outlined by Lord. The estimation equation used is shown below:

$$\theta_{n+1} = \theta_n + \frac{\sum_{i=1}^{n} S_i(\theta_n)}{\sum_{i=1}^{n} I_i(\theta_n)} \qquad \textbf{(3)}$$

where

$$S_i(\theta) = [u_i - P_i(\theta)]\frac{P_i'(\theta)}{P_i(\theta)[1 - P_i(\theta)]} \quad \textbf{(4)}$$

where θ is the skill level after n questions, and $u_i = 1$ if the response is correct and $u_i = 0$ for the incorrect response.

3.4. Stopping Rule

One important characteristic of CAT is the test termination criterion. The termination criterion is generally based on the accuracy with which the examinees' ability has been assessed. In most CATs, the termination of the test may be based on the number of items administered, the precision of measurement or a combination of both [3]. Measurement precision is usually assessed based on error associated with a given ability. The standard error associated with a given ability is calculated by summing the values of the item information functions (IIF) at the candidate's ability level to obtain the test information. Test information, TI(θ), is given by the equation:

$$TI(\theta) = \sum_{i=1}^{N} I_i(\theta) \quad \textbf{(5)}$$

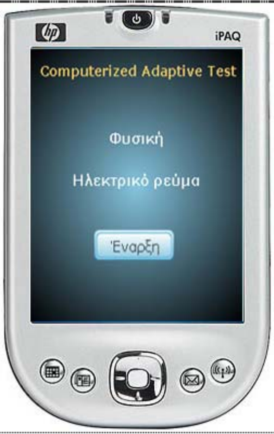
Next, the standard error is calculated by using the equation:

$$SE(\theta) = \frac{1}{\sqrt{\mathrm{TI}(\theta)}} \quad \textbf{(6)}$$

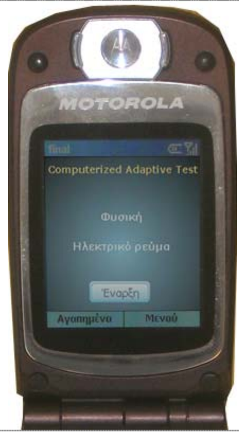
After each administration of an item, the standard error associated with a given ability is calculated to determine whether a new item must be selected or whether the administration of the test can be terminated. It is common in practice to design CATs so that the standard errors are about .33 or smaller [11]. In CAT-MD the test terminates for each examinee when the standard error associated with a given ability (θ) is less than 0.30 or when the maximum number (that is 20) of items has been administered.

4. System Implementation

Currently, the basic architecture of the system has been implemented. The prototype software has been developed using Macromedia Flash as it offers competitive advantages. It is a lightweight, cross-platform runtime that can be used not just for enterprise applications, but also for communications, and mobile applications. According to Macromedia Company the 98 percent of all Internet enabled computers and 30 million mobile devices use the Flash technology (www.macromedia.com). To date, many manufacturers license Macromedia Flash on their branded consumer electronics devices, such as mobile phones, portable media players, PDAs, and other devices. These licensees include leading mobile device manufacturers such as Nokia, Samsung, Motorola, and Sony Ericsson.



The CAT-MD on HP iPAQ (PDA)



The CAT-MD on Motorola MPx220 (mobile)

Fig. 1 Interface of CAT-MD

Figure 1 presents two screenshots of the implementation of CAT-MD on a mobile phone and on a PDA. Moreover, the CAT-MD is portable to any device that has installed the Macromedia Standalone-Flash Player. In addition, if a Macromedia plug-in for the web browser (Internet Explorer, Mozilla, etc.) is installed, the CAT-MD can be also accesses as flash shockwave film.

5. Conclusion

This article describes the design and development of the CAT-MD (Computerized Adaptive Test on Mobile Devices), a prototype CAT on mobile devices such as PDAs. Formative evaluation is the next step of our research in order to investigate the effectiveness and efficiency of the system and also to assess its usability and appeal.

Acknowledgments

The work presented in this paper is partially funded by the General Secretariat for Research and Technology, Hellenic Republic, through the E-Learning, EL-51, FlexLearn project.

References

[1] Abdullah, S.C. (2003). Student Modelling By Adaptive Testing - A Knowledge-Based Approach. Unpublished PhD Thesis, University of Kent at Canterbury, June 2003.

[2] Baker, F. (2001). The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.

[3] Boyd, A. M. (2003). Strategies for Controlling Testlet Exposure Rates in Computerized Adaptive Testing Systems. Unpublished PhD Thesis, The University of Texas at Austin, May 2003.

[4] Dodd, B. G., De Ayala, R. J. and Koch W. R. (1995). Computerized adaptive testing with polytomous items. Applied Psychological Measurement, **19** (1), 5-22.

[5] Eggen, T.J.H.M. (2001). Overexposure and underexposure of items in computerized adaptive testing. Measurement and Research Department Reports 2001-1, Citogroep Arnhem.

[6] Hambleton, R. K., Swamination, H. and Rogers, H. J. (1991). Fundamentals of Item Response Theory. California: Sage Publications Inc.

[7] Jones V. and Jo H. J. (2004).Ubiquitous learning environment: An adaptive teaching system using ubiquitous technology. Proceedings of the 21st ASCILITE Conference, Perth, Western Australia, 5-8 December 2004.

[8] Lilley, M. and Barker, T. (2002). The Development and Evaluation of a Computer-Adaptive Testing Application for English Language, 6th Computer Assisted Assessment Conference, July 2002, Loughborough.

[9] Linacre, J. M. (2000). Computer-Adaptive Testing: A Methodology whose Time has Come. MESA Memorandum No. 69. Published in Sunhee Chae, Unson Kang, Eunhwa Jeon and J.M. Linacre. Development of Computerised Middle School Achievement Test (in Korean). Seoul, South Korea: Komesa Press.

[10] Lord, F.M. (1980). Applications of Item Response Theory to Practical Testing Problems. New Jersey: Lawrence Erlbaum Associates, Publishers.
[11] Rudner, L.M. (1998). "An online, interactive, Computer Adaptive Testing Tutorial". 11/98. Available at http://EdRes.org/scripts/cat
[12] Sharples, M. (2000). The Design of Personal Mobile Technologies for Lifelong Learning, Computers and Education, **34** (2000), 177-193.
[13] van der Linden W.J. and Glas, C.A.W. (2003). Preface. In van der Linden, W.J., Glas, C.A.W (Eds). Computerised Adaptive Testing: Theory and Practice. Dordrecht, Boston, London: Kluwer Academic Publishers, vi-xii.
[14] Wainer, H. (1990). Computerized Adaptive Testing (A Primer). New Jersey: Lawrence Erlbaum Associates.