# Computerized Adaptive Test – Adapting to What?

*Evangelos Triantafillou*

*Edutech, Dodekanisou 27, Thessaloniki 55131, Greece, vtrianta@edutech.gr*

**The use of Computerized Adaptive testing has expanded rapidly the last two decades mainly due to the advancements in communication and information technology that made computers with high power and speed affordable and effortlessly connected to broad bandwidth networks. Since the mid-80s when the first Computerized Adaptive Testing became operational much research and many technical challenges have made new assessment tools possible. The scope of this paper is to review and examine the different variables that have been used in adaptive educational systems and then discuss their potential use to a hypothetical user model for Computerized Adaptive Testing.**

**Keywords**
Adaptive variables, Computerized Adaptive Test, Computerized assessment, User model

## 1. Computerized Adaptive Test

Testing is one of the most widely used tools in higher education. The main goal of testing is to measure student knowledge level in one or more concepts or subjects, i.e. in pieces of knowledge that can be assessed. Since education was established as an institution, different methods of assessment were used in different contexts, such as class presentations, essays, projects, practicum, etc. However, the most common tools of measuring performance are the oral test and the 'paper-and-pencil' test. Given that the computer has been an educational tool over the last few decades and its use has spread rapidly in all levels of education and training, the use of computer-based tests (CBTs) has increased significantly over the last few years. CBTs became feasible for licensure, certification and admission.

The most common type of CBT is the linear one that is a fixed-length computerized assessment that presents the same number of items to each examinee in a specified order and the score usually depends on the number of items answered correctly. A linear test consists of a full range of easy and difficult test items that are either randomly selected from a larger pool or are the same for all examinees. Evidently the type of CBT described here imitates a 'paper-and-pencil' test that is presented in a digital format and pays little or no attention to the ability of each individual examinee.

By contrast, in computerized adaptive testing (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 to giving each examinee the same fixed test, CAT item selection adapts to the ability level of individual examinees. After each response the ability estimate is updated and the next item is selected such that it has optimal properties according to the new estimate [1]. The CAT first presents 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. If the next item is again answered correctly, it re-estimates the ability as higher still and presents the next item to match the new ability estimate. 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 correctly answer more difficult items.

There are many advantages recorded in the literature with regards to CAT such as flexibility of test management; scores immediate availability; increased test security; increased motivation etc. However, the main advantage of CAT over any other computerized based test is efficiency. Since fewer questions 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 [2],[3].

Since the mid-80s when the first CAT systems became operational, i.e. the Armed Services Vocational Aptitude Battery for the US Department of Defense account using adaptive techniques to administer multiple-choice items, much research and many technical challenges have made new assessment tools possible [1]. Several recognized testing programs such as GRE (Graduate Record Exam), GMAT (Graduate Management Admission Test), SAT (Scholastic Aptitude Test), Microsoft's qualifications, etc. have adopted adaptive testing as their current method for testing.

Most CAT systems include a user model. The variables of the user model describe characteristics of examinees, such as knowledge, skills and abilities, about which the user of the assessment wants to make inferences. However, the main goal of the vast majority of CAT systems is to arrange examinees on a problem complexity scale that is relevant for graduation/admission decisions. As a result, user models used by these systems do not include a large array of user variables. They usually contain variables representing the aspects of proficiency that are the targets of inference in the assessment.

Current research in CAT is not limited to educational admissions, yet, focus on applications, in small and large scale, that address self-assessment, training, employment, teacher professional development for schools, industry, military, assessment of non-cognitive skills etc. Moreover, dynamic item generation tools and automated scoring of complex constructed-response examinations reaches operational status [4]. Therefore, it is important to extend CAT's functionality to include more variables in its user model that define the examinee as an individual beyond the mastery level, for improved performance and more efficient test delivery.

Research on personalised hypermedia applications and especially Adaptive Educational Hypermedia Systems (AEHS) has identified a number of variables that can prompt adaptivity. Contributions from general areas such as user modelling, student modelling, intelligent tutoring systems are also relevant to this issue. Evidence of the interconnection of the above research fields with CAT is that AEHS incorporate CAT in their architecture in order to extend the adaptive capabilities of the systems and support learning (e.g. INSPIRE [5], ELMART [6], DCG [7]). Moreover CAT is used as a student modelling technique in Intelligent Tutoring Systems [8], [9].

## 2. Adapting to What?

According to literature, a number of different adaptive variables acknowledged by researchers in the area of personalized adaptive systems [10], [11], [12], [13]. Adaptive variables refer to the features of the user that are used as a source of the adaptation, i.e. to what features of the user the system can adapt its behaviour. The variables that user models include can be classified to 'user dependent' that includes those directly related to the user and define him/her as an individual, and to 'user independent' that affect the user indirectly and are related mainly to the context of a user's work with a hypermedia application [14]. The user dependent variables are: (a) knowledge on the domain presented, (b) background - experience, (c) preferences, (d) interests, (e) individual traits, (f) user personality, (g) mental model, (h) personal data, (i) abilities/disabilities, (j) social-group. On the other hand, the user independent variables are: (a) current goal/task, (b) environment-work, and (c) situation variables.

A CAT in order to be more efficient than a fixed-length computerised test, initially assess each individual's level by presenting first an item of moderate difficulty. However, if the *Knowledge of the domain* variable is modelled for each individual then this initial question could be more closer to the examinee's ability estimation and this will result possibly in cutting down testing time, as fewer items can be administered to evaluate the aptitude of the examinee. Self-adaptive testing (SAT), a variation of CAT, can also be used to determine the starting difficulty level of the CAT [15]. In SAT the examinee, rather than a computerised algorithm, chooses the difficulty of the next item to be presented [16].

In IRT based CAT systems the item selection process 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. The computer continuously re-evaluates the ability of the examinee until the accuracy of the estimate reaches a statistically acceptable level. If we consider the response in previous item as an interaction behaviour aspect and the fact that as the user gains experience the task is more automatic and will require fewer resources, in terms that less items will be needed to assess performance, then we can suggest that most IRT based CAT systems while modelling Knowledge of the domain in a sense they take into account the *User performance* and *User Cognitive Workload* variables described by Rothrock et al. [12] and the *Usage data* one described by Kobsa et al. [17].

Modelling *Background and hyperspace Experience* variable could result in simpler interfaces for the examinees that are familiar with the information space and more explanatory ones for the unfamiliar ones. This combined with the modelling of *Preferences* variable that can basically indicate interface elements (preferred colours, fonts, navigation ways etc.) allow examinees to focus on the assessment process. Further, more clear and self-explicit interfaces may result by taking into account the *Personal data* variable. For example, in examining gender, males and females appear to have different preferences in terms of media presentation, navigation support, attitudes, and information seeking strategies. Some examinees might feel frustrated or discouraged when they cannot work confidently with the assessment's interface or when the interface is not designed to suit their individuality. In turn, this will result in poorer performance, since more time will be needed to process information. This is an important issue as in most assessments time is an essential factor for measuring the overall performance.

*Individual traits* variable refers to stable features of the user such as personality factors, cognitive factors, and learning styles. Not much research exists, according to own

knowledge, on user personality factors. Richter and Salvendy [18] suggested that users perceive the computer software as having personality attributes similar to those of humans. Interfaces designed with introverted personality can result in most cases fastest performance for extroverted and introverted individuals. Moreover, modelling of cognitive or learning styles for CAT can result in more efficient systems. In interface design terms, with regards to cognitive style for example, a rigid structure should be provided for field dependent (FD) users as they need navigation and orientation support; while a more flexible (or customisable) interface should be made available for field independent (FI) users. Furthermore, studies have shown that FD are holistic and require external help while FI people are serialistic and possess internal cues to help them solve problems. FD learners are more likely to require externally defined goals and reinforcements while FI tend to develop self-defined goals and reinforcements [19]. These implications of style characteristics in CAT design could result in clear, explicit directions, maximum amount of guidance and extensive feedback to FD examinees, and on the other hand minimal guidance and direction and least feedback to FI examinees.

The modelling of the *Interests* variable for CAT systems can offer items closer to the long-term interests of each individual examinee. By knowing what interests a particular user, adaptive algorithms can be set to rule out certain items. However, this could be problematic in some cases, for example general knowledge assessments, as examinees will not face items that represent the whole range of the domain.

Kobsa et al. [17] suggest that besides “knowing what”, a user’s “knowing how” can also play an important role in adapting systems to user needs. In a CAT system modelling of *User skills and capabilities* variable can give examinees with different skills, when needed, help messages and explanations according to their familiarity with the domain presented. Further, in examinee population almost always included people with disabilities. If a mechanism exists to assist such individuals on demand disable people will feel less disadvantaged as they could easily take part in any examination process.

Furthermore, the modelling of *Groups of users* variable will be important in cases of group adaptive testing systems. Computer supported collaborative learning is currently at the focus of educational attention, however, according to our knowledge there are no examples of CAT systems for group evaluation so far.

The independent variables have an effect on the user indirectly, in terms that are not defining him/her as an individual. The most complicated variable to model is *User’s goal* as it change constantly from session to session and in many cases there are simultaneous goals within the same session. For example the main goal of taking a test is to pass it, however, simultaneously several goals exist, one for each item that is included in the test. In simple CAT systems modelling of User’s goal is not of a particular weight because it complicates the development of the test without any significant benefits for the examinee. However, in assessing non-cognitive skills modelling of User’s goal variable is important as examinees will always face items that closely match their own individual goals resulting in better individual performance.

A user is not tied to a particular hardware platform. S/he can work in one instance from a personal computer attached to a desk and on the other instance from a mobile device such as a Personal Digital Assistant (PDA). As a result dependent variables remain the same with regards to the student modelling. The independent variable of *Environment* cannot affect the content, yet it seriously affects the presentation mode. Systems can adapt to the user

platform by selecting appropriate ways in terms of bandwidth, media etc. for presenting the information. For educational courseware modelling of Environment variable may facilitate teaching and learning for disciplines related to outdoors activities such as zoology, botany, sailing etc. Nevertheless, it is quite unusual to model this variable for testing purposes as there are not many situations when an examinee will need to be assessed for the same subject using a PC and a PDA.

However, it is important to consider at this point the effort of Kinshuk and Lin [20] who explored how to improve learning process by adapting course content presentation to student learning styles in multi-platform environments such as PC and PDA. They develop a framework and a mechanism to comprehensively model student's learning styles and present the appropriate subject matter, including the content, format, media type, and so on, to suit individual student based on the Felder-Silverman Learning Style Theory.

Summarising, most IRT based CAT systems employ in their student model the Knowledge of the domain variable. This variable is closely associated with User performance, Usage data, and User Cognitive Workload. Besides these variables, modelling of Background-experience, Preferences, Personal data and Individual traits can produce well-organized CAT systems since fewer items will be needed to assess performance. Moreover, it could affect items' quality, since items can be more complex taking into account user characteristics. As a result, testing sessions would not be limited to measure performance but they can contribute to the learning process in terms of using evidence of examinee's performance gathered using complex tasks to support learning activities. In advanced CAT modelling of User's goal can also contribute to the test's quality. Modelling of Interests need careful implementation as it may result in false measurements, as examinees will be presented with items that always fall in their individual interests' domain and not in the whole knowledge domain examined with a CAT.

## 3. Future Trends

Currently, research in CAT moves beyond admission programs to address many aspects of measuring performance in education and training. This combined with new dynamic item generation tools and advances in profile scoring can facilitate computerised assessments that take into consideration more individual differences of the user than the mastery level, resulting in improved individual performance and more efficient test delivery.

Modelling multiple variables is important as users have complex characteristics that ultimate affect their performance. Student models must incorporate multiple variables of the user; dependent and independent. However, adding additional variables will not always increase the accuracy of the student model but will always increase its complexity and the requirements to collect additional user information [21]. Media elements are difficult to generate and are not flexible to automatic recombination as text is. Therefore, multimedia adaptation adds additional complexity and requires a greater implementation effort.

There are many research questions related to multiple variables modelling such as 'what is the proper type and number of variables to measure?'; 'which techniques should be used to model the variables?'; 'how we could modify the weights associated with different variables in order to represent the user more accurately?'; 'how can we maintain a balance between the number of variables, model complexity, and the accuracy of the model?' etc. In this paper several studies that are aiming to answer such issues were referenced.

However, besides research questions the key issue remains; taking into account individual characteristics in test design can benefit the users resulting in better performance. The essence of testing is to measure performance and consequently an elaborated student model for CAT that will include a large array of variables must be the way ahead. The type and number of variables that each CAT would comprise in the student model depend heavily on the subject matter and the way that the test is implemented.

## References


**1** van der Linden, W.J. & 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.

**2** Rudner, L. M. (1998). An online, interactive, Computer Adaptive Testing Tutorial. 11/98. http://EdRes.org/scripts/cat [viewed at 15/5/2007]

**3** 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.

**4** Williamson, D. M. Bejar, I. I. & Sax, A. (2004). Automated Tools for Subject Matter Expert Evaluation of Automated Scoring. Research and Development Report. March 2004, RR-04-14, ETS: Prinveton, NJ.

**5** Gouli, E., Papanikolaou K. and Grigoriadou M. (2002). Personalizing Assessment in Adaptive Educational Hypermedia Systems. *In P. De Bra, P. Brusilovsky, and R. Conejo (Eds.): AH 2002, LNCS 2347*, pp 153–163. Springer-Verlag Berlin Heidelberg.

**6** Weber, G. and Brusilovsky, P. (2001). ELM-ART: An Adaptive Versatile System for Web-based Instruction. *International Journal of Artificial Intelligence in Education,* 12, 351-384.

**7** Vassileva, J. (1996). A task-centered approach for user modeling in a hypermedia –based information system. *In Proceedings of the 4th International Conference on User Modeling*, Hyannis, MA, 115-120.

**8** Dowling, C.E., Kaluscha, R. (1995). Prerequisite Relationships for the Adaptive Assessment of Knowledge. In: J. Greer, (Ed.), *Proceedings of AI-ED'95, 7th World Conference on Artificial Intelligence in Education*, Washington, DC, 16-19 August 1995, AACE. pp. 43-50.

**9** Ríos, A., Millan, E., Trella, M., Perez-de-la-Cruz, Conejo, R. (1999). Internet Based Evaluation System. In: S.P. Lajoie, M. Vivet (Eds.), *Artificial Intelligence in Education*. Open Learning Environments: New Computational Technologies to Support Learning, Exploration and Collaboration. Volume 50 in Frontiers in Artificial Intelligence. IOS Press, Amsterdam, pp. 387-394.

**10** Brusilovsky, P. (1996). Methods and Techniques of Adaptive Hypermedia. Journal of user modeling and user-adapted interaction.6 (2-3), 87-129.

**11** Brusilovsky, P. (2001) Adaptive hypermedia. User Modeling and User Adapted Interaction, Ten Year Anniversary Issue (Alfred Kobsa, ed.) 11 (1/2), 87-110.

**12** Rothrock, L., Koubek, R., Fuchs F., Haas, M. & Salvendy, G. (2002). Review and Reappraisal of Adaptive Interfaces: Toward Biologically-Inspired Paradigms. Theoretical Issues in Ergonomic Science.

**13** Magoulas, G. D. & Dimakopoulos, D. N. (2005). Designing Personalised Information Access to Structured Information Spaces. Proceedings of the 1st International Workshop Workshop on New Technologies for Personalized Information Access.

**14** Triantafillou, E., Georgiadou, E., & Economides, A. (2007). The Role of User Model in CAT: Exploring Adaptive Variables. Journal of Technology, Instruction, Cognition and Learning, 5, 69-89.

**15** Frosini,G., Lazzerini,B., Marcelloni,F. (1998). Performing automatic exams. *Computers & Education* 31 (3), pp. 281-300.

**16** Rocklin, T., O' Donnell, A.M. (1987). Self-Adapted Testing: A Performance-Improving Variant of Computerized Adaptive Testing. *Journal of Educational Psychology*, 79 (3), pp. 315-319.

**17** Kobsa, A. Koenemann, J. & Pohl, W. 2001. Personalised hypermedia presentation techniques for improving online customer relationships, *The Knowledge Engineering Review*, Vol. 16:2, 111–155.

**18** Richter, L.A. & Salvendy, G. (1995). Effects of Personality and Task Strength on Performance in Computerized Tasks. *Ergonomics,* 38(2), 281-291.

**19** Witkin, H.A, Moore, C.A., Goodenough, D.R., Cox, P.W. (1977). Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research*, 47(1), 1-64.
**20** Kinshuk & Lin T. (2004). Application of learning styles adaptivity in mobile learning environments. *Third Pan Commonwealth Forum on Open Learning*, 4-8 July 2004, Dunedin, New Zealand
**21** Carver, C., Hill, M. & Pooch, U. (1999). Third Generation Adaptive Hypermedia Systems. WebNet 99, Honolulu, Hawaii, 1999.