Triantafillou, E., Georgiadou, E., & Economides, A.A.: **Applying Adaptive Variables in Computerised Adaptive Testing.** *Australasian Journal of Educational Technology*, *AJET*, Vol. 23, No. 3 2007

Applying Adaptive Variables in Computerised Adaptive Testing

Evangelos Triantafillou
Center of Educational Technology
Dodekanisou 21, Thessaloniki 55131, Greece
vtrianta@edutech.gr

Elissavet Georgiadou
Center of Educational Technology
Karaoli 46, Thessaloniki 57001, Greece
elisag@otenet.gr

Anastasios A. Economides
University of Macedonia, Department of Computer Networks
Egnatia 156,Thessaloniki 54006, Greece
economid@uom.gr

ABSTRACT

Current research in Computerised Adaptive Testing (CAT) 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. Therefore, it is important to extend CAT's functionality to include more variables in its student model that define the examinee as an individual beyond the mastery level, for improved performance and more efficient test delivery. This paper is aiming to look at different variables that can prompt adaptation and then discusses their potential use to a hypothetical student model for CAT. The objective of this effort is to provide researchers, designers, and developers of CAT a perspective to exploit research outcomes from the area of personalised hypermedia applications.

Keywords

CAT, Student model, Adaptive variables, User modelling, Computerised assessment

1. Introduction

Due to the advancements in communication and information technology, the popularity of computer-based testing has increased in recent years. Computer delivery of tests became feasible for licensure, certification and admission. Moreover, computers can be used to increase the statistical accuracy of test scores using computerized adaptive testing (CAT). 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 (van der Linden & Glas, 2003). 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 is presented. 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. The vast majority of CAT systems rely on Item Response Theory as the underlying model (Lord, 1980; Wainer, 1990). However, the Decision Theory provides an alternative underlying model for sequential testing (Rudner, 2002), and the Knowledge Space Theory (Doignon & Falmagne, 1985) is another basis of development for small-scale construction of adaptive tests.

Regardless of some disadvantages reported in the literature –for example, high cost of development, item calibration, item exposure control (Eggen, 2001; Boyd, 2003), the effect of a flawed item (Abdullah, 2003), or the use of CAT for summative assessment (Lilley & Barker, 2002, 2003)– 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 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 (Rudner, 1998; Linacre, 2000).

Since the mid 80's when the first CAT systems became operational ([Armed Services Vocational Aptitude Battery for US Department of Defence account], (van der Linden & Glas, 2003) using adaptive techniques to administer multiple-choice items, much research and many technical challenges make possible new assessment tools. Currently, analysis of the results can go deeper than just calculate the right and wrong answers. Contemporary research in profile scoring involves the design and generation of enhanced score reports focus on the interpretation of score report components, feedback about skills (e.g. most promising skills for the student to work on), and educational advice, i.e. suggestions for improvement (Gitomer & Bennet, 2002). Moreover, as research advances in the field new item generation tools that will further increase the efficiency of test creation process appear (e.g. Higgins, Futagi & Deane, 2005; Guzmán, Conejo & García-Hervás, 2005; Lilley, Barker, & Britton, 2004; Gonçalves, Aluísio, de Oliveira, & Oliveira, 2004; Bejar, Lawless, Morley, Wagner, Bennett, & Revuelta, 2002).

Most CAT systems include a student model. Paiva, Self and Hartley (1995, page 509) define a student model as "representations of some characteristics and attitudes of the learners, which are useful for achieving the adequate and individualised interaction established between computational environments and students". Replacing the term student by user this definition is also applicable to a user model. A user model is constituted by descriptions of what is considered relevant about the actual knowledge and/or aptitudes of a user, providing information for the system environment to adapt itself to the individual user (Koch, 2000).

Student model variables 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, student 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 (Williamson, Bejar & Sax, 2004). Therefore, it is important to extend CAT's functionality to include more variables in its student 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 (Gouli, Papanikolaou & Grigoriadou, 2002), ELMART (Weber & Brusilowsky, 2001), DCG (Vassileva, 1996)). Moreover CAT is used as a student modelling technique in

Intelligent Tutoring Systems (Dowling & Kaluscha, 1995; Ríos, Millan, Trella, Perez-de-la-Cruz & Conejo, 1999).

This paper is aiming to look at different variables that can prompt adaptation and then discuss their potential use to a hypothetical student model for CAT. The objective of this effort is to provide researchers, designers, and developers of CAT a perspective to exploit research outcomes from the research area of personalised hypermedia applications. Next, the paper will proceed to examine the different variables that can prompt adaptation and following that it will discuss their potential use to a hypothetical student model for CAT.

2. Adaptive variables

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. Brusilovsky (1996) identifies the following features which are used by existing adaptive hypermedia systems: users' goals, knowledge, background and hyperspace experience, and preferences. Furthermore, Brusilovsky (2001) adds two more variables to this list: the user's interests and individual traits. Moreover, the author indicates the importance of adaptation in user's environment (user's location, user's platform).

Kobsa, Koenemann & Pohl (2001) in reviewing techniques for personalised hypermedia presentation, they describe the following categories of user data that have been the basis for adaptation in a number of systems developed since 2001: a) demographic data, b) user's knowledge, c) user's skills and capabilities, d) user's interests and preferences, and e) user's goals and plans. In addition, they underline the significance of the computer usage (interaction behaviour, current task, and interaction history) and the physical environment (locale, software and hardware) that can be taken into account when adapting hypermedia pages to the needs of the current user.

Moreover, Rothrock, Koubek, Fuchs, Haas & Salvendy (2002) in reviewing adaptive interfaces argue that "an adaptive interface autonomously adapts its displays and available actions to current goals and abilities of the user by monitoring user status, the system task, and the current situation" (p. 9). They identify the following variables calling for adaptation: 1. user performance, 2. user goals, 3. user workload, 4. user situation awareness, 5. user knowledge, 6. groups of users, 7. user personality and cognitive style, 8. task variables (situation variables and system variables).

Further, Magoulas and Dimakopoulos (2005) in exploring the dimensions of individual differences that should be included in a student model specification to meet personalisation services requirements and create personalised information access identify the following nine dimensions of a user data model for structured information spaces: (i) personal data, such as gender, age, language, and culture, (ii) cognitive or learning styles, (iii) device information (the hardware used for access), (iv) context-related data capture the physical environment from where the user is accessing the information and can be used to infer the user's goals, (v) user history data capture user past interaction with the system and can be used under the assumption that users' future behaviour will be almost similar to their past behaviours, (vi) user preferences and interests, (vii) goal-related data, (viii) system experience indicates the knowledge of that particular user about the information space, and (ix) domain expertise relates to the existing level of understanding of a particular user on the domain knowledge.

A list of the different variables acknowledged from the researchers above are presented below in Table 1.

Table 1. Adaptive variables identified in the literature from 1996 to 2005

	Brusilovsky, 1996	Brusilovsky, 2001	Kobsa, Koenemann & Pohl, 2001	Rothrock, Koubek, Fuchs, Haas & Salvendy, 2002	Magoulas & Demakopou-los, 2005

Users' goals	✓	✓	✓	✓	✓
Knowledge of the domain	✓	✓	✓	✓	✓
Background and hyperspace experience	✓	✓			✓
Preferences	✓	✓	✓		✓
User's interests		✓	✓		✓
Individual traits (Cognitive or learning style, user personality)		✓		✓	✓
Environment (location, locale, software, hardware) (User Situation Awareness)		✓	✓	✓	✓
Personal data			✓		✓
User skills and capabilities			✓		
User performance				✓	
Usage data (user history)			✓		✓
User Cognitive Workload				✓	
Groups of Users				✓	

As shown in Table 1 thirteen different adaptive variables have been identified from 1996 to 2005 in total. Some of the variables are under the same or similar terminology. For example, Brusilovsky (2001) is arguing about Individual traits that include user personality factors, cognitive factors and learning styles, while Rothrock et al. (2002) refer to user personality and cognitive style and Magoulas and Dimakopoulos (2005) argue for learning or cognitive styles. The investigation of the thirteen adaptive variables included in Table 1 guide the authors of this paper to classify them under two broad categories: *user dependent* and *user independent.* Next the paper will proceed to examine the variables included in these two broad categories.

3. User dependent and user independent variables

The user dependent variables are those directly related to the user and strictly define him/her as an individual. These variables generally concern with individual user characteristics such as the user's knowledge state, the user's background, the user's demographics, user's mental model etc. In particular, from the research reviewed in this paper the user dependent variables are identified as follows: (a) Knowledge of the domain, (b) Background and Hyperspace Experience, (c) Preferences, (d) User interests, (e) Individual traits, (f) Personal data, (g) User skills and capabilities, (h) User Performance, (i) Usage data, (j) User Cognitive Workload, and (k) Groups of users.

The user independent variables generally affect the user indirectly and are related mainly to the context of a user's work with a hypermedia application rather than with the user as an individual. In particular, from the research reviewed in this paper the user independent variables are identified as follows: (a) User's goal and (b) Environment.

3.1 Dependent variables

a) Knowledge of the domain

User's *knowledge of the domain* is a variable for a particular user. This means that an adaptive hypermedia system which relies on user's knowledge has to recognize the changes in the user's knowledge state and update the student model accordingly. There are many established techniques for modelling student knowledge in relation to domain or course knowledge (for a detailed account see Abdullah, 2003). However, user's knowledge of the subject is most often represented by an *overlay model* which is based on the structural model of the subject domain. Generally, the structural domain model is represented as a network of domain *concepts*. The concepts are related with each other thus forming a kind of semantic network which represents the structure of the subject domain. These concepts can be named differently in different systems - topics, knowledge elements, objects, learning outcomes - but in all the cases they are just elementary pieces of knowledge for the given domain. In most of the existing CAT systems user's knowledge of the domain is the basic variable in their student model since item selection adapts to the ability level of individual examinees.

b) Background and Hyperspace Experience

Background and Hyperspace Experience in the given hyperspace are two features of the user which are similar to user's knowledge of the subject but functionally differ from it. *User's Background* describes all the information related to the user's previous experience outside the subject of the hypermedia system, which is relevant enough to be considered. This includes the user's profession, experience of work in related areas, as well as the user's point of view and perspective. On the other hand, *User's Experience* in the given hyperspace describes the familiarity of the user with the structure of the hyperspace and how easy can the user navigate in it. Sometimes, the user who is generally quite familiar with the subject itself is not familiar at all with the hyperspace structure. Vice versa, the user can be quite familiar with the structure of the hyperspace without deep knowledge of the subject. *Background* and *Experience* are usually modelled using stereotype model (e.g. experience stereotype, background stereotype for profession).

c) Preferences

Preferences are user features that relate to the user's likes and dislikes. This variable describes that a user can prefer some types of nodes and links to others or some parts of a page over others. Moreover, preferences can indicate interface elements such as preferred colours, fonts, navigation ways, etc. User preferences are not assumed by the system; instead the user has to notify the system, directly or indirectly by providing feedback. Usually, the user through checklists can select preferred interface elements. Once the preferences are determined the system generalises and applies them for adaptation in new contexts.

d) Interests

The *Interests* variable is in a way similar to *Preferences*, but it is not the same as it refers mostly to web-based information retrieval systems. It concerns with the user's long-term interests, that are used in parallel with the user's short-term search goal in order to improve the information filtering and recommendations. Interests can be modelled through navigation monitoring, for example, by observing which links the user visits more often.

e) Individual traits

User's *Individual traits* is a group name for user features that together define a user as an individual. Examples are user personality factors (e.g. introvert/extravert), cognitive factors, and learning styles. Like user background, individual traits are stable features of a user that either cannot be changed at all, or can be changed only over a long period of time. Unlike user background however, individual traits are traditionally extracted not by a simple interview, but by specially designed psychological tests.

User Personality
Murray and Bevan (1985) argue that human-computer interaction would improve if computers were assigned personalities, as the best way for a human to interact with a computer should closely resemble the interaction between two humans. On that view, Richter and Salvendy (1995) compared the

performance of introverted and extroverted users using “extroverted” and “introverted” interfaces. The extroverted interface they design had more words, more “fun” pictures, more sounds, bold fonts and exclamation marks than the introverted interface. The subjects used in their empirical study were classified as introverted or extroverted according to the Eysenck Personality Inventory score. The main findings from this study suggest that users perceive the computer software as having personality attributes similar to those of humans and also using software designed with introverted personality results in general fastest performance for both individuals with extroverted and introverted personalities (Rothrock et al., 2002).

Cognitive Style- Learning Style
Cognitive or learning styles refer to a user’s information processing behaviour and have an effect on user’s skills and abilities, such as preferred modes of perceiving and processing information, and problem solving. They can be used to personalise the presentation and organisation of the content, the navigation support, and search results (Magoulas & Dimakopoulos, 2005).

Cognitive style is the way individuals organize and structure information from their surroundings and its role is critically important. It is associated with student success in any learning situation. Cognitive style is usually described as a personality dimension, which influences attitudes, values, and social interaction. It also refers to the preferred way an individual processes information. There are many different definitions of cognitive styles as different researchers emphasize on different aspects. However, Witkin’s definition of field dependent (FD) and field independent (FI) is the most well known division of cognitive styles and is more relevant to hypermedia research than others (Witkin, Moore, Goodenough & Cox, 1977). Many experimental studies have showed the impact of field dependence /independence on the learning process and academic achievement and identified a number of relationships between cognitive style and learning, including the ability to learn from social environments, types of educational reinforcement needed to enhance learning, amount of structure preferred in an educational environment (Summerville, 1999, Ford & Chen, 2000, Weller, Repman & Rooze, 1994, Triantafillou, Demetriadis, Pombortsis & Georgiadou, 2004).

Learning style is an important issue that affects the learning process and therefore the outcome. Many definitions and interpretations of learning styles appeared in literature the past decades (Bedford, 2006). However, in general terms, learning styles is the individual preferences for how to learn (Sternberg, 1997). When designing instructional material, it is imperative to accommodate elements that reflect individual differences in learning as every learner has a unique way of learning. Papanikolaou and Grigoriadou (2004) suggest that important decisions underlying the incorporation of learning style characteristics in educational adaptive hypermedia systems demand the synergy of computer science and instructional science, such as: (i) the selection of proper categorizations, which are suitable for the task of adaptation, (ii) the design of adaptation, including the selection of appropriate adaptation technologies for different learning style categorizations and of apposite techniques for their implementation, (iii) the design of the knowledge representation of such a system in terms of the domain and the learner model, (iv) the development of intelligent techniques for the dynamic adaptation of the system and the diagnosis process of learners’ learning style including also the selection of specific measurements of learners’ observable behaviour, which are considered indicative of learners’ learning style and studying attitude.

f) Personal data

Personal data, such as gender, age, language, and culture should be taken into account when designing adaptive educational interfaces to optimise learner’s potential to benefit from the system’s design in terms of knowledge acquisition. For example males and females appear to have different preferences in terms of media presentation, navigation support, attitudes, and information seeking strategies (Magoulas & Dimakopoulos, 2005).

An empirical study into gender differences in collaborative web searching reveals that males formulate queries comprising fewer keywords, spent less time on individual pages, click more hypertext links per minute and in general are more active while online than females (Large, Beheshti & Rahman, 2001). Moreover, research suggests that males significantly outperform females in navigating virtual environments. Tan, Robertson and Czerwinski (2001) suggest that special navigation techniques when combined with a large display and wide field of view appeared to reduce that gender bias.

Kobsa, Koenemann and Pohl (2001) extend the term personal data to demographic data about the user which are “objective facts” like the following: record data (e.g., name, address, phone number), geographic data (area code, city, state, country), user characteristics (e.g., age, sex, education, disposable income), psychographic data (data indicating lifestyle), customer qualifying data (e.g., frequency of product/service usage), registration for information offerings, participation in raffles and so on as their research is focus on online customer relationships.

g) User skills and capabilities

Kobsa et al. (2001) suggest that besides “knowing what”, a user’s “knowing how” can also play an important role in adapting systems to user needs. Adaptive help systems are typical representatives of this approach. For instance, the Unix Consultant (Chin, 1989) tailors its help messages and explanations to the user’s familiarity with UNIX commands. Peter and Rösner (1994) tailor repair instructions to the user’s familiarity with the operations involved in the suggested repair plan. Küpper and Kobsa (1999) go further and distinguish between the actions a user is familiar with and the actions he or she is actually able to perform. It is possible that a user knows how to do something but is not able to perform the action due to lack of required permissions or to some physical handicap. Therefore, the tourist information system AVANTI (Fink, Kobsa, & Nill, 1998), which takes into account the needs of different kinds of disabled people (wheelchair-bound, motor-impaired and vision-impaired), recommends only actions that these users are actually able to perform.

This variable is important as people with disabilities often find difficulty to use computer-based systems, since the vast majority of these systems have no design considerations for them. These different users have varying needs regarding content and presentation of the information. For example, information for the blind should be presented in audio mode and a Braille display and speech synthesiser is needed so as to interact with the learning material; information for the deaf should never be in audio format.

h) User Performance

Rothrock et al. (2002) consider adaptation useful in not only the correction, but also in the prevention of poor performance. The user’s performance is mainly defined by his error rate in performing a task, as well as the time required to perform the task. If there are concurrent tasks, they must be assessed separately. Examples of inputs to infer the user’s performance include computer data entry speed, latency of response to a verbal request, reaction time to capture a simple target, and tracking deviation. User performance is difficult to measure as it is complicated to specify accurately all user goals and reactions. For example, highly cognitive tasks, like decision-making, are very difficult to measure, because the performance outcome does not necessarily reflect the complexity of the mental process.

i) Usage data

Kobsa et al. (2001) suggest that usage data can be used by the system to adapt to user preferences, habits and levels of expertise. Usage data may be directly observed and recorded, or acquired by analysing observable data (e.g. what pages and files have been requested from the server, mouse clicks and movements). In addition to interaction behaviour, the usage context may also be considered as a source for adaptation. Among the relevant items are the current task and the interaction history. Magoulas and Dimakopoulos (2005) refer to *User history data* that capture user past interaction with the system, e.g. visited pages that contain pointers to specific keywords, or browsing habits, and can be used under the assumption that users’ future behaviour will be almost similar to their past behaviours.

j) User Cognitive Workload

Rothrock et al. (2002) consider *User Cognitive Workload* as another variable that calls for adaptation. The class of input variables associated with workload is important because it provides a direct link to user performance. The predominant theory used to infer user workload in multi-task processing is the multiple-resource theory (Wickens, 1992). In the multiple-resource theory, the user has multiple pools of resources at his/her disposal from which to perceive, decide, and act. A limitation of the multiple-resource model is that it does not take into account the learning that takes place as the user gains experience. Thus, as the user is more experienced, the task is more automatic and will require fewer resources. A predictive workload measure can be calculated from models using time-line analysis. The objective of these models is to calculate the global workload. The global workload is the sum of the measurable workloads for each task spanning across all time intervals, which is then weighted by the theoretical overlap between human resources. If the workload calculated is greater than 100%, the task can be reallocated or postponed.

k) Groups of users

Computer Supported Collaborative Learning (CSCL) and groupware applications are at the focus of educational research lately. Group models are important for collaborative work, since a standard group model should serve as a starting point for interaction for the new member that enters a group (Brusilovsky, 1996). While the new user starts to interact with the system, the user profile can be formed including those characteristics that are in common with, and are different from, the group profile. To build the group profile, information from users can be acquired using similar techniques with those used for the individual student model: stereotypes, interviews, monitoring users' behaviour. However, these techniques take into account adaptive variables such as Individual traits in order to select users for the group construction. The group profile is quite important for web-based systems as the web facilitates collaborative activities.

3.2 Independent variables

a) User's goal

The most changeable user feature that activates adaptation is the *User's goal*. It is related to the context of a user's work with a hypermedia application rather than with the user as an individual. It informs what the user wants to accomplish by using the application. For example, in information retrieval systems, a user's goal is a search goal; in educational systems is a learning goal; in testing systems might be a problem-solving one. User's goal is not firm but it constantly changes from session to session and frequently changes several times within a session. However, there can also be simultaneous goals i.e. simple, multiple, concurrent. General or high level goals are more stable than local or low-level goals. For example, in educational systems the learning goal is a high-level goal, while the problem solving goal is a low-level goal which changes from one educational problem to another several times within a session.

b) Environment

The importance of adaptation to user's *Environment* is acknowledged by all researchers examined above. It is a new kind of adaptation that was brought by Web-based systems. Users of web-based systems can work irrespective of time and location using different equipment and as a result adaptation to the user's environment can result in better use of the system and yet better performance. Systems can adapt to the user platform, such as hardware, software and network bandwidth. Such adaptation usually involves selecting the type of material and media to present the content, for example, still image vs. movie, text vs. sound (Joerding, 1999).

Kobsa et al. (2001) suggest that Web usage may be influenced by both the software (browser version and platform, availability of plug-ins etc.) and the hardware (bandwidth, processing speed, input device etc.) of the individual user, and by the characteristics of the user's current locale (current location and usage locale: the noise level and brightness of the surroundings, and information about places and objects in the immediate environment).

Magoulas and Dimakopoulos (2005) use the terms *Device information* and *Context-related data* to describe the environment variable. *Device information* concerns the hardware used for access and affects personalisation services in terms of screen layout and bandwidth limitations, and *Context-related data* capture the physical environment from where the user is accessing the information and can be used to infer the user's goals.

Moreover, changes in the environment or changes in the system can call for an adaptation of the interface. Rothrock et al. (2002) use the term task variables that include *situation* and *system variables.* Situation variables that influence user abilities as well as task requirements include: time pressure, location in space and presence and location of targets; situation in time; weather conditions; visibility; and vibration and noise. Like the situation variables, some changes of the task represent critical system events. Moreover, variables that cause the system changes (e.g., loss of engine power and failures) are often interdependent with the user and the situation variables. Environment variable is closely associated with *User Situation Awareness*, also suggested by Rothrock et al. (2002). Situation awareness is the perception of the elements in the environment within a volume of time, and the comprehension of their meaning, and the projection of their status in the near future (Endsley, 1997).

Current Information and Communication Technologies developments focus on mobile information technology that allows for mobility in the physical space. Given the user and the information is connected to a network this technology facilitates accessibility of information from any point in the physical space. For communication purposes the user employs different devices that have, however, specific characteristics and limitations in terms of bandwidth and information presentation. For mobile information technology the particular challenge for adaptivity is the support of users at different locations. To achieve this, mobile information technology can be combined with technologies to identify the users' working environment and his or her position in the physical space such as infrared or General Positioning Systems (GPS) (Oppermann & Specht, 1999).

4. Discussion

In section 3 the paper examines different adaptive variables acknowledged by researchers in the area of personalized adaptive systems. This section will discuss whether the application of these variables to a student model of a CAT system will be of a benefit to such system in terms of increasing its efficiency.

A CAT in order to be more efficient than a fixed-length computerised test initially assesses 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 (Frosini, Lazzerini, Marcelloni, 1998). In SAT the examinee, rather than a computerised algorithm, chooses the difficulty of the next item to be presented (Rocklin & O' Donnell, 1987).

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. (2002) and the *Usage data* one described by Kobsa et al. (2001).

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 (1995) suggest 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 (Witkin et al. 1977). 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. (2001) 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 (2004) who explore 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.

5. Conclusion

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. Moreover, graphical modelling extents the IRT-based CAT inferential framework to accommodate richer tasks and more complex student models (Almond & Mislevy, 1999).

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 (Carver, Hill & Pooch, 1999). 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 and several studies that attempted to address such questions are referenced in this paper. Nevertheless, the key issue is that 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. Mislevy, Steinberg and Almond (1999, p.7) argue that "the factors that determine the number and the nature of the student model variables in a particular application are the conception of competence in the domain and the intended use of the assessment".

The scope of this paper is to review and examine the different variables that can prompt adaptation and discuss their potential use to a hypothetical student model for CAT in order to provide researchers, designers, and developers of CAT a perspective to exploit research outcomes from the area of personalised hypermedia applications.

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

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

Almond, R. G., & Mislevy, R. J. (1999). Graphical models and computerized adaptive testing. *Applied Psychological Measurement*, *23*(3), 223–237.

Bedford, T. (2006). *Learning styles: A review of the English language literature*. In R.R. Sims & S.J. Sims (eds.), Learning styles and learning: A key to meeting the accountability demands in education. New York: Nova Science Publishers Inc.

Bejar, I. I., Lawless, R., Morley, M., Wagner, M., Bennett, R., & Revuelta, J. (2002). *A feasibility study of on-the-fly item generation in adaptive testing* (ETS RR-02-23). Princeton, NJ: ETS.

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.

Brusilovsky, P. (1996). Methods and Techniques of Adaptive Hypermedia. *Journal of user modeling and user-adapted interaction*.6 (2-3), 87-129. http://www.sis.pitt.edu/~peterb/papers/UMUAI96.pdf

Brusilovsky, P. (2001) Adaptive hypermedia. *User Modeling and User Adapted Interaction*, Ten Year Anniversary Issue (Alfred Kobsa, ed.) 11 (1/2), 87-110. http://www2.sis.pitt.edu/~peterb/papers/brusilovsky-umuai-2001.pdf

Carver, C., Hill, M. & Pooch, U. (1999). Third Generation Adaptive Hypermedia Systems. *WebNet 99*, Honolulu, Hawaii, 1999.

Chin, D. N. (1989). KNOME: Modeling What the User Knows in UC. In Kobsa, A. & Wahlster, W. (eds.) *User Models in Dialog Systems*, 74-107. Springer-Verlag: Berlin.

Doignon, J. P. & Falmagne, J. C. (1985). Spaces for the Assessment of Knowledge. *International Journal of Man-Machine Studies* 23 (2), pp.175-196.

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.

Eggen, T. J. H. M. (2001). Overexposure and underexposure of items in computerized adaptive testing. *Measurement and Research Department Reports* 2001-1, Citogroep Arnhem. http://www.cito.nl/share/poc/reports/Report01-01.pdf

Endsley, M. R. (1997). Automation and situation awareness. In *Automation and Human-Performance, Theory and Applications*, In R. Parasuraman and M. Mouloua (Eds.). Mahwah, N.J.: Lawrence Erlbaum.

Fink, J., Kobsa, A. & Nill, A. (1998). “Adaptable and adaptive information provision for all users, including disabled and elderly people” *The New Review of Hypermedia and Multimedia* 4 163–188.

Ford, N. & Chen, S. (2000). Individual Differences, Hypermedia Navigation, and Learning: An Empirical Study. *Journal of Educational Multimedia and Hypermedia*, 9 (4), 281-311.

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

Gitomer, D. H. & Bennett, R. E. (2002) Unmasking Constructs Through New Technology, Measurement Theory, and Cognitive Science. *Research Memorandum*, February 2002, RM-02-01, Educational Testing Service, Princeton, NJ.

Gonçalves, J. P., Aluísio, S. M., de Oliveira, L. H. M., & Oliveira, O. N. (2004). A Learning Environment for English for Academic Purposes Based on Adaptive Tests and Task-Based Systems. *Lecture Notes in Computer Science, 3220*, 1-11.

Gouli, E., Papanikolaou, K. & 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.

Guzmán, E., Conejo, R. & García-Hervás, E. (2005). An Authoring Environment for Adaptive Testing. *Educational Technology & Society*, 8 (3), 66-76.

Higgins, D., Futagi, Y. & Deane, P. (2005). *Multilingual Generalization of the Model Creator Software for Math Item Generation ETS Research and Development* (ETS RR-05-02) Princeton, NJ: ETS.

Joerding, T. (1999). A temporary user modeling approach for adaptive shopping on the Web. *Proceedings of Workshop on Adaptive Systems and User Modelling on World Wide Web at WWW '99 Conference*, Toronto, Canada, 75-79.

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. http://www.col.org/pcf3/Papers/PDFs/Kinshuk_Lin_1.pdf

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.

Koch, N. (2000). Software Engineering for Adaptive Hypermedia Systems. *Unpublished PhD thesis*, Mathematics and Informatics Department, University of Munich. http://www.pst.informatik.uni-muenchen.de/~kochn/PhDThesisNoraKoch.pdf

Küpper, D. & Kobsa, A. (1999). “User-tailored plan generation” in J Kay (ed.) *UM99 User Modeling: Proceedings of the Seventh International Conference* Springer-Verlag 45–54.
Large, A., Beheshti, J. & Rahman, T. (2001). Gender differences in collaborative Web searching behavior: an elementary school study. *Information Processing & Management*, 38(3), 427-443. http://wotan.liu.edu/dois/data/Articles/juljuljiqy:2001:v:38:i:3:p:427-443.html
Lilley, M. & Barker, T. (2002). The Development and Evaluation of a Computer-Adaptive Testing Application for English Language, *6th Computer Assisted Assessment Conference*, July 2002, Loughborough. http://www.caaconference.com/pastConferences/2002/proceedings/lilley_m1.pdf
Lilley, M. & Barker, T. (2003). An Evaluation of a Computer Adaptive Test in a UK University Context, *7th Computer Assisted Assessment Conference*, 8th and 9th July 2003, Loughborough. http://www.caaconference.com/pastConferences/2003/procedings/lilley.pdf
Lilley, M., Barker T. & Britton, C. (2004). The development and evaluation of a software prototype for computer-adaptive testing. *Computers & Education, 43*, 109-123.
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. http://www.rasch.org/memo69.pdf
Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems.* Lawrence Erlbaum Associates, Publishers, New Jersey.
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*.
Mislevy, R. J., Steinberg, L. S. & Almond, R. G. (1999). On the Roles of Task Model Variables in Assessment Design, *CSE Technical Report 500*, January 1999, Educational Testing Service, Princeton, New Jersey. http://eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED431804
Murray, D. & Bevan, N. (1985). The social psychology of computer conversations. In B. Shackel (Ed.) *Human-Computer Interaction – INTERACT 84*. New York: Elsevier Science.
Oppermann, R. & Specht, M. (1999). Adaptive information for nomadic activity: a process oriented approach. *Proceedings of the Software-Ergonomie ’99*, Stuttgart: Teubner, 255 –264. http://fit.fraunhofer.de/~oppi/publications/SE99.AdaptiveActivitySuppor.pdf
Paiva, A., Self, J. & Hartley, R. (1995). Externalizing Learner Models. In J Greer (ed.) *Proceedings of AIED95*, AACE publication, 509-519.
Papanikolaou, K.A. & Grigoriadou, M. (2004). Accommodating learning style characteristics in Adaptive Educational Hypermedia Systems. *Proceedings of the AH 2004 Workshop “Individual Differences in Adaptive Hypermedia”*.
Peter, G. & Rösner, D. (1994). “User-model-driven generation of instructions” *User Modeling and User-Adapted Interaction* **3**(4) 289–319.
Richter, L.A. & Salvendy, G. (1995). Effects of Personality and Task Strength on Performance in Computerized Tasks. *Ergonomics,* 38(2), 281-291.
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.
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.
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.*
Rudner, L. M. (1998). *An online, interactive, Computer Adaptive Testing Tutorial*. 11/98. http://EdRes.org/scripts/cat [viewed at 15/5/2007]
Rudner, L. M. (2002). *An Examination of Decision-Theory Adaptive Testing Procedures*. Paper presented at the annual meeting of the American Educational Research Association New Orleans, LA April 1-5, 2002.
Sternberg, R. J. (1997). *Thinking style*s. New York: Cambridge University Press.
Summerville, J. (1999). Role of awareness of cognitive style in hypermedia. *International Journal of Educational Technology*, 1(1). http://www.ed.uiuc.edu/ijet/v1n1/summerville/

Tan, D. S., Robertson, G. G. & Czerwinski, M. (2001). Exploring 3D Navigation: Combining Speed-Coupled Flying with Orbiting. In *Proceedings of CHI 2001, Human Factors in Computing Systems,* (Seattle, WA April 1-6, 2001), ACM press, 418-424.

Triantafillou, E., Demetriadis S., Pombortsis, A. & Georgiadou, E. (2004). The value of adaptivity based on cognitive style: an empirical study. *British Journal of Educational Technology,* 35 (1), 95-106.

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.

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.

Wainer, H. (1990). *Computerized Adaptive Testing: A Primer.* Lawrence Erlbaum Associates, Publishers, New Jersey.

Weber, G. & Brusilovsky, P. (2001). ELM-ART: An Adaptive Versatile System for Web-based Instruction. *International Journal of Artificial Intelligence in Education,* 12, 351-384. http://www2.sis.pitt.edu/~peterb/papers/JAIEDFinal.pdf

Weller, H. G., Repman, J. & Rooze, G. E. (1994). The relationship of learning behaviour, and cognitive styles in hypermedia-based instruction: Implications for design of HBI. *Computers in the Schools*, 10, 401-420.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance* (2nd ed.). New York: Harper Collins.

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.

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.