

## **Computer Adaptive Assessment and its use in the development of a student model for blended learning.**

TREVOR BARKER University of Hertfordshire  
MARIANA LILLEY University of Hertfordshire  
CAROL BRITTON University of Hertfordshire

**Abstract** This paper presents an overview of our work on the development and testing of an automated feedback tool based on Computer-Adaptive Testing. Computer-adaptive tests (CATs) are software applications that adapt the presentation of test questions to the learner's proficiency level, so that those performing well are given more difficult questions and vice versa. In this paper, we present and describe the development of the models used in a feedback tool based on this approach. The model includes a proficiency level estimation based on Item Response Theory and also a questions' database. The questions in the database are classified according to topic area and difficulty level. The difficulty level is initially set by expert evaluation based upon Bloom's taxonomy and adapted according to students' performance over time. The output from our adaptive test is a continuously updated student model that estimates proficiency in each of the domain areas covered in the test, relating not only to performance, but also to cognitive ability, based on Bloom's levels. Earlier work has shown that the approach we adopt is reliable and fair to students and provides useful and important measures of ability. Potentially these measures may be used, not only in formative and summative assessment, but also to help in the delivery of learning or remedial activities based on individual ability. We describe our student model based on adaptive testing and show how it was used to provide automated feedback for students in a summative assessment context. The evaluation of our feedback tool by groups of learners and teachers suggested that our approach was a valid one, capable of providing useful advice for individual development. A survey of staff attitude supported this view. The results of these evaluations are presented in this paper. In the concluding section of the paper we suggest ways that the student profiles created by our method are likely to be useful in a variety of learning contexts.

### **1 Introduction**

In Higher Education today, increasing reliance is being placed upon the use of online learning systems. Often these are used to manage learning, present information and test learners in an entirely undifferentiated way, all users

having exactly the same view of the system. With the development of increasingly large and complex computer applications and greater diversity in learner groups, consideration of individual differences has become an important issue in designing usable and useful applications.

User modelling is a technique that is often employed to this end, allowing users to perform tasks and interact with systems differentially, depending on some feature of their personalities, abilities, preferences or performance. Student models are a sub set of user models, in the domain of teaching and learning. User models are models that systems have of users that reside inside a computational environment (Fischer, 2001). User models are used to configure the presentation of information depending on the specific requirements of individual users. To be able to adapt itself to an individual user, the system has to (i) be aware of the domain, individual users and their knowledge and (ii) monitor usage or progress within the system to keep the model updated (Kavcic, 2000).

Global description models are an attempt to configure the presentation of information and performance based upon assumptions about individual characteristics. These may include personality differences, cognitive skills, motivation, reasoning abilities, perceptual speed, memory, language skills, listening skills, visual skills, intelligence, age, gender and similar personal characteristics. Barker and colleagues (2002) used language ability and other cognitive skills in order to configure their student model. Such global description models are more difficult to adapt automatically as it is often difficult to link psychological characteristics directly with a presentation model. For this reason, global description models often benefit from user collaboration in order to adapt them (Barker et al., 2002). User collaboration, although useful, was found to be slow and inefficient due to the complexity of the decisions that had to be made within the domain.

Intelligent software applications that adapt to their users based on models other than cooperation have been gaining rapidly in importance within the field of human-computer interaction. It is hoped that such applications would present similar qualities to those offered by a cooperative model – such as the ability to provide good and useful information – but in a more efficient way. Examples of such adaptive interfaces include systems that help users to filter web query results (Vouros, 2001), recommendation systems that help users to make choices (Langley et al., 1999), intelligent tutoring system applications that adapt to the knowledge of individual users (Brusilovsky, 2001). Computer-adaptive tests are a further example of a software application within the intelligent interfaces domain (Jettmar & Nass, 2002), and this type of adaptive software is the focus of this paper.

## 2 Computer Adaptive Tests

Computer adaptive tests (CATs) are computer-assisted assessment applications that differ from traditional computer-based tests (CBTs) mainly in the way that the questions administered during a session of assessment are selected. Whilst in a CBT all test-takers are presented with the same predefined set of questions, in a CAT questions are dynamically selected based on each test-taker's performance. A typical CAT starts with a question of average difficulty. If the test-taker answers the question correctly, a more difficult question is administered next. Conversely, an incorrect response will cause a less difficult question to follow. Wainer (2000) argues CATs mimic aspects of an oral interview where the tutor would judiciously modify the oral examination by choosing questions appropriate to the ability and knowledge of individual learners.

### 2.1 Item Response Theory

At the University of Hertfordshire, a CAT prototype has been designed, developed and evaluated over the past three years. This work is described in full by Lilley & Barker (2002; 2004) Barker & Lilley (2003) and Lilley et al., (2002).

In this prototype, the functions to be served by the prototype were what Brusilovsky (2001, p. 5) called "kind of content" and "interface sequencing". The rationale was to tailor the interface content and sequencing – in other words, the questions administered and their order – to the proficiency level of each user. To this end, the statistical Three-Parameter Logistic (3-PL) Model from Item Response Theory (IRT) was employed. By employing such statistical mode, knowledge on user proficiency levels was obtained by unobtrusively monitoring and evaluating user performance during regular tasks rather than by prompting users to specify their levels of confidence as in a cooperative approach.

Equation 1 (Lord, 1980), shows the 3-PL Model function used to predict the probability of a test-taker with an unknown proficiency level  $\theta$  correctly answering a question of difficulty  $b$ , discrimination  $a$  and pseudo-chance  $c$ . In Equation 1, questions with greater values for the difficulty  $b$  parameter require greater proficiency on the part of the test-taker to answer the question correctly than those questions with lower values. The discrimination  $a$  parameter describes the question's usefulness when distinguishing amongst test-takers near a proficiency level  $\theta$  (Wainer, 2000). The pseudo-chance  $c$  parameter indicates the probability of a test-taker answering a question correctly by chance.

$$P(\theta) = c + \frac{1-c}{1+e^{-1.7a(\theta-b)}} \quad (1)$$

A detailed description of IRT is beyond the scope of this paper and the interested reader is referred to Lord (1980) and Wainer (2000).

## 2.2 Prototype overview

One of the central elements of the 3-PL Model is the level of difficulty of the task being performed by the user. Indeed important assumptions of the model include the need to provide a questions database that is accurately ranked according to question difficulty. This was an interesting problem and two approaches were used to achieve the task's difficulty estimate. Firstly, subject experts were used to rank the questions in order of difficulty, based upon their experience of the subject domain and Bloom's taxonomy of the cognitive skills (Bloom, 1956; Anderson & Krathwohl, 2001). An initial value of the difficulty parameter was established in this way, ranging from -2 to +2. Secondly, at the end of each test session, the order of ranking was modified according to user performance per question. Questions that were answered correctly more often had their ranking lowered and questions that were answered incorrectly more often had difficulty levels increased. A similar approach to the calibration of tasks' difficulty was also effectively employed by Fernandez (2003). One important difference between our approach and Fernandez's approach was that use of Bloom's taxonomy of the cognitive skills (Bloom, 1956; Anderson & Krathwohl, 2001). Table 1 shows the differentiation of the database according to level of difficulty.

Difficulty parameter	Cognitive skill	Skill involved
$-2 \leq b < -0.6$	Remember	Ability to recall taught material
$-0.6 \leq b < 0.8$	Understand	Ability to interpret and/or translate taught material
$0.8 \leq b \leq 2$	Apply	Ability to apply taught material to novel situations

Table 1 Values assigned to the difficulty parameter. In the 3-PL Model from Item Response Theory, the difficulty parameter is typically denoted by the letter  $b$  (Lord, 1980).

A further requirement of our CAT prototype was the generation of a user profile differentiated according to topic area within the subject domain. To this end, the questions in the database were also classified according to topic area. Table 2 illustrates the 6 topics areas covered by our database of questions.

<b>Topic area</b>	
1	Issues related to the use of sound in interfaces
2	Graphical representation in interfaces
3	User-centred approaches to requirements gathering
4	Design, prototyping and construction
5	Usability goals and User experience goals
6	Evaluation paradigms and techniques

Table 1 Topic areas within the subject domain covered by the database of questions

Our previous research in the area of Computer Adaptive Testing has been related to the following aspects:

- The establishment of test conditions for a CAT
  - E.g. ability to review questions,
  - Test stopping conditions
  - Stability of CAT levels
- The reliability of CAT measures
  - Test-retest (reliability studies)
- The fairness of the method
  - Comparison to other testing methods (validity studies)
  - CAT levels and their relationship to individual cognitive style
- Student perception of test difficulty
- Calibration of the adaptive questions database
- A comparison of the use of CAT in formative and summative tests

The use of a composite student model to configure learning in blended learning objects is described by Barker (2006). In the research reported in this paper however, we look at the use of a student model based on CAT to configure a blended learning object designed provide automated feedback for learners. We also present the findings of two small studies that assess the attitude of students and staff to the feedback provided, according to a student model based on CAT levels. The diagram below (Figure 1) shows how the components of the CAT-based student model are used to present automated feedback.

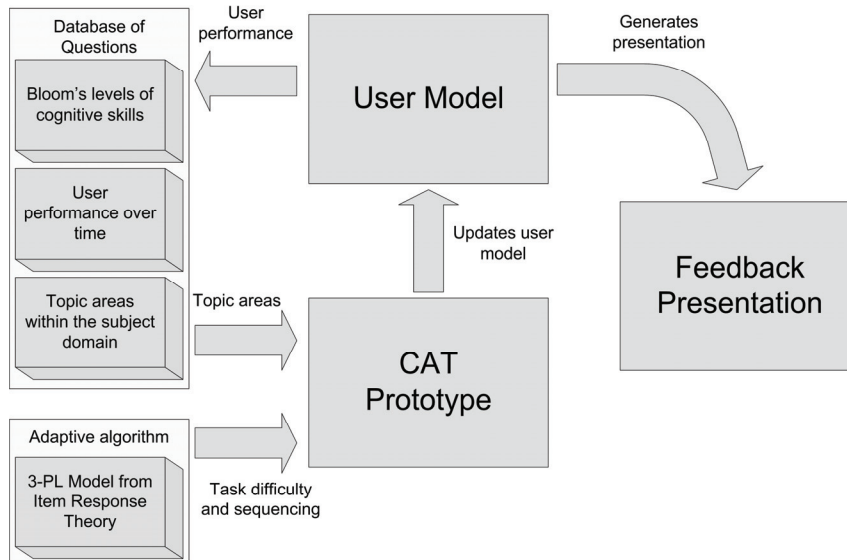


Figure 1 The diagram illustrates the relationship between the mathematical model, the database of questions and the student model

### 3 The Study

The generation of automated, meaningful feedback to individual users of computer-assisted assessment software applications is an area of great interest to academic staff and researchers. In spite of the growth in the use of computer-assisted assessments, the generation of automated feedback for individual development is still a relatively unexplored field (Denton, 2003). Thus, the focus of the empirical study reported here was to investigate how the knowledge gained about the proficiency levels of each individual using our CAT application can be employed to provide learners with personalised recommendations on how to increase their individual proficiency levels.

#### 3.1 Subjects

One hundred and twenty-two second year students enrolled on a programming module of the BSc in Computer Science degree at the University of Hertfordshire participated in a summative assessment session using the CAT application.

#### 3.2 Procedure

The assessment session took place in the University's computer laboratories, under supervised conditions. The participants had 30 minutes to answer 20 questions within the Human-Computer Interaction domain. The test covered the 6 topic areas listed in Table 2.

The test started with a set of predefined questions of average difficulty, one question for each one of the 6 topics covered by the assessment. The remaining 14 questions administered were dynamically selected using our adaptive algorithm. In this study, we assumed that a test-taker's proficiency level in one topic area within the subject domain would be a suitable indicator of proficiency level in a related topic area in the same domain.

### 3.3 Summary of test-takers' performance

In our CAT prototype, the values for the proficiency level  $\theta$  ranged from -2 (lowest) to +2 (highest). The mean for the overall proficiency level was -1.236 (N=122, SD=1.20). The mean for the number of correct responses in the non-adaptive section of the assessment was 41.25% (N=122, SD=22.74). As anticipated, this mean value was lower than the one observed in the adaptive section of the assessment (Mean=46.92%, N=122, SD= 22.35).

This study differs from previous work by the authors (Lilley et al., 2004; Lilley & Barker, 2002; 2004; Barker & Lilley, 2003), in that, in addition to the estimation of an overall proficiency level, test-takers' responses were grouped by topic and a proficiency level calculated for each set of topic responses as shown in Table 2. The mean value for the overall proficiency level was -0.70 (N=122, SD=1.61) for *Issues related to the use of sound in interfaces*; -1.20 (N=122, SD=1.33) for *Graphical representation in interfaces*; -1.26 (N=122, SD=1.35) for *User-centred approaches to requirements gathering*; -0.49 (N=122, SD=1.79) for *Design, prototyping and construction*; -0.78 (N=122, SD=1.58) for *Usability goals and User experience goals* and -0.98 (N=122, SD=1.65) for *Evaluation paradigms and techniques*.

### 3.4 Feedback tool

Our approach to the provision of automated feedback It was one of our assumptions that a tutor-led feedback session would typically comprise the provision of an overall score, general comments on proficiency level per topic and recommendations on which concepts within the subject domain should be revised. It was then planned that the feedback would be made available via a web-based application. In the following section, the feedback tool based on the CAT student model is described.

#### 3.4.1 Overall score

The overall score, or overall proficiency level, would be estimated by the CAT algorithm using the complete set of responses for a given test-taker and the adaptive algorithm introduced in section 2.1. Figure 2 illustrates how this information was displayed within our automated feedback prototype.

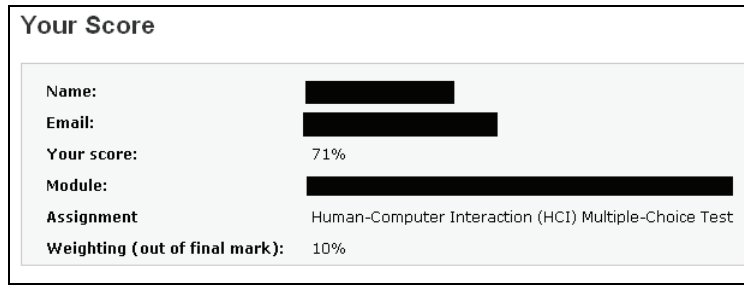


Figure 2 Screenshot illustrating how overall score was displayed within our automated feedback prototype. The student's name and module have been omitted

### 3.4.2 Performance summary per topic

Test-takers' responses would be grouped by topic and a proficiency level calculated for each set of topic responses. Proficiency level estimates per topic would then be mapped to Bloom's taxonomy of cognitive skills. The underlying idea was to inform learners about their degree of achievement for each topic domain. Some learners reported that they would also like to compare their test performance with the performance of the whole group. This information was also made available in this section of the feedback, as illustrated in Figure 3.



Figure 3 Screenshot of screen containing information regarding performance per topic



### 3.4.3 Recommended points for revision

An important assumption of our feedback tool was that tutors providing feedback on an objective test during a face-to-face session were likely to provide students with directive feedback rather than simply indicating what the correct options for each question were. As an initial attempt to mimic some aspects of how a subject domain expert would provide learners with recommendations on how to increase their individual proficiency levels, a database of feedback sentences was designed and implemented. This database comprised statements relating to each one of the questions. For each individual student, only those questions answered incorrectly were selected. Figure 4 illustrates the approach to directive feedback employed in this study.

Recommended Points for Revision	
Your personalised revision plan comprises four sections:	
<ul style="list-style-type: none"> <li>• <a href="#">Identifying needs and establishing requirements</a></li> <li>• <a href="#">Design, prototyping and construction</a></li> <li>• <a href="#">Implementation issues</a></li> <li>• <a href="#">Evaluation paradigms and techniques</a></li> </ul>	
Identifying needs and establishing requirements	
Did you know this?	Further action that you should do...
Use cases are usually employed for capturing the functional requirements of a system. A typical use case should convey a scenario's typical course of events.	Read Sections 7.6.2 and 7.6.3 from "Interaction Design: Beyond Human-Computer Interaction" and identify the difference between a <b>use case</b> and an <b>essential use case</b> .
<b>Utility</b> is a usability goal that refers to the extent to which the system provides the right kind of functionality.	Read Section 1.5.1 from "Interaction Design: Beyond Human-Computer Interaction".
"The system must support a user who is likely to be a well-trained engineer or scientist who is competent to handle technology" is an example of a <b>user requirement</b> .	Identify a user requirement for your Semester B [REDACTED] project.
A user-centered approach is characterised by "Early focus on users and tasks", "Empirical measurement" and "Iterative design".	Read Chapter 9 from "Interaction design: beyond human-computer interaction", focusing on Section 9.3 ("What is a user-centered approach?").
Back to top.	

Figure 4 Example of 'Recommended Points for Revision' for the topic 'Identifying needs and establishing requirements'. The module name has been omitted.

## 4 Learner attitude towards the feedback format used

Laurillard (1993) and Barker & Barker (2002) suggest that the evaluation of educational software is complex and, to be relevant, should be undertaken in a real educational context. This study involved a session of summative assessment for a group of second year students on a Computer Science degree and therefore it was important to investigate their attitude towards the approach used.

With this in mind, an evaluation of learner attitude towards the feedback format adopted was undertaken. This was an optional activity, and 58 out of 122 participants volunteered to provide this information. Participants classified the feedback they received after the assessment as "very useful", "useful" or "not useful". Twenty-nine out of 58 learners classified the feedback received as being "very useful" and the remaining 50% as being "useful". No participant considered the feedback to be "not useful". Thirty-eight out of 58

respondents indicated their satisfaction with the provision of specific points for revision. The provision of feedback according to topic area was also listed as a positive aspect by 22% of the respondents.

Thirteen out of 58 participants suggested the inclusion of the questions answered incorrectly and their key-answers in the feedback document. An underlying assumption was that this would not motivate learners to reflect on their mistakes and thus gain a deeper conceptual understanding of the subject domain. Nevertheless, it can be argued that learners value what Kayashima, Inaba & Mizoguchi (2004) called "learning by awareness" or, in other words, "learning by being aware of one's own mistakes". To address this learners' concern as well as encourage reflection skills, we are planning to extend our feedback database and create one distinct feedback sentence per question. It is anticipated that these sentences would hold greater resemblance to the actual questions than the current comments do.

Some learners suggested an increase on the level of personalisation provided. To this end, we are intending to compare a learner's performance in previous assessments with his or her performance in the most recent assessment. This strategy is likely to provide learners with more meaningful and personalised information on their progress than that offered at present. In spite of its limitations, the feedback approach employed in this study was well-received by learners and fosters future work.

It was important to ensure that the attitude of learners to the automated feedback tool was positive. In CAA 2005, we provided a report of an evaluation of a feedback session with a group of 113 Computer Science undergraduates participated in a session of summative assessment using our CAT prototype (Lilley and Barker, 2005). In that study, students received feedback on test performance via the automated feedback tool.

Students then completed a questionnaire in which they rated a series of statements using a Likert Scale from 1 (Strongly disagree) to 5 (Strongly agree). A group of 97 students answered the questionnaire and their answers are summarised in Table 3.

The results shown in Table 3 suggest that the automated feedback approach was favourably received by the learners who participated in the study. It was therefore important to investigate tutors' attitude towards the automated feedback approach proposed here. It was important to be sure that the approach was also acceptable to staff.

Question	Strongly disagree 1	2	3	4	Strongly agree 5	Mean (SD)
Overall, the feedback tool was effective at providing helpful advice for individual development.	4	5	15	43	30	3.93 (1.02)
Overall, the feedback tool was effective at providing feedback on performance.	4	4	13	44	32	3.99 (1.01)
The "Overall Score" section was useful at providing information on how successfully I have learned.	6	9	23	31	28	3.68 (1.17)
The "Performance Summary per Topic" was useful at providing information on how successfully I have learned in each topic area.	6	6	19	34	32	3.82 (1.15)
The "Points for Revision" section was useful at providing information on how successfully I have learned.	8	9	14	35	31	3.74 (1.24)
Overall, I was satisfied with the degree of personalisation offered by the application.	10	7	19	35	26	3.62 (1.25)
The content of the feedback was appropriate for my individual performance.	6	6	20	39	26	3.75 (1.11)

Table 3 Learners' perceived usefulness of the feedback approach employed (N=97)

## 5 Staff attitude towards the feedback format used

In earlier work, we have shown that a system of automated feedback, based on student performance in a Computer Adaptive Test was useful, efficient and generally well regarded by students as described in Lilley and Barker (2002; 2003; 2004) and Lilley et al., (2004). Barker and colleagues (2002) noted the importance of all major stakeholders in design, implementation and evaluation of projects related to online learning. For this reason, it was important to consider also the views and attitudes of teaching staff to the provision of automated feedback based on a CAT. In order to measure staff attitude a short study was undertaken to obtain views and suggestions from staff related to our automated feedback prototypes.

A group of fifty university lecturers and support staff at an internal University conference presentation on Managed Learning Environments were given a 25 minute presentation on the feedback tool followed by a 5 minute discussion session and a short questionnaire. The presentation session involved a short

presentation of the automated feedback prototype, including sample output screens, examples of feedback and also research data related to student performance and attitude to the feedback provided. Data obtained from nineteen tutors who returned the questionnaire was summarised and collated and is presented in Tables 4 and 5 below.

Question	Not useful 1	2	Useful 3	4	Very useful 5	Mean (SD)
In the context of summative assessment, the automated feedback approach that I have just seen is:	1	1	10	1	6	3.53 (1.17)
In the context of formative assessment, the automated feedback approach that I have just seen is:	0	0	8	3	8	4.00 (0.94)
In the context of objective testing (i.e. multiple-choice questions), the automated feedback approach that I have just seen is:	0	1	7	2	9	4.00 (1.05)
In the context of written assignments, the automated feedback approach that I have just seen is:	6	5	5	0	3	2.42 (1.39)

Table 4 Tutors' perceived usefulness of the feedback approach proposed in this study (N=19)

Question	Poor 1	2	Good 3	4	Very good 5	Mean (SD)
With regards to its <i>speed</i> , the automated feedback approach that I have just seen is:	0	0	4	3	12	4.42 (0.84)
With regards to its <i>quality</i> , the automated feedback approach that I have just seen is:	1	1	8	4	5	3.58 (1.12)
With regards to its <i>appropriateness</i> to enhance students' learning experience, the automated feedback approach that I have just seen is:	1	0	6	4	8	3.95 (1.13)

Table 5 Tutors' perceived speed, quality and appropriateness of the feedback approach proposed in this study (N=19)

It can be seen from tables 4 and 5 that tutors in general considered the approach to be a useful method for the provision of feedback. This is an important finding, since it will be important that tutors as well as students

value the method. Table 4 shows that it is valued more highly in the context of formative, rather than summative, assessment. The use of such automated methods for written assignments was considered the least useful. It was not clear whether this was because of the difficulty of providing automated feedback for written work, or that tutors feel that providing feedback themselves was a better approach. Table 5 shows that on average tutors thought the automated approach to be fast, appropriate and of good quality, though the quality dimension achieved the lowest mean score. All in all tutors' attitude to the approach was positive, which was an important finding.

## 6 Discussion

The work reported in this paper follows from earlier research on the use of psychological student models in an intelligent tutoring system (Barker et al., 2002). In this work, it was found that the approach was beneficial for learners and tutors, yet some global descriptors employed in the model had to be obtained co-operatively and thus inefficiently as explained earlier. To overcome these efficiency issues, one of the focal points of this empirical study was the use of a statistical method to dynamically estimate learners' proficiency levels. In this way, the interface content and level of difficulty of the tasks was interactively modified to match the proficiency level of each individual user. In previous work [Lilley et al., 2003; Lilley & Barker 2002; 2004; Barker & Lilley 2003], we were able to show that the CAT approach was reliable and accurate and that learners were not disadvantaged by its use.

In this study we have shown that a user model based on learners' proficiency levels was effective when applied to the generation of automated feedback. The importance of feedback as a tool to enhance learner's motivation and engagement is well-known and widely reported in the literature (Freeman & Lewis, 1998; Kayashima et al., 2004; Laurillard, 1993; Mathan & Koedinger, 2002). To assist in enhancing learning, feedback needs to be timely, meaningful and constructive. An effect of increased class sizes is that academic staff are often unable to give feedback on learners' assessment performance to the extent they may wish. One solution would be the development of software applications that enable the provision of individual, meaningful feedback to learners. However, the potential benefits of automated feedback have not yet been fully explored by academic staff, even by those who are already making use of computer-assisted assessment tools (Denton, 2003). In this paper, we have described an approach to the provision of automated feedback based on a user model developed using Item Response Theory (IRT) and Bloom's model of learning (Bloom, 1956; Anderson & Krathwohl, 2001).

The automated feedback tool introduced here was a useful application of the user modelling approach. It comprised a proficiency level estimation algorithm

based on Item Response Theory and an adaptive differentiated database of questions. The user profile generated by the CAT was used with a database of feedback sentences. Sentences were selected from the database based on estimated proficiency levels and questions answered incorrectly. For each individual test-taker only those sentences that applied to his or her profile were selected and sent as an attachment to their personal email account. Our automated feedback software prototype was positively evaluated by a group of learners. From the learner's perspective, the provision of a list of points for revision and comments on performance per topic area were the most positive aspects of the approach. Their views were taken to indicate that, at the very least, our automate feedback tool identifies areas for improvement and provides useful advice for individual development.

The application of our user model to provide individual feedback for learners was an important and useful one which was valued by learners in general. In the current phase of this research we are investigating the attitudes of teaching staff to the provision of automated feedback in the way described. In this study we will employ a range of evaluation techniques to obtain information on how staff views the method and to understand the potential benefits and limitations from the tutors' point of view.

The development of an accurate, reliable and efficient user model that covered a range of subject areas was an important outcome of this research. Such a user profile is likely to have many potential applications in the establishment of individual user interfaces in managed learning environments. The goal of an individual presentation strategy for learners will be fostered by the development of such profiles. We intend in future to investigate how they may be obtained efficiently using CAT in formative as well as summative testing and also how they may be used in the presentation of differentiated learning materials and the management of learning.

#### References

Anderson, L.W. & Krathwohl, D.R. (Eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Barker, T. (2006) Attending to Individual Students: How student modelling can be used in designing personalised Blended Learning objects. *Journal for the Enhancement of Learning and Teaching* 3(2) pp. 38-48. Hatfield: University of Hertfordshire Press

Barker, T. & Barker, J. (2002). The evaluation of complex, intelligent, interactive, individualised human-computer interfaces: What do we mean by reliability and validity? In *Proceedings of the European Learning Styles Information Network Conference*, University of Ghent, June 2002.

Barker, T. & Lilley, M. (2003). Are Individual Learners Disadvantaged by the Use of Computer-Adaptive Testing? In *Proceedings of the 8th Learning Styles Conference*.

University of Hull, United Kingdom, European Learning Styles Information Network (ELSIN), pp. 30-39.

Barker, T., Jones, S., Britton, C. & Messer, D. (2002). The use of a co-operative student model of learner characteristics to configure a multimedia application. *User Modelling and User Adapted Interaction* 12 (2/3), pp. 207-241.

Bloom, B. S. (1956). Taxonomy of educational objectives. *Handbook 1, Cognitive domain: the classification of educational goals*. London: Longmans.

Brusilovsky, P. (2001) Adaptive Hypermedia. *User Modeling and User-Adapted Interaction* 11, pp. 87-110.

Denton, P. (2003). Evaluation of the 'Electronic Feedback' marking assistant and analysis of a novel collusion detection facility In *Proceedings of the 7th Computer-Assisted Assessment Conference*. Loughborough University, United Kingdom, pp. 127-134.

Fernandez, G. (2003). Cognitive Scaffolding for a Web-Based Adaptive Learning Environment. *Lecture Notes in Computer Science* 2783, pp. 12-20.

Fischer, G. (2001). User Modeling in Human-Computer Interaction. *User Modeling and User Adapted Interaction* (11), pp. 65-86.

Freeman, R. & Lewis, R. (1998). *Planning and implementing assessment*. London: Kogan Page.

Jettmar, E. & Nass, C. (2002). Adaptive testing: Effects on user performance In *Proceedings of the Human Factors in Computing Systems Conference (CHI 2002)*, April 20-25, 2002, Minneapolis, Minnesota, USA, pp. 129-134.

Kavcic, A. (2000). The Role of User Models in Adaptive Hypermedia Systems In *Proceedings of the 10th Mediterranean Electrotechnical Conference (MEleCon) 2000*, May 2000, Lemesos, Cyprus.

Kayashima, M., Inaba, A. & Mizoguchi, R. (2004). Towards shared understanding of metacognitive skill and facilitating its development. *Lecture Notes in Computer Science* 3220, pp. 251-261.

Langley, P., Thompson, C., Elio, R., & Haddadi, A. (1999). An adaptive conversational interface for destination advice. *Lecture Notes in Artificial Intelligence (LNAI)* 1652, pp. 347-364.

Laurillard, D. (1993). *Rethinking university teaching: A framework for the effective use of educational technology*. London: Routledge.

Lilley, M. & Barker, T. (2002). The Development and Evaluation of a Computer-Adaptive Testing Application for English Language In *Proceedings of the 6th Computer-Assisted Assessment Conference*. Loughborough University, United Kingdom, pp. 169-184.

Lilley, M. & Barker, T. (2003) "Comparison between Computer-Adaptive Testing and other assessment methods: An empirical study", *Proceedings of the 10th International Conference of the Association for Learning Technology (ALT-C)*, University of Sheffield, United Kingdom.

Lilley, M. & Barker, T. (2004). A computer-adaptive test that facilitates the modification of previously entered responses: an empirical study. *Lecture Notes in Computer Science (LNCS) 3220*, pp. 22-33.

Lilley, M., Barker, T. & Britton, C. (2004). The development and evaluation of a software prototype for computer adaptive testing. *Computers & Education Journal* 43(1-2), pp. 109-122.

Lord, F. M. (1980). Applications of Item Response Theory to practical testing problems. New Jersey: Lawrence Erlbaum Associates.

Mathan, S. A. & Koedinger, K. R. (2002). An empirical assessment of comprehension fostering features in an Intelligent Tutoring System. *Lecture Notes in Computer Science* 2363, pp. 330-343.

Vouros, G. (2001). Conceptual modelling of multimedia objects for user-tailored information presentations. *Journal of Applied Artificial Intelligence* 15, pp. 521-559.

Wainer, H. (2000). *Computerized Adaptive Testing (A Primer)*. 2nd Edition. New Jersey: Lawrence Erlbaum Associates.

### **Biography**

Trevor Barker is a lecturer in Computer Science at the University of Hertfordshire. He obtained his PhD in Computer Sciences at this university for research into student models and multimedia learning applications.