Personalizing Assessment in Adaptive Educational Hypermedia Systems

Evangelia Gouli, Kyparisia Papanikolaou, and Maria Grigoriadou

Department of Informatics and Telecommunications, University of Athens,
Panepistimiopolis, GR-15784 Athens, Greece
{lilag, spap, gregor}@di.uoa.gr

Abstract. In this paper, we present a comprehensive framework for assessment, developed through the web-based module named PASS-Personalized ASSessment, which can be integrated in an Adaptive Educational Hypermedia System to provide personalized assessment. PASS estimates learner's performance through multiple assessment options - pre-test, self-assessment and summative assessment - tailored to learner's responses. The adaptive functionality of PASS, which is mainly based on the adaptive testing and the adaptive questions techniques, is described. The first results from the formative evaluation of PASS are encouraging, concerning the total number of questions posed to estimate learner's knowledge level, which is usually less than the maximum needed and the accuracy of the outcome results compared to the estimations of the expert-tutor.

1 Introduction

Educational assessment is a process of drawing reasonable inferences about what learners know on the basis of evidence derived from observation of what they say, do or make in selected situations [8]. An important consideration affecting the design of the assessment is the purpose on which it will be used. Two broad purposes served by assessment are *the assessment to assist learning* usually referred as *formative assessment* and *the assessment of individual achievement* usually referred as *summative assessment.* Formative assessment is part of the developmental process of learning and assesses the quality of learning while summative assessment assesses the quantity and retention of learning following the completion of a unit of instruction [5, 8].

In a web-based distance learning setting, information derived from both assessment approaches as well as the learner's interaction with the system can be stored by the learning environment for further exploitation: (*i*) by *learners* to observe their personal learning progress and to decide how to further direct their learning process, and (*ii*) by *tutors* to individually support learners and formulate judgments about the quality and the effectiveness of the provided content.

Moreover, when the learning environment is an Adaptive Educational Hypermedia System (AEHS) [1], assessment provides a way to estimate learners' knowledge level

P. De Bra, P. Brusilovsky, and R. Conejo (Eds.): AH 2002, LNCS 2347, pp 153–163.
© Springer-Verlag Berlin Heidelberg 2002

and skills through learners' responses in tests, characteristics that are usually used as sources of adaptation. Thus, the generation of assessment tests tailored to learners' responses could mainly enhance the adaptation of a system in this educational context and extend its adaptive capabilities. Several techniques have been recently used in learning environments introducing adaptation to the assessment process such as adaptive testing [11] and adaptive questions [2].

In this paper, we propose a comprehensive framework for assessment, developed through the web-based module named PASS-Personalized ASSessment. Multiple assessment options tailored to the learner's responses are provided: (*i*) pre-testing and self-assessment that correspond to the formative assessment, and (*ii*) summative assessment. The assessment process is based on the central idea of the adaptive testing and the adaptive questions techniques. The PASS module can be integrated in any AEHS with certain characteristics.

The paper is organized as follows. In Sect. 2, a description of the adaptive testing and the adaptive questions techniques is provided. Then, in Sect. 3, the PASS module is presented and in Sect. 4, the functionality of PASS module is described according to the proposed framework of the assessment. In Sect. 5, experimental results from the formative evaluation of PASS module are discussed and in Sect. 6, the paper ends with concluded remarks and plans for further research.

2 Adaptation in Assessment

The enhancement of the assessment process with adaptive capabilities is meaningful for at least two reasons: *(i)* the assessment process becomes dynamic and individualized, as it is adapted to the learner's performance, and *(ii)* the number of questions required to estimate learner's knowledge level is usually reduced, resulting in a less tedious assessment process. The *adaptive testing* and the *adaptive questions* techniques, which can be used to introduce adaptation in the assessment process, are presented below.

Adaptive Testing involves a computer-administered test in which, the selection/presentation of each question and the decision to stop the process are dynamically adapted to the learner's performance in the test [11], [4].

Computerized adaptive testing relies on Item Response Theory [11]. In the adaptive testing procedure, a learner answering questions correctly (or incorrectly) will be gradually administered more difficult (or easier) questions. Questions are selected so that their difficulty matches the learner's estimated knowledge level. The questions that provide most amount of "information" about the actual knowledge level of the learner are usually those with difficulty similar to the learner's knowledge level, as it is estimated by the system, and low guessing factor. The learner's knowledge level estimation depends on the number of questions answered correctly and on the difficulty level of the answered questions. The assessment process will hopefully reach an accurate estimation of the learner's knowledge level after a number of questions posed.

Several approaches exploit the idea of adaptive testing. Huang in [4] describes an adaptive testing algorithm, CBAT-2, that generates content-balanced questions and the SIETTE system [10] is a Web-based testing system with adaptive capabilities.

Adaptive questions technique defines a dynamic sequence of questions depending on learner's responses. Specific responses or learner's performance to certain questions trigger the next series of questions according to several predefined rules, in contrast to the adaptive testing technique, where the triggered question is the one that provides the most amount of "information" about the learner's actual knowledge level.

The adaptive questions technique uses a pool of questions, which are highly structured. Questions are grouped in different classes according to criteria specified by the tutor, such as specific learning outcomes. Furthermore, each class may contain a number of subclasses with different characteristics, such as the difficulty level of their questions. Classes/subclasses are triggered following a certain sequencing mainly determined by the learner's responses. Each time a class/subclass is triggered, all its questions are posed.

Adaptive questions have been used mainly in computer-assisted surveys. Pitkow and Recker in [9] showed that Web-based adaptive questionnaires can reduce the number and complexity of questions posed to users. Furthermore, in CATES [2], "adaptive questionnaires" are used to assess Web users' attitudes.

3 PASS - A Personalized ASSessment Module

PASS is a web-based assessment module, aiming to estimate learner's performance and to assess specific learning outcomes, which are congruent with the learner's learning goal. To this end, it provides multiple assessment options. Tutors can use it in order to define the assessment specifications and to have a detailed overview of the learners' performance and progress. Learners can use PASS for taking a pre-test, self-assessment and summative assessment tests (multiple assessment options) tailored to their responses. PASS can be integrated to any AEHS with a structured domain knowledge and a learner model that keeps information about the navigational behaviour of the learner.

PASS module has been integrated and evaluated in a simulated environment of INSPIRE [7]. INSPIRE is a web-based Adaptive Educational Hypermedia System, which aims to facilitate distance learners during their study, providing them with personalized support. Based on the learning goals that the learner selects, INSPIRE generates lessons that correspond to specific learning outcomes, accommodating the knowledge level of the learner and his/her learning style.

In order to follow the functionality of the different components comprising the PASS module, we provide a brief description of the domain knowledge of INSPIRE. This is represented in three hierarchical levels of knowledge abstraction [7]: learning goals, concepts and educational material. The outcome concepts, which are the most important concepts of a learning goal, are associated with a number of prerequisite concepts. The educational material of an outcome concept is organized in three dif-

ferent levels of performance [6]: *Remember level* (associated with learner's ability to recall the provided content), *Use level* (associated with learner's ability to apply the provided content in specific problems) and *Find level* (associated with learner's ability to propose and solve original problems). On each particular level of performance, one or more educational material pages, comprising of multiple types of knowledge modules, such as examples, theory presentations, exercises, activities, are provided.

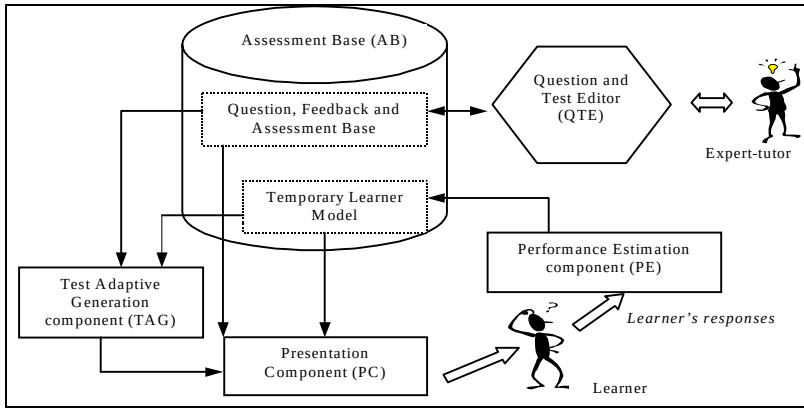

Fig. 1. Schematic of the PASS module

The schematic of the PASS module is illustrated in Fig 1. It is comprised of the following components:

– *Assessment Base* (AB): a data storage, which contains a pool of questions with a variety of parameters, assessment parameters and a Temporary Learner Model (TLM). The TLM stores all the information regarding the learner's interaction with the PASS module and the temporary estimations concerning the knowledge level of the learner.
– *Question and Test Editor* (QTE): a tool that allows tutors to define: (*i*) the questions, (*ii*) the feedback provided to the learner for each of the alternative answers of the question, (*iii*) the question's parameters such as the initial difficulty level, the number of alternative answers, the level of performance (in this case: Remember, Use, Find) that the question assesses, (*iv*) the assessment parameters such as the maximum number of questions in each particular test, the minimum number of questions for each level of performance, the degree of confidence in the estimated knowledge level, the termination criteria, and (*v*) the weight of each educational material page and of each prerequisite concept denoting their importance for the outcome concept. The QTE allows the tutor to concentrate on authoring questions and relieves him/her of the strain of technological details.
– *Test Adaptive Generation* (TAG): selects the appropriate question to be posed to the learner according to: (*i*) the questions' parameters, (*ii*) the assessment parameters, and (*iii*) the current estimation of the learner's knowledge level recorded on the TLM. It takes all the required information from the AB and following the algo-

rithm of the selected assessment option, it selects the "best" next question/class of questions. Furthermore, the TAG uses multiple termination criteria in order to terminate the assessment procedure (see Sect. 4 for more details).
– *Performance Estimation* (PE): estimates the learner's performance level and updates the TLM (see Sect.4 for more details).
– *Presentation Component* (PC): presents the question selected by TAG and the appropriate feedback according to the learner's response. Also, it provides a graphical and a text-based representation of the learner's learning progress.

4 PASS Module's Functionality

PASS, as an assessment module, provides a variety of evidence to support educational decision making, offering three assessment options: (*i*) pre-testing, (*ii*) self-assessment on the content that the learner has already studied, and (*iii*) summative assessment on the overall content. These options, in the context of AEHSs, can provide essential information about the learner's performance in different stages of the learning process. The adequate exploitation of this information can mainly enhance the dynamic adaptation of an AEHS towards: *(i)* personalizing the content that will be delivered to the learner, and *(ii)* providing adaptive navigational support through the content.

Pre-testing functionality. Pre-testing has been designed: (*i*) to provide a preview of the subject matter, (*ii*) to investigate the learner's prior knowledge in the provided content, (*iii*) to diagnose learning difficulties in the prerequisite material, and (*iv*) to initialise the learner's temporary model.

To fulfil the above aims, the three following categories of questions have been specified: (*i*) questions assessing learners' prior knowledge on the subject matter by providing a preview of it, (*ii*) open questions providing information about learners' experience and prior knowledge on related subjects, and (*iii*) questions assessing learners' knowledge on the prerequisite material.

Specifically, the technique of adaptive questions is adopted for the selection/presentation of questions assessing learners' knowledge on the prerequisite material (3rd category of questions). This technique is adopted in order to enable the learners to refrain from frustration, which is usually caused when they are asked to answer questions the level of which is above their knowledge. In the proposed approach, a class of questions is specified for each prerequisite concept of an outcome concept. For each particular class, a number of subclasses is defined, aiming to: (*i*) assess the level of learners' performance, and/or (*ii*) draw meaningful inferences through learners' responses. The rules that trigger the next class/subclass of questions are based on: (*i*) the percentage of learners' correct responses, and/or (*ii*) qualitative criteria defined to draw specific inferences about the gaps in the learners' knowledge.

An overview of the way the pre-testing is provided through the PASS module is described below (not yet fully implemented). Once the learner has decided to take a pre-test, the TAG selects the first two categories of questions and the PC presents

them to the learner. Next, the TAG applies the adaptive questions procedure, for questions assessing learners' knowledge on the prerequisite concepts, taking into account the learner's responses. The TAG triggers the appropriate classes/subclasses of questions following certain rules. All the questions of each triggered subclass are posed. When the assessment procedure is completed for a class, the PE estimates the learner's knowledge level for the particular prerequisite concept with which the class is associated. Then, a new class of questions is triggered. A qualitative model, which classifies learner's knowledge level to one of the four levels of proficiency {Insufficient (In), Rather Insufficient (RI), Rather Sufficient (RS), Sufficient (S)} is used.

Following the termination of the pre-testing, the PE estimates the prior knowledge level of the learner on each outcome concept by taking into account: (*i*) the percentage of correct responses on questions presenting a preview of the subject matter, (*ii*) the learner's knowledge level estimation for the prerequisite concepts of the outcome, and (*iii*) the weights of the prerequisite concepts denoting their importance for the outcome with which they are associated. This estimation is used as an initial estimation of the learner's knowledge level in the adaptive testing procedure, which is performed during the self-assessment and the summative assessment.

Self-assessment and summative assessment functionality. Self-assessment aims to stimulate the learner to contemplate and reflect on the content that s/he has studied, to assess the learner's performance and to provide immediate and informative feedback. Summative assessment intends to evaluate the learning outcomes of the instruction by indicating an end mastery of the learning outcomes for each outcome concept. The learner has always the option to select self-assessment or summative assessment. Moreover, the system has the ability to propose the most appropriate one, according to the navigational behaviour of the learner through the content i.e. in case the learner has not visited all the provided material of an outcome concept, self-assessment is proposed for the outcome.

The construction of the self-assessment and the summative assessment tests is dynamic, depending on the current knowledge level of the learner. Moreover, in self-assessment, the navigational behaviour of the learner through the content is also considered. The assessment procedure takes into account the content that the learner has visited as well as the time that s/he has spent studying the material. Thus, the AEHS needs to keep such kind of information in the learner model. In summative assessment, questions associated with the prerequisite concepts and questions relevant to each outcome concept, are posed.

The technique of adaptive testing is adopted for self-assessment and summative assessment. The assessment procedure in both these approaches works as follows:

– 1st step: The TAG looks for an initial estimation of the learner's knowledge level in the TLM. If no estimation exists, an average is assumed.
– 2nd step: Depending on the assessment option selected, i.e. self-assessment or summative assessment, the maximum number of questions to be posed for each performance level (Remember, Use, Find) is estimated. In the case of self-assessment, the weight of each educational material page (see in Sect. 3, QTE) and the navigational behaviour of the learner, are taken into account. In the case of

summative assessment, just the weight of each educational material page is considered.

- 3rd step: The TAG selects the candidate questions according to the assessment option, the navigational behaviour of the learner and the minimum and maximum number of questions for each performance level. Questions that are associated with important educational material pages (pages with higher weight) have a greater chance of being chosen.
- 4th step: For each candidate question, the TAG calculates the Item Characteristic Curve (ICC) and the Item Information Function (IIF) based on the current estimation of the learner's knowledge level, the difficulty level of the question and the guessing factor [11] (the discriminatory power is omitted). The ICC represents the probability that the learner with a certain knowledge level will be able to provide a correct answer and the IFF is a representation of the amount of "information" provided by each question about the learner's knowledge level. According to the weight of each educational material page and the amount of "information" provided by each question, the TAG selects the "best" question. Usually, questions with difficulty similar to the learner's knowledge level and low guessing factor provide the most amount of "information". In the proposed approach, the difficulty level of each question is initially assigned by the tutor and as the question is used in the assessment procedure, it is re-estimated according to the number of times that it has been answered correctly or incorrectly [4]. The calibration of the difficulty level of questions is necessary.
- 5th step: The PE, according to the learner's response, estimates his/her knowledge level. The knowledge level is adjusted by a quantity, which depends on the previous estimation of learner's knowledge level and on all his/her previous responses. The PE classifies learner's knowledge level for each outcome concept of the selected learning goal, to one of the four levels of proficiency: {In, RI, RS, S}.
- 6th step: The TAG terminates the assessment procedure when any of the following predefined criteria is met: (*i*) when the number of questions posed exceeds the maximum number of questions defined for the test (see Sect. 3, QTE), (*ii*) when all the questions in the AB have already been posed, (*iii*) when the degree of confidence in the estimated learner's knowledge level is high, (*iv*) when the number of questions posed exceeds the maximum number of questions estimated for each particular level of performance (see 2nd step).

Steps 4 and 5 are repeated until the termination criteria are met. For a detailed description of the algorithm, which is followed by the above assessment procedure, see [3]. At the end of the assessment procedure, information on the learner's achievement and charts showing the learner's progress over the learning process are provided and stored in the system for further exploitation by the tutor.

5 Experimental Results

The first stage of the formative evaluation of the PASS module aims to check the validity of its estimations concerning the knowledge level of the learner and the effi-

ciency of the proposed procedures. The study, performed in this context, focused on the module's estimations provided on the self-assessment and summative assessment options, which were compared to: (*i*) the diagnostic process of an expert-tutor, and (*ii*) the simplified process of accounting the percentage of right answers, a method adopted in many AEHSs. In order to check the validity of the estimations, we examined the accuracy of the outcome results, i.e. the percentage of learners who were classified by PASS into the same level of proficiency as classified by the expert-tutor. The efficiency of the assessment procedure used in the summative assessment option (the same procedure is used in self-assessment) was examined through the total number of questions posed to the different learners.

The students' data used in this study have been obtained from an experiment, which was performed as a part of the evaluation of INSPIRE at the Dept. of Informatics and Telecommunications, University of Athens. In this experiment, twenty undergraduate students used INSPIRE for two hours in order to study the learning goal "Describe the Role of Cache Memory and its basic operations". Once the students had studied the educational material of the outcome concept "Mapping Techniques", they were asked to submit an assessment test on the particular outcome (summative assessment). The professor of the "Computer Architecture" course who had the role of the expert-tutor assessed these tests. The estimations of the expert-tutor were based on the general impression given by the tests, the difficulty of the questions, the number of correct answers on each performance level and the comparison between students' answers.

Students' answers on the test were also provided to PASS and the expert-tutor assigned the required parameters. The maximum number of questions posed in the test was assigned to 15. The initial difficulty of the questions was assigned according to the level of performance that the question assessed, e.g. for questions assessing the Remember level of performance, the initial difficulty was 0.3. For question's difficulty calibration, data from the final exams of the course "Computer Architecture" were used. The initial knowledge level of the students was assumed as {RI}. The scale of the four levels of proficiency, i.e. {In, RI, RS, S} was experimentally set. This scale provides results, which are closer to the estimations of the expert-tutor. The study was also performed by classifying learner's knowledge level to five and to six levels of proficiency. We found that if four levels of proficiency are considered then the accuracy of the estimations is higher compared to the other two approaches.

Additionally, we estimated the students' knowledge level based on the percentage of correct answers, according to heuristic rules i.e. if the percentage of correct answers is between 0-25% or between 26-50% or between 51-75% or over 75% then the proficiency level is estimated as {In} or {RI} or {RS} or {S} correspondingly.

Figure 2 shows the summative assessment results obtained from the classification of 20 students into the four levels of proficiency using the three estimation approaches. Unfortunately, none of the students' proficiency level has been characterized as {S}. The reader may notice that for 17 out of 20 students, the proficiency level estimations, resulted from the PASS module, coincide with the expert-tutor's estimations. On the other hand, for only 4 students, the expert-tutor's estimations coincide with estimations resulted from the percentage of correct answers approach.

Although the sample is rather small to reach a safe conclusion, the experimental results imply that the student's proficiency level estimations provided by PASS follow the expert-tutor's estimations with high accuracy.

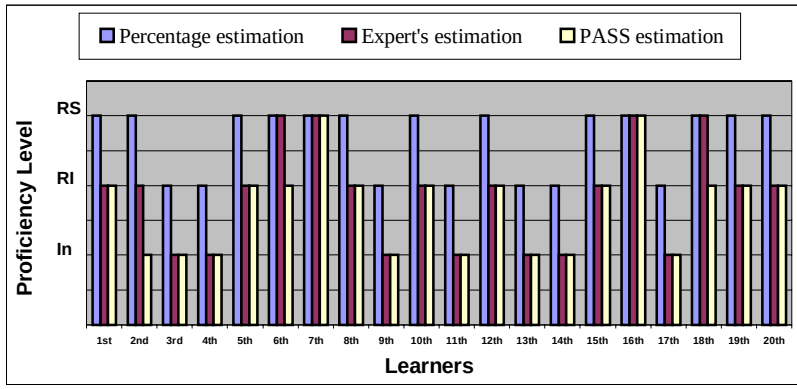

Fig. 2. The proficiency level of 20 students on the concept "Mapping Techniques" as it is estimated by: *(i)* calculating the percentage of students' correct responses, *(ii)* the expert-tutor, and *(iii)* the PASS module.

Furthermore, another interesting point, which we considered, was the total number of questions that were posed to each student, a parameter that mainly affects the testing time. Figure 3 shows the total number of questions posed to students in order to estimate their proficiency level. The reader may notice that 11/20 students answered less than 15 questions. These results imply that less time is needed to take the assessment test since fewer questions are required to achieve an acceptable accuracy. Furthermore, Fig. 4 shows the relationship between the average number of questions posed and the estimation of student's level of proficiency. The initial students' proficiency level was assumed as {RI} and as we have mentioned above, questions were selected in a way that their difficulty was similar to the student's knowledge level. The reader may notice, in Fig.4, that for students with a proficiency level estimated as {RS}, the average number of questions posed was less, compared to those with a

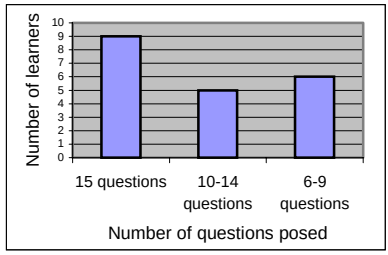

Fig. 3. Number of questions posed by the PASS module

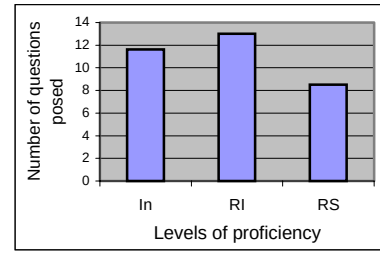

Fig. 4. Average number of questions posed for each level of proficiency

lower level of proficiency, i.e. {In} or {RI}. One possible explanation is that students with lower levels of proficiency usually guess the answers, so the assessment procedure needs to pose more questions for estimating accurately their knowledge level.

6 Conclusions and Further Research

In this paper, we presented PASS, a web-based and easy to integrate in an AEHS, assessment module. PASS offers three assessment options tailored to learner's responses: pre-testing, self-assessment and summative assessment. The technique of adaptive questions has been adopted for pre-testing. The discriminative characteristics of the self-assessment and summative assessment procedures adopted in PASS are: the adoption of the adaptive testing technique, the consideration of the learner's navigational behaviour, the re-estimation of the difficulty level of each question each time it is posed (initially assigned by the expert-tutor), and the consideration of the importance of each educational material page.

The experimental results from the formative evaluation of PASS, even performed on a limited test group, have been encouraging and indicate that the estimations on learners' proficiency are close to the expert-tutor estimations. We have also shown that a reduction in the number of questions posed can be achieved, especially for learners with higher levels of proficiency, resulting in a reduction of the testing time.

Our future plans, apart from the evaluation of the PASS module with a wider group of learners, include the completion of the pre-testing implementation and the adaptation of the feedback provided to the learners' according to their performance.

References

1. Brusilovsky, P.: Adaptive and Intelligent Technologies for Web-based Education. In: Rollinger, C., Peylo, C. (eds.): Special Issue on Intelligent Systems and Teleteaching. Kunstliche Intelligenz, Vol. 4 (1999) 19-25
2. Chou, C.: Constructing a computer-assisted testing and evaluation system on the World Wide Web–the CATES experience. IEEE Transactions on Education 43(3)(2000) 266-272
3. Gouli, E., Kornilakis, H., Papanikolaou, K., Grigoriadou, M.: Adaptive Assessment improving interaction in an Educational Hypermedia System. In: Avouris, N., Fakotakis, N. (eds.): Advances in Human-Computer Interaction I. Proceedings of the PanHellenic Conference with International Participation in Human-Computer Interaction (2001) 217-222
4. Huang, S.: A Content-Balanced Adaptive Testing Algorithm for Computer-Based Training Systems. In: Frasson, C., Gauthier, G., Lesgold, A. (eds.): Intelligent Tutoring Systems, Third International Conference, Springer (1996) 306-314
5. Kommers, P., Grabinger, S., Dunlap, J.: Hypermedia Learning Environments. Instructional Design and Integration. Lawrence Erlbaum Associates (1996)
6. Merril, D.: Component Display Theory. In: Reigeluth, C.M. (eds.): Instructional design theories and models: An overview of their current status. Lawrence Erlbaum Associates, Hillsdale NJ (1983) 279-333

7. Papanikolaou, K., Grigoriadou, M., Kornilakis, H., Magoulas, G.: INSPIRE: An INtelligent System for Personalized Instruction in a Remote Environment. In: Reich, S., Tzagarakis, M.M., De Bra, P. M.E. (eds.): Hypermedia: Openness, Structural Awareness and Adaptivity. Lecture Notes in Computer Science Vol. 2266, pp. 215-225. Springer-Verlag, Berlin (2002)
8. Pellegrino, J., Chudowsky, N., Glaser, R. (eds): Knowing what students know: The Science and Design of Educational Assessment. National Academy of Sciences, National Academy Press, Washington (2001)
9. Pitkow, J., Recker, M.: Using the Web as a Survey Tool: Results from the Second WWW User Survey. Computer Networks ISDN Systems 27(6) (1995) 809-822
10. Rios, A., Millan, E., Trella, M., Perez-de-la-Cruz, J., Conejo, R. : Internet Based Evaluation System. In: Lajoie S., and Vivet M. (eds.): Artificial Intelligence in Education, Open Learning Environments, New Computational Technologies to Support learning Exploration and Collaboration. Frontiers in AI and Applications, Vol 50. IOS Press (1999) 387-394
11. Wainer, H., Dorans, N., Eignor, D., Flaugher, R., Green, B., Mislevy, R., Steinberg, L., Thissen, D. (eds): Computerized adaptive testing: A primer. 2nd edition. Lawrence Erlbaum Associates, Hillsdate NJ (2000)