How could pedagogy make a choice between personalized learning systems?

Julie McDonald[1], and Jari Timonen[2]

[1]Dundee University, Centre for Medical Education, Dundee Scotland
[2]University of Joensuu, Department of Computer Science and Statistics, Finland
j.v.mcdonald@dundee.ac.uk
jari.timonen@cs.joensuu.fi

Abstract. The paper considers usability evaluation methods developed for adaptive educational systems. Furthermore, the application of pedagogy and instructional design is discussed in relation to the evaluation process of adaptive systems. By combining the strengths of existing usability evaluation methods and considering a criteria for instructional design, two adaptive systems, MOT and ELM-ART, are evaluated. In the evaluation we combine pedagogical and heuristic evaluation. Conclusions are made in order to support pedagogical decision-making between adaptive systems.

Keywords: Adaptive Systems, evaluation, pedagogy, instructional design

1 Introduction

Interest towards adaptive systems is increasing. Instead of 'one-size-fits all' systems (Brusilovsky, 2001) more personalized systems are receiving considerable attention. However, the evaluation of adaptive systems is challenging due to non-standardized methods and an increasing multi professional approach to delivering instruction. There is also concern that focusing on producing and administering systems in terms of usability may stop short of considering the instructional process and pedagogy in evaluation.

There is a need to look at systems, which are "context- aware" (Abowd et al., 2002) and account for pedagogy (Kinshuk et al., 2000; Dagger et al., 2004). This is not a new problem and indeed not limited to evaluation of systems in terms of usability. The "no significant difference" phenomenon (Phipps & Merisotis, 1999; Joy & Garcia, 2000; Lockee, Moore & Burton, 2001), for example has been the drive behind extending evaluation approaches in the field of instructional design[1] to include not only how a system works but also the pedagogy involved. As Kinshuk et al. (2000) point out "Initially, the system should be evaluated for its overall effectiveness and usability."

Evaluation is a two-phase process, which can be seen as formative i.e. at the time of development and summative i.e. after implementation (Scriven, 1967). Usability evaluation is more commonly carried out in the formative phase while the educational initiative is forming in order to improve the system, while evaluating pedagogic effectiveness is common in both formative and summative phases (evaluating the impact of the educational initiative). The scope of this paper is how to incorporate an evaluation of pedagogy into the formative usability evaluation.

The paper first looks at common methods for evaluating adaptive systems. The pedagogical dimension in relation to implementing learning within an adaptive learning environment is then examined and an evaluation heuristics chosen for its consistency with an instructional design approach.

[1] Instructional design has a long history and can be described as the systematic process of translating general principles of learning and instruction into plans for instructional materials and learning. As a discipline it is based on two foundations. Firstly a systems design model for managing the instructional development process and secondly pedagogic models that specify how instruction can effectively promote learning (Reigeluth, 1983, 1987).

The paper goes on to look at the application of the evaluation framework of two adaptive systems: MOT and ELM-ART. The former enables the user to create concept maps and use them for goal map creation that is finally structured into study lessons. The latter is an intelligent system that provides the framework for interactive learning of LISP using a web interface. The adaptive systems and their evaluations are then described and finally conclusions and possible future directions are presented.

2 Evaluation of Adaptive Systems

Several evaluation approaches designed for adaptive systems have been proposed and the benefits of the layered evaluation described (Brusilovsky et al. 2001; Weibelzahl, Lippitsch, & Weber, 2002; Chin, 2001). The most common approach is the heuristic evaluation, which has a wide application in usability studies not only in industry but also in education (Baker et al., 2001). However despite the benefits of heuristics, (the short training period needed to apply this approach and its ability to evaluate the whole system) this approach has been criticised for its focus on learner behaviour rather than the elements of adaptivity (Höök, 2000). A response to this is the layered approach to the evaluation of adaptive learning systems, which accounts for adaptivity components (Brusilovsky et al. 2001).

While Chin (2001) looks to overall performance in the layered evaluation, the main advantage is the ability to break the utility into several functions. Brusilovsky et al. (2001) define the success of the adaptive system based on the evaluation of two layers: user-computer interaction and adaptation decision-making. The idea is to evaluate the system with and without adaptive components. In the assessment of the first layer, conclusions for adaptation are based on the validity of adaptive components and the possibility for further adaptivity.

The evaluation of adaptive learning systems using this approach therefore looks to the validity and meaningfulness of adaptive decisions made by the system. But there still remains a gap in evaluation methods, which fail to take into consideration the multiple dimensions of evaluating adaptive learning systems which are more learner centred rather than function centred.

Magoulas et al. (2003) propose addressing this shortcoming by combining heuristic evaluation (Nielsen,1994a and 1994b) to a multi layered evaluation. In doing this heuristics are extended to cover a six layer evaluation model proposed by Weibelzahl and Lauer (2001). The layers of the latter consist of (1) correctness of input data acquisition, (2) correctness of inference, (3) appropriateness of adaptation decisions, (4) change of system behaviour when the system adapts, (5) changes in user behaviour when system adapts, and (6) change and quality of total interaction. Based on the previous layers Magoulas et al. (2003) extend this model by the addition of pedagogical heuristics. This extended model defines detailed criteria tables in order to evaluate adaptive educational systems, which reflect a pedagogical utility in adaptive learning environments. The following list contains the main items of 'pedagogical heuristics' by Magoulas et al. (2003).

1. Visibility of the Adaptive Learning Environment
2. Match between Adaptive Learning Environment and the real world
3. Various levels of learner control
4. Consistency and standards
5. Help users recognize, diagnose and recover from errors
6. Error prevention
7. Recognition rather than recall
8. Flexibility and efficiency of use
9. Aesthetic and minimalist design
10. Searchable help functions and documentation

There is much to be said for the strengths of the layered and the heuristic approach to evaluation of adaptive learning. However, despite considering Magoulas et al.`s (2003) layered evaluation as a feasible method, evaluation of adaptive systems still remains challenging and non-straightforward. Cristea (2004) discusses adaptivity and adaptive (educational) hypermedia and questions the viability of evaluation solutions where the dimensions to be considered are multiple. Moreover, as Cristea (2004) points out, the increasing use of different roles in adaptive

hypermedia also indicates a need to look at each role separately. This would support the layered approach. Herder (2003) puts forward a utility-based evaluation method for adaptive systems, which considers the limitations of a layered model. Herder (2003) proposes that the learner should be central to the evaluation, interaction being an important factor. He further stresses the need to take the cognitive, affective, conative, and demographic qualities of the learner into account. His rational is that such a utility-based model then measures the overall satisfaction of learner interaction as well as adaptive decisions made by system. Taking this further, Dagger et al. (2004) emphasize the importance of considering a pedagogical approach, which goes beyond heuristics and supports the instructional strategies of learning environments. The question remains how to evaluate adaptive learning environments acknowledging the obvious strengths of usability and pedagogical heuristics as with Magoulas et al.`s (2003) approach and at the same time considering the advantage of evaluating instructional strategies.

3 Adaptivity and Pedagogy

Evaluation of instructional design is based on a selection of instructional strategies in relation to the proposed educational environment (Reigeluth, 1983). Gagné (1985) outlines an instructional process of learning with the following nine items.

(1) *Gain attention*: in order to focus your learners on the task at hand, instruction should seek to gain interest and gain attention. (2) *Inform learners of objectives*: after gaining learners' attention, establish the purpose of the instruction, this helps to facilitate learning and directs the learners. (3) *Stimulate recall of prior learning*: in order to make instruction relevant and meaningful, topics should be linked to prior knowledge. This maintains and stimulates interest. (4) *Present the* content: by using a variety of media, including text, graphics, audio narration, and video content which is well organized. Provide intuitive navigation and a user-friendly interface. (5) *Provide learner guidance*: provide learner guidance, advise learners of resources available and guide them through the use of some effective learning strategies e.g. matching to knowledge level, navigation. (6) *Elicit learning/practice*: provide learners with the opportunities to practice what they have learnt and feedback on their understanding. (7) *Provide feedback*: provide useful and immediate feedback if a learner makes an error, and provide guidance to further learning following errors. (8) *Assess learning*: summative assessment occurs at the end of an instructional unit but may be formative. Make sure that there is feedback and further guidance. (9) *Enhance retention and transfer*: by using metaphors and analogies, adaptive and connect new information to existing knowledge, future knowledge, career, work, or life experiences.

4 Evaluations of Adaptive Educational Systems

Based on the review of evaluations of adaptive systems we found considerable challenges in terms of choosing an appropriate evaluation model. Many of the proposed methods describe their functionalities in a high abstraction level that require expanding actions in order to apply them to the evaluation of real applications. Consequently, it is necessary to ask what is the abstraction level required for conducting evaluations without losing generalization of methods.

In order to address both the issue of functionality and pedagogical issues of instructional design we selected the layered evaluation method proposed by Magoulas et al. (2003) and the pedagogic strategies according to Gagnés (1985) instructional events. The evaluation heuristics described by Magoulas et al. (2003) was further chosen for its consistency with Gagnés (1985) instructional design approach. We scored each evaluation criteria with the scale from 0 to 3 and allocated following descriptors to each score: 3 = requirements met or possible; 2 = minor repairs suggested; 1 = repairs suggested; and 0 = extensive repairs suggested. In order to put this into practice we carried out evaluations of two adaptive educational systems MOT and ELM-ART based on the selected method.

4.1 MOT

MOT is an editing tool creating adaptive hypermedia for educational purposes. Although MOT can be applied to other domains in this paper we describe and evaluate the system from the perspective of the development of educational content. MOT enables learning material creation without technical HTML skills thus it is targeted in particular for teachers who might lack such skills. MOT enables the user to create concept maps, which are used to create goal maps, and finally structured into adaptive lessons. The MOT system is based on the LAOS model, a five-layer model for adaptive hypermedia authoring (Christea and Mooij, 2003a). Although the evaluation of adaptive systems is challenging MOT has been evaluated based on its planned goals and their fulfilment in the actual system (Cristea & Mooij, 2003b). One of the strengths of the MOT system is to enable a pedagogical labelling and weighting of concepts (Christea et al, 2005). MOT presents authors with a high degree of freedom to make choices according to their own pedagogical approach. In this way MOT attempts to solve the problem of satisfying multiple teaching styles and ultimately multiple learning styles.

4.2 ELM-ART

ELM-ART (ELM Adaptive Remote Tutor) is an adaptive WWW learning system, which implements different types of adaptivity and adaptability. A multi-layered model permits an advanced link annotation and sequencing of content. Its strength lies in an episodic learner model, which allows the system to identify and analyse learner problems with subsequent individualized solutions (procedural knowledge). Furthermore ELM-ART uses domain knowledge to adapt to the learner's knowledge state.

ELM-ART provides the framework for interactive learning of LISP using a web interface. The course starts by assessing the students' experience of using the Internet, programming and use of computers. In this way the preliminary lesson is adapted to the learner's knowledge base and subsequent lessons are introduced according to an inferred knowledge state. The main advantages of the learning system are the adaptivity to the learner's knowledge state as well as feedback on completed items and visited pages via its advanced annotation of links.

4.3 Evaluating MOT and ELM-ART as adaptive systems

The authors felt the choice of systems appropriate to evaluate the heuristic elements and pedagogic elements of adaptive systems. MOT being a tool for creating an adaptive system, which presents the possibility of a pedagogical approach according to the chosen criteria and ELM-ART being an adaptive system that has already been defined in pedagogic terms. In this way, we are able to consider evaluation as a formative process i.e. at the time of development rather than as summative after implementation (Scriven, 1967). Draper, Brown, Henderson & McAteer, (1996) for example argue for an 'integrative' evaluation of teaching strategies which address student learning and the ultimate pedagogic goals of the educational intervention. Therefore, this evaluation approach not only considers the importance of a formative evaluation of the utility and pedagogy of a system such as ELM-ART, which may be further developed, but also the potential for pedagogical goals of a system such as MOT in the development of educational content.

4.4 Results

The results of the evaluation based on Gagné `s (1985) instructional events and the previously described heuristic evaluation from Magoulas et al. (2003) are shown below.

Table 1: Pedagogic evaluation according to Gagnés instructional events

Instructional events	Criterion	MOT	ELM-ART
Gain attention	Gain interest and gain attention.	Simple UI and instructions (2)	Good introduction and overview (3)
Inform learners of objectives	Establish the purpose of the instruction, facilitate learning and direct the learners.	Objectives described. However, the descriptions from research perspective does not necessarily benefit the user of system (2)	Objectives clearly explained (3)
Stimulate recall of prior learning	Link topics to prior knowledge.	No recall present (2)	Recall present (3)
Present the content	Use a variety of media, including text, graphics, audio narration, and video content which is well organized. Provide intuitive navigation and a user-friendly interface.	Media mix not used, navigation not always straightforward (1)	Media mix not used, navigation intuitive and user friendly (2)
Provide learner guidance	Provide learner guidance, advise learners of resources available and guide them through the use of some effective learning strategies e.g. matching to knowledge level, navigation.	User manual of the system guides the usage (3)	Guidance present with learning strategies(3)
Elicit learning/practice	Provide learners with the opportunities to practice what they have learnt and feedback on their understanding	Practicing enabled, no feedback on understanding (2)	Opportunities to practice present (3)
Provide feedback	Provide useful and immediate feedback if a learner makes an error, and provide guidance to further learning following errors.	Error handling awkward in some situations (1)	Useful and immediate feedback present, guidance for further learning absent (2)
Assess learning	Summative assessment and formative with feedback and further guidance.	Not present (0)	Summative and formative with feedback present. (3)
Enhance retention and transfer	Using metaphors and analogies, adaptive and connect new information to existing knowledge, future knowledge.	Not present (0)	Not present (0)
SCORING		**13/27**	**22/27**

Tables 2-7 below show the evaluation results of MOT and ELM-ART based on a method proposed by Magoulas et al. (2003). Table columns show the heuristic evaluation criteria, and observations based on the functionality of MOT and ELM-ART.

Table 2. Criteria for layer 1: Correctness of input data acquisition

Heuristic	Criterion	MOT	ELM-ART
H6: Error Prevention	Data inputs are case-blind whenever possible.	Non case-blind inputs exist. Not causing errors. (3)	No data input possibility(1)
	When learners navigate between multiple windows, their answers are not lost.	Navigation and redirecting require repairs.(0)	Navigation and redirecting require repairs.(0)
	Buttons that can cause serious consequences are located far away from low-consequence and high-use keys.	Use of confirmation boxes instead of decentralization of buttons. Equalization of logging buttons required. (2)	Error prevention required. (2)
	The de-activation of the adaptivity always requires users' confirmations.	Continuous adaptation. (3)	Continuous adaptation. No de-activation possibility (2)
	The change of learning model always requires learner confirmation.	Stable learning model for a student. Teacher responsible for changes. (3)	No possibility for change of learning model. (2)
	The input fields to identify the learner preferences are not set to default values.	No learner preference identification. (3)	No user preference identification. (3)
H7: Recognition rather than Recall	Required data entry fields are clearly marked	Teacher entries clearly marked. No data entries for students. (3)	User entries clearly marked. (3)
SCORING		**17/27**	**13/27**

Table 3. Criteria for layer 2: Correctness of inference

Heuristic	Criterion	MOT	ELM-ART
H3: Various Levels of Learner Control	The same content is presented in various formats according to the learning profile.	Student view similar to all. Multiple content creation (for the same topic) possible for a teacher (1)	User view can be personalized. (3)
	Multiple levels of explanation are applied to help students find the correct answers.	Student view similar to all. (1)	System gives feedback in response to given solution (3)

	The system provides both basic and advanced subject content.	Creation of basic and advanced content possible for a teacher. (3)	Advanced searching (Boolean) possible (3)
H6: Error Prevention	Users are prompted to confirm actions that have drastic and destructive consequences.	Alert and confirmation windows used. (3)	Alert and confirmation windows not noted. (3)
	When learners are taking assessments, they can change their previous answers.	Authoring tool. Student view similar to all. (3)	Students can retake assessments. (3)
SCORING		**11/15**	**15/15**

Table 4. Criteria for layer 3: Appropriateness of adaptation decisions

Heuristic	Criterion	MOT	ELM-ART
H2: Match between systems and the Real World	The relevant concepts are organized in an understandable way.	Teacher responsible for organizing. (2)	Concepts are organized. (3)
	Users can easily identify the logical order of the relevant concepts.	Teacher responsible for organizing. (2)	Logical order of the relevant concepts realized. (3)
	The arrangement of the menu items matches with learner's profile.	Teacher responsible for arrangement. (2)	User menu matches with learner's profile (3)
	The concepts are explained in a way that matches with learner's profile.	Teacher responsible for organizing. (2)	User menu matches with learner's profile (3)
	The explanation of the assessment result is meaningful to the learner.	Authoring tool, no assessment in system.(3)	Feedback given which fulfils this criteria (3)
H4: Consistency and Standard	The layout is consistent for different learning models.	Limited learning model support or instructions. (0)	User model supported with instructions. (3)
	Online instructions are displayed in a consistent location across different learning models.	Teacher manual available. Limited support for learning models. (0)	Criteria fulfilled (3)
	There are not too many types of icons to identify the difficulty levels of the content.	No. icon types minimal. (3)	Criteria fulfilled (3)
SCORING		**14/24**	**24/24**

Table 5. Criteria for layer 4: Change of system behaviour when the system adapts

Heuristic	Criterion	MOT	ELM-ART
H1: Visibility of Learning Environments	There is visual feedback to inform learners about their current progress.	Authoring tool. Limited support for progress feedback. (2)	Visual feedback is present. (3)
	After a task is completed, learners are informed about the next step.	Sequential progression, no next step information. (2)	Sequential progression, no next step information. (2)
	The system can let learners easily identify content that is not suitable for his/her current status.	Teacher responsible for content. Learner cannot influence this. (1)	Identification possible (3)
	Adaptation always suggests a route through the program.	Sequential progression. (3)	Suggested progression (3)

H8: Flexibility and Efficiency of Use	It is easy to select relevant information directly.	No shortcuts (hyperlinks) in text. (1)	Direct selection of relevant information (3)
	Students are allowed not to read the subject from beginning to end.	Allowed to read. (3)	Not allowed to read (3)
SCORING		**12/18**	**17/18**

Table 6. Criteria for layer 5: Change of system behaviour when the system adapts

Heuristic	Criterion	MOT	ELM-ART
H5: Help users Recognise, Diagnose, and Recover from Errors	Error messages inform learners of errors severity.	Adequate messages to teachers, no messages to students. (3)	Adequate messages to user (3)
	Error messages let learners understand the cause of the problem.	No messages to students. (3)	Messages not always conclusive (2)
	When learner chooses a wrong concept, the system allows them to de-select it.	De-select for teacher allowed. No concept selection control for students. (3)	De-select possible (3)
	When learner chooses an unsuitable learning model, the system allows them to go back	Limited learning model support. (1)	Limited model support. (1)
H3: Various Levels of Learner Control	The system allows learners to select/deselect the adaptivity.	No possibility. (3)	No possibility. (3)
	The system allows learners to choose their own learning model.	Limited learning model support. (1)	Limited model support. (1)
SCORING		**14/18**	**13/18**

Table 7. Criteria for layer 6: Change and quality of total interaction

Heuristic	Criterion	MOT	ELM-ART
H9: Aesthetic and Minimalist Design	Menu items are brief, yet long enough to describe the subject content.	Menu items brief but long enough. (3)	Menu items brief and long enough. (3)
	The title of each page is short, simple, and distinctive	Criteria met. (3)	Criteria met. (3)
	The colour scheme matches with the preference of each individual student.	No colour adaptation. (2)	Colour adaptation. (3)
H10: Help Functions and Documentations	There is an alphabetical index to search information within the help system.	Alphabetical glossary available. (3)	No Alphabetical glossary available. (3)
	Learners are allowed to switch between the help system and their tasks.	Glossary opens to same window. (2)	Criteria met (3)
SCORING		**13/15**	**15/15**

Table 8 summarizes both the pedagogical evaluation and heuristics evaluation as well as shows total evaluation scores for both the systems.

Table 8. Summary of the evaluation.

Evaluation method	MOT	ELM-ART
Pedagogic	13/27 *(48 %)*	22/27 *(81%)*
Heuristics	81/111 *(73 %)*	102/111 *(92 %)*
Total	94/138 *(68 %)*	124/138 *(89%)*

4.5 Discussion

Based on Gagnés instructional events (1985) and the heuristic criteria from Magoulas et al. (2003) the individual scoring criteria demonstrates a tendency for each system to be preferable in terms of its use and pedagogical application according to individual criteria applied. Using the criteria for Gagnés instructional events all criteria are well served in ELM-ART with a weakness identified in feedback as well as enhanced retention and transfer mechanisms.

The pedagogical evaluation of MOT in comparison demonstrated that it gives limited pedagogical support for the teacher in developing a course which meets the majority of Gagnés instructional events criteria. However, as Okamoto, Cristea & Kayama (2001) point out, systems should consider the potential of collaboration and feedback from tutors and peers as well as feedback through adaptivity. Therefore functions of feedback in particular might be better served through the collaborative environment than a standard response derived from a system.

Turning to the heuristic evaluation, high scores were found in the majority of the criteria, weaknesses lying in MOT in the areas of different presentations, multiple levels of explanation, inconsistency in the display of online instructions (H3), content identification by learners versus knowledge level and learner navigation relating to content level (H6). ELM-ART on the other hand was relatively consistent in all areas and demonstrated weaknesses, in contrast to MOT, only in prevention of errors on data input (H6). Otherwise ELM-ART presented a similarity with the MOT system in limited learning model support (H3).

The individual weaknesses of MOT, however, are not directly comparable to those of ELM-ART. The strengths of the MOT system lie in the opportunity to apply the criteria identified rather than a finished product as with ELM-ART. Therefore the evaluation is an indication only of the direction for pedagogical and usability issues when making a decision between systems. For example, an extremely high global score as with the ELM-ART system indicates that the system has taken essential educational or usability issues into consideration and an extremely low global score would indicate that certain educational or usability criteria are absent. On balance the MOT system demonstrated an overall lower global score than ELM-ART, which may be of consideration for development of the adaptive content and its application.

One of the major limitations of this evaluation is the use of a relative scoring system rather than an absolute score for the systems. There is no absolute score for either system and the evaluation is based on subjective results rather than an empirical measurement. This subjectivity may be a weakness in that the evaluation judgments are open to debate. For example, the heuristic evaluation of "Stimulate recall of prior learning" of ELM-ART could have resulted in a high score if the episodic learner model had been given more weight. The episodic learner model of ELM-ART enables the system to refer back to actual code examples that learners have previously seen and may be considered a route to stimulate prior learning. Further examples of interpretative judgment could be described to illustrate this point.

One of the major concerns of creating educational content is the individual usability and pedagogical possibilities of developed and to be developed adaptivity projects and their subsequent impact on student learning. Okamoto et al. (2001) have stressed the need to “balance the usage of new technologies in view of the learning goals” and further highlight the problems of evaluation and comparison of educational systems. Following on from this it could be said that as a holistic decision making process the evaluation approach described above, although less than accurate, could be used as a basis for further investigation on implementation of created content for educational use. In this respect while the relative scales demonstrated in this paper are presented as suitable for a global evaluation of “possible” pedagogical and heuristic qualities, critical to both systems would be their application in practice and subsequent evaluation of the identified criteria.

5 Conclusions

Evaluation of the pedagogical soundness of adaptive systems is challenging. In this paper we reviewed existing evaluation methods, discussed pedagogical viewpoints, proposed a scoring

based evaluation method that combines pedagogy and heuristics, and evaluated two systems with the combined method.

We recognized that the identified pedagogical possibilities and heuristics are applicable only in this formative phase and address the usability of the systems only as a snap shot scenario. However evaluation, both formative i.e. at the time of development as well as in the summative phase i.e. after implementation (Scriven, 1967) are important for adaptive systems as is the value of pedagogy and heuristics in both phases. It would be valuable to extend the evaluation to real life contexts and to consider the impact and outcomes of learners using the systems. In particular, this means applying evaluation criteria to a "gold standard" case to test for its utility on implementation and impact on learning.

Our future aim is to extend lessons learnt from current methods for a summative evaluation method, which considers the strengths of the evaluation criteria. That is, the inclusion of pedagogy as well as heuristics. In order to achieve this we need to further develop and apply both the heuristic and pedagogic criteria in the evaluation of learner/user outcomes and impact.

References

1. Abowd, G.D., Mynatt, E.D., Rodden, T.: The Human Experience.IEEE Pervasive Computing, Vol. 1, Jan-Mar, 48-57 (2002)
2. Baker, K., Greenberg, S. and Gutwin, C.: Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration. Proceedings of the 8th IFIP Working Conference on Engineering for Human-Computer Interaction (2001)
3. Brusilovsky, P., Karagiannidis, C., and Sampson, D.: Benefits of Layered Evaluation of Adaptive Applications and Services. In S. Weibelzahl, D. Chin and G. Weber (Eds.) Empirical Evaluation of Adaptive Systems. Proceedings of workshop held at the Eighth International Conference on User Modeling in Sonthofen, Germany, July 13th, 2001, 1-8, Freiburg: Pedagogical University of Freiburg (2001)
4. Brusilovsky, P.: Adaptive hypermedia. User Modeling and User Adapted Interaction, 11 (1/2), 87-110 (2001)
5. Chin, D. N. :Empirical evaluation of user models and user-adapted systems. User Modeling and User-Adapted Interaction, 11(1-2), 181–194 (2001)
6. Cristea, A.I., De Mooij, A., LAOS: Layered WWW AHS Authoring Model And their corresponding Algebraic Operators, In Proceedings of WWW'03 Budapest, Hungary, ACM, IW3C2, (2003a)
7. Cristea, A.I., De Mooij, A.: Evaluation of MOT, an AHS Authoring Tool: URD Checklist and a special evaluation class. International Conference on Computers and Advanced Technology in Education, Rhodos, Greece, 241-246 (2003b)
8. Cristea, A. I.: What can the Semantic Web do for Adaptive Educational Hypermedia? Educational Technology & Society, 7 (4), 40-58 (2004)
9. Cristea, A., Stewart, C.: Automatic Authoring of Adaptive Educational Hypermedia, book chapter to appear in "Web-Based Intelligent e-Learning Systems: Technologies and Applications", Zongmin Ma. IDEA Publishing group (2005)
10. Dagger, D., Wade, V., Conlan, O.: Developing Adaptive Pedagogy with the Adaptive Course Construction Toolkit (ACCT), published in the workshop proceedings of the AH2004 conference (online), Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH2004) Proceedings, Eindhoven, The Netherlands (2004)
11. Draper, S.W. Brown, M.I. Henderson, F.P McAteer, E.: Integrative evaluation:An emerging role for classroom studies of CAL [WWW document]. URL http://www.psy.gla.ac.uk/~steve/IE.html (visited 2007 Jan 10), (1996)
12. Gagné, R. M. The conditions of learning and theory of instruction, Fourth Edition. New York: Holt, Rinehart and Winston (1985)
13. Herder, E.: Utility-Based Evaluation of Adaptive Systems. In: Weibelzahl, S. and Paramythis, A. (eds.). Proceedings of the Second Workshop on Empirical Evaluation of Adaptive Systems, held at the 9th International Conference on User Modeling UM2003, Pittsburgh, pp. 25-30 (2003)
14. Höök, K.: Steps to take before intelligent user interfaces become real. Interacting with Computers, Vol. 12, No. 4, 409-426 (2000)
15. Joy, E. H. & Garcia, F. E.: Measuring Learning Effectiveness: A New Look at No- Significant-Difference Findings. Journal of Asynchronous Learning Networks, 4(1): 33- 39 (2000)
16. Kinshuk, Patel A., Russell D.: A multi-institutional evaluation of Intelligent Tutoring Tools in Numeric Disciplines. The Evaluation of Learning Technology Conference Proceedings (Ed. M. Oliver), University of North London, London, 38-40 (2000)
17. Lockee, B., Moore, M., Burton, J.: Old Concerns with New Distance Education Research. EDUCAUSE Quarterly, 24(2): 60-62 (2001)
18. Magoulas, G. D., Chen, S. Y. and Papanikolaou, K. A.: Integrating Layered and Heuristic Evaluation for Adaptive Learning Environments. In: Weibelzahl, S. and Paramythis, A. (eds.). Proceedings of the Second Workshop on Empirical Evaluation of Adaptive Systems, held at the 9th International Conference on User Modeling UM2003, Pittsburgh, pp. 5-14 (2003)
19. Nielsen, J.: Enhancing the explanatory power of usability heuristics. Proc.of CHI'94 Conference. 152-158 (1994a)

20. Nielsen, J.: Heuristic Evaluation. Usability Inspection Methods. Wiley, New York (1994b)
21. Okamoto T., Cristea A. & Kayama M.: Future integrated learning environments with multimedia. Journal of Computer Assisted Learning, 17, 4-12 (2001)
22. Phipps, R., Merisotis, J.: What's the Difference? A Review of Contemporary Research on the Effectiveness of Distance Learning in Higher Education. The Institute for Higher Education Policy, Washington, D.C. (1999)
23. Reigeluth, C. M. (Ed.): Instructional-design theories and models: An overview of their current status. Hillsdale NJ: Erlbaum (1983)
24. Reigeluth, C. M. (Ed.): Instructional-design theories in action: Lessons illustrating selected theories and models. Hillsdale NJ: Erlbaum (1987)
25. Scriven, M. The methodology of evaluation. In Gredler, M. E. Program Evaluation. (p. 16) New Jersey: Prentice Hall, (1996)
26. Weibelzahl, S. & Lauer C. U.: Framework for the evaluation of adaptive CBR-systems. In I. Vollrath, S. et al. (eds.), Experience Management as Reuse of Knowledge. Proc. 9th German Workshop on Case Based Reasoning, pp. 254-263. Baden-Baden, Germany (2001)
27. Weibelzahl, S., Lippitsch, S., Weber, G.: Advantages, opportunities, and limits of empirical evaluations: Evaluating adaptive systems. Künstliche Intelligenz, 3/02, 17-20 (2002)