Integrating Layered and Heuristic Evaluation for Adaptive Learning Environments

George D. Magoulas, Sherry Y. Chen, and Kyparissia A. Papanikolaou

[1] Department of Information Systems and Computing, Brunel University
Uxbridge, UB8 3PH, UK
Email: {Sherry.Chen, George.Magoulas}@brunel.ac.uk
[2] Department of Informatics and Telecommunications, University of Athens
Panepistimiopolis, GR-15784 Athens, Greece
Email: spap@di.uoa.gr

In this paper we argue about the need to develop an educational-evaluation model and a methodology that include usability testing as standard procedure capable to determine the impact of adaptation on learners' behavior in an educational environment. To this end, heuristic evaluation is modified and criteria that diagnose potential usability problems related to adaptation are introduced and then integrated into the layered evaluation framework.

1. Introduction

Adaptive Learning Environments (ALEs) increase the functionality of conventional educational hypermedia, reducing learners' cognitive overload and disorientation by combining free browsing with personalization . They can support all the continuum of learning modes, from pure system-controlled to full learner-controlled (Brusilovsky, 1995). A variety of techniques can be used for flexible delivery of course material, including adaptive presentation and adaptive navigation support (De Bra, Brusilovsky, and Houben, 1999).

Although, it is acknowledged that the design of ALEs should be based on scientific knowledge or theoretical frameworks related to learning, transferring pedagogical considerations into software design specifications is not straightforward. In ALEs, adaptive techniques usually focus on providing user with navigation flexibility, tailoring the content to learners' needs, goals, interests and preferences, and simplifying task execution. However, there is still limited knowledge as to the effectiveness of these techniques, which, in turn, influence the usefulness of ALEs. Arguably, this is due to the lack of systematic works for the evaluation of adaptive systems (Weibelzahl, 2001).

Evaluative works are recognized as critically important to validate existing approaches and to provide concrete prescriptions for developing learner-centered ALEs. Most current evaluative works mainly focus on a particular aspect, e.g. search time (e.g. Boyle and Encarnacion, 1994) or search scope (e.g. Höök, 2000). Recently, modular evaluation (Paramythis et al., 2001), layered evaluation (Karagiannidis, and Sampson, 2000), and the evaluation framework of Weibelzahl and Lauer (2001) have been proposed to assess the adaptation process. Although, these approaches acknowledge the importance of taking into account usability considerations, these are not their main targets when evaluating adaptive systems. Nevertheless usability issues are considered of great importance when introducing and evaluating educational technology, as they can have both positive and negative implications on the teaching and learning process. For example, usability problems can affect the student's cognitive processes and the accomplishment of educational objectives (Bourges-Waldegg et al. 2000). Lack of usability consideration may cause inconvenience for students to use an ALE, which, in turn, has impacts on the correctness of adaptivity. Thus, for example, the interface of an ALE is no longer considered as a mere superficial layer as differences in interface design can affect student learning experience and may lead to the construction of misconceptions and misunderstandings (Balacheff and Kaput, 1996).

In this vein, this paper attempts to integrate usability aspects into the layered evaluation of ALEs by introducing usability criteria that specially cater for the adaptation. Among the various usability inspection techniques, the primary technique used here is heuristic evaluation, which involves each interface being scrutinised against a set of recognized usability principles, or 'heuristics' (Nielsen and Molich, 1990). This technique is usually used to obtain a detailed evaluation of the whole interface and ensures that the entire problem space is covered. Research has shown that problems overlooked in user testing can be in fact identified through heuristic evaluation. Yet, there is a clear lack of empirical evaluation that applies such criteria to examine ALEs.

The paper begins by building a theoretical background to present the importance of evaluation of ALEs and the value of layered evaluation, and further discusses the effectiveness of heuristic evaluation. It then explains how heuristic evaluation is modified and integrated into the layered evaluation. Subsequently, a set of criteria is developed in order to assess the impacts of usability on user behaviors and consequently their impacts on the adaptation. Finally, the paper ends with a discussion about the value of these criteria for the development of ALEs and other adaptive web-based applications.

2. Theoretical Background

2.1 Evaluation of Adaptive Learning Environment

In the context of adaptive systems evaluation remains an open issue (Weibelzahl and Weber, 2002; Weber and Brusilovsky, 2001). Recently, Weibelzahl and Weber (2001) developed EASy-D, an online database for evaluation studies of adaptive systems. EASy-D can serve as a reference for the usefulness of adaptation techniques and adaptive systems in general, and for the comparison of evaluation studies, in terms of experimental designs, evaluation criteria, system functions, etc. When it comes to the evaluation of ALEs, although several empirical studies have been performed (Boyle and Encarnacion, 1994; Weber and Brusilovsky, 2001), it seems that there is no standard methodology, including designs and criteria, for evaluating ALEs. Existing attempts usually focus on studying learner's behaviour with and without adaptation, which does not seem appropriate particularly when adaptivity is an inherent property of the system (Höök, 2000), and do not take into account the different aspects of adaptation.

The concept of usability in ALEs is indelibly associated with the provision of meaningful adaptation. Usability evaluation identifies the lack of consistency that may be caused by the adaptive behavior of the environment and the difficulty of users to form a consistent mental model of the application and its content.

2.2 Layered Evaluation of Adaptive Systems

A number of researchers recommend evaluating adaptive systems on a layer by layer basis as a more comprehensive approach (Brusilovsky et al., 2001; Karagiannidis, and Sampson, 2000; Weibelzahl and Lauer, 2001; Weibelzahl and Weber, 2002). Current approaches, such as layered evaluation (Brusilovsky et al., 2001; Weibelzahl and Weber, 2002), and modular evaluation (Paramythis et al., 2001) mainly evaluate adaptive systems by deriving appropriate models or architectures of adaptation, which differ in the resolution degree of their decompositions or processing steps. In contrast to approaches that focus on the overall user's performance and satisfaction, (Chin 2001), layered evaluation in particular assesses the success of adaptation by decomposing it into different layers, and evaluating them one by one. The different layers reflect the various stages/aspects of the adaptation, starting from low-level input data acquisition or user monitoring, (Karagiannidis, and Sampson, 2000) and going up to high-level assessment of the behavioral complexity of the users (Weibelzahl and Lauer, 2001). This approach provides a series of advantages over previous attempts, such as useful insight into the success or failure of each adaptation stage separately, facilitation of improvements, generalization of evaluation results, and re-use of successful practices.

2.3 Heuristic Evaluation

Among a wide range of the techniques for usability evaluation, heuristic evaluation is a widely accepted method for diagnosing potential usability problems and is popular in both academia and industry (Baker et al., 2001) since it can be completed in a relatively short amount of time (i.e. a few hours). This methodology involves an expert evaluating the interface against a set of recognized usability principles - the 'heuristics' (Nielsen, 1994a). Heuristics are general rules used to describe common properties of usable interfaces (Nielsen, 1994b). They can help evaluators to discover sources of trouble, making detection of usability problems easier.

Ideally, an evaluator should have a broad background in usability evaluation and interface design as well as specific knowledge of the subject domain. However, it is also true that heuristic evaluation techniques provide little difficulty to those who want to use them, regardless of their prior knowledge in usability evaluation and subject domain. In fact, it is possible to use heuristic evaluation after only a few hours of training (Nielsen, 1995). The prescription of structured techniques can also facilitate effective evaluation by novices, more so than for other usability methods available (Levi and Conrad, 1996).

The other advantage of heuristic evaluation is that it can be used to address some of the gaps in past evaluation works, which either focused on a particular aspect of the system in detail or considered the entire system without enough depth (Baker et al., 2001). The use of heuristics ensures that both the entire system can be evaluated in depth and specific problems can be discovered at an early design stage before releasing a running prototype of a system (Fu, et al., 2002). This is potentially very useful in the context of ALEs as it allows gathering evidence about the appropriateness of our hypothesis before implementing adaptation. Heuristic evaluation was first formally described in presentations in the Human-Computer Interaction conference through papers published by Nielsen and Molich (1990). Since then, they have refined the heuristics based on a factor analysis of 249 usability problems (Nielsen, 1994a) to derive a revised set of heuristics with maximum explanatory power. The revised sets of 10 heuristics are as follows (Nielsen, 1994b):

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose and recover from errors
10. Help and documentation

3. An Integrated Methodology

3.1 Development of Detailed Criteria

Nielsen's (1994b) set of heuristics is used as a benchmark in this study, as their usefulness has already been studied and validated (e.g. Kahn and Prail, 1994). The set of heuristics, although very useful, is suitable for general usability studies, instead of particularly designed for pedagogy-based applications. In order to have more explicit criteria for evaluating an ALE, Nielsen's heuristics need to be modified to reflect pedagogical considerations. The proposed modification is governed by the previous research in the field of education and human-computer interaction (Chen and Ford, 1998). The modified heuristics are summarized below:

1. *Visibility of the ALE*, e.g. the adaptation of the system attracts user attention; the environment always informs learners what content is suitable for their current status; feedback and information on learner progress are provided to encourage reflection.
2. *Match between ALE and the real world*, e.g. the learning environment present information in a way that matches with each individual's learning preferences.

3. *Various levels of leaner control*, e.g. learners are free to develop their personal strategies, select and sequence tasks, activate or de-activate adaptivity feature, change their learning models.
4. *Consistency and standards*, e.g. the adaptive behavior does not alter the usage of the system, allowing learners to use the same interaction approach regardless of the adaptivity features and/or the learning model.
5. *Help users recognize, diagnose and recover from errors*, e.g. the system can help user to recover from error; users are not stuck when they make a mistake.
6. *Error prevention*, e.g. the appropriateness of adaptation decisions support learners to avoiding the errors.
7. *Recognition rather than recall*, e.g. instructions and cues that the system provides for users to identify results of adaptations easily.
8. *Flexibility and efficiency of use*, e.g. the system accommodates learners' preferences by providing alternative ways to accessing information quickly and experiment with various concepts.
9. *Aesthetic and minimalist design*, e.g. the system offers learners an enjoyable experience by adopting minimalist and aesthetical design.
10. *Searchable help functions and documentation*, e.g. help focuses on learner's activities currently undertaken and provides a list of concrete steps to be carried out.

Furthermore, to facilitate a detailed, structured and thorough evaluation, it was necessary to develop associated criteria for each heuristic. The criteria were developed by an analysis of past and present usability studies (e.g. Nielsen and Mack, 1994) and the empirical studies of adaptive hypermedia learning environments (Brusilovsky, et al., 1998; Papanikolaou et al. 2003). The findings of these studies were applied to identify the adaptivity aspects that might affect learners' performance or cause difficulties in interacting with the ALE and then develop appropriate criteria, which are associated with the modified heuristics.

3.2 Integration of Criteria into Layered Evaluation

This study adopts the layered evaluation framework proposed by Weibelzahl and Lauer (2001) because it breaks the process of delivering adaptivity into more detailed steps, including (1) correctness of input data acquisition, (2) correctness of inference, (3) appropriateness of adaptation decisions, (4) change of system behavior when the system adapts, (5) changes in user behavior when system adapts, and (6) change and quality of total interaction.

In order to present suitable criteria for each layer, the tasks within each layer were defined and their relationships with students' learning activities were also identified by integrating the learning models of Chen (2002) and Magoulas et al. (submitted). Subsequently, the criteria of heuristic evaluation were assigned into a suitable layer depending on the purpose of each criterion and its impact on students' learning activities. For example, one of the criteria for Heuristic No 6, Error Prevention, is that data inputs should be case-blind whenever possible. This criterion is assigned into the first layer, correctness of input data acquisition, because case-blind fields can reduce typing errors and increase the reliability of input data.

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

Heuristic	**Criterion**	**Purpose**
H6: Error Prevention	Data inputs are case-blind whenever possible.	Reduce the likelihood of errors
	When learners navigate between multiple windows, their answers are not lost.	Avoid data lost
	Buttons that can cause serious consequences are located far away from low-consequence and high-use keys.	Prevent errors
	The de-activation of the adaptivity always requires users' confirmations.	Ensure that the de-activation has not been done by mistake
	The change of learning model always requires learner confirmation.	Check if the learner has made an accidental change
	The input fields to identify the learner preferences are not set to default values.	Identify learners preferences correctly.
H7: Recognition rather than Recall	Required data entry fields are clearly marked	Attract learners' attention

Table 2: Criteria for layer 2: correctness of inference

Heuristic	**Criterion**	**Purpose**
H3: Various Levels of Learner Control	The same content is presented in various formats according to the learning profile.	Help learner to choose suitable formats and then identify learners learning strategies correctly.
	Multiple levels of explanation are applied to help students find the correct answers.	Make explanation understandable to learners with different levels of prior knowledge
	The system provides both basic and advanced subject content.	Allow learners to choose suitable concepts and then assess the level of their understanding correctly.
H6: Error Prevention	Users are prompted to confirm actions that have drastic and destructive consequences.	Make sure that the action was not done by mistake.
	When learners are taking assessments, they can change their previous answers.	Allow learners undo their actions.

4. Generation of guidelines

Based on the previous discussion, a framework of guidelines is generated that include formative criteria for each layer using the adapted heuristics presented in Section 3.1. Tables 1 to 4 presents the criteria associated with each layer

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

Heuristic	**Criterion**	**Purpose**
H2: Match between systems and the Real World	The relevant concepts are organized in an understandable way.	Easily identify concept relationships.
	Users can easily identify the logical order of the relevant concepts.	Avoid cognitive overload
	The arrangement of the menu items matches with learner's profile.	Allow menu items to be readily understandable
	The concepts are explained in a way that matches with learner's profile.	Information can be easily remembered.
	The explanation of the assessment result is meaningful to the learner.	Explanations are easily understandable
H4: Consistency and Standard	The layout is consistent for different learning models.	Help learners interact with the learning environments and form a consistent mental model of the application.
	Online instructions are displayed in a consistent location across different learning models.	Help learners locate instructions easily.
	There are not too many types of icons to identify the difficulty levels of the content.	Reduce the complexity of the mental task.

Thus by integrating heuristics into the layered evaluation framework, the success of adaptation in ALEs not only relates directly to the efficiency and effectiveness of the task execution but also to the support of expression, thinking, reflection and metacognitive activity (Vosniadou, 1994), i.e. with the effectiveness and efficiency of learning that should occur during interaction.

5. Discussion and Concluding Remarks

Heuristic evaluation has been proved to be a good approach for the evaluation of web-based applications. By integrating heuristic evaluation into layer evaluation, this paper proposed guidelines for the improvement of ALEs and diagnosis of design problems for each layer at an early design stage. This approach emphasizes on user's perspective as our goal is to develop successful adaptation that can meet learners' needs and promote their learning performance. We are currently investigating the applicability of our approach in ALEs. Future research will focus on conducting user testing to assess the robustness and reliability of the guideline suggested by this paper, which could ensure in turn the development of learner-centered ALEs.

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

Heuristic	**Criterion**	**Purpose**
H1: Visibility of Learning Environments	There is visual feedback to inform learners about their current progress.	Provide indication of the size and length of the subject content.
	After a task is completed, learners are informed about the next step.	Avoid disorientation.
	The system can let learners easily identify content that is not suitable for his/her current status.	Provide content tailored to learner needs.
	Adaptation always suggests a route through the program.	Avoid get lost in hyperspace.
H8: Flexibility and Efficiency of Use	It is easy to select relevant information directly.	Allow selecting relevant content tailored to learner's needs.
	Students are allowed not to read the subject from beginning to end.	Speed up the interaction.

Table 5: Criteria for layer 5: Change of system behavior when the system adapts

Heuristic	**Criterion**	**Purpose**
H5: Help users Recognise, Diagnose, and Recover from Errors	Error messages inform learners of errors severity.	Imply that learner has control
	Error messages let learners understand the cause of the problem.	Make learners reflect on how to correct errors
	When learner chooses a wrong concept, the system allows them to de-select it.	Ensure the appropriateness of adaptations.
	When learner chooses an unsuitable learning model, the system allows them to go back	Ensure system behavior is appropriate.
H3: Various Levels of Learner Control	The system allows learners to select/deselect the adaptivity.	Provide appropriate system features.
	The system allows learners to choose their own learning model.	Allow learners to set their own preferences

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

Heuristic	Criterion	Purpose
H9: Aesthetic and Minimalist Design	Menu items are brief, yet long enough to describe the subject content.	Help learners select relevant items.
	The title of each page is short, simple, and distinctive	Help learners identify the subject content of the current page.
	The color scheme matches with the preference of each individual student.	Enjoyment of working with the system.
H10: Help Functions and Documentations	There is an alphabetical index to search information within the help system.	Help learners find relevant instructions for a particular task.
	Learners are allowed to switch between the help system and their tasks.	Allow learners to do their tasks and consult the help facility at the same time.

References

Baker, K., Greenberg, S. and Gutwin, C. (2001) Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration. Proc. 8th IFIP Conf. Engineering for HCI.

Balacheff N. and Kaput J. (1996). Computer-based learning environments in Mathematics. International Handbook of Mathematics Education. Bishop A.J. et al. (eds.) Amsterdam: Kluwer, pp. 469-501.

Bourges-Waldegg, P., Moreno L., Rojano T. (2000). The role of usability on the implementation and evaluation of educational technology, Proc. 33rd Hawaii International Conference on System Sciences, pp. 1-7.

Boyle, C., and Encarnacion, A. O.(1994) An Adaptive Hypertext Reading System. User Modeling and User-Adapted Interaction 4(1) 1-19.

Brusilovsky, P. (1995). Intelligent tutoring systems for World-Wide Web. In R. Holzapfel, Poster proceedings 3rd International WWW Conference. Darmstadt, April 10-14, 42-45.

Brusilovsky, P., Karagiannidis, C., and Sampson, D. (2001). The Benefits of Layered Evaluation of Adaptive Applications and Services. In: S. Weibelzahl, D. Chin, G. Weber (eds.): Proceedings of UM2001 Workshop on Empirical Evaluation of Adaptive Systems, Sonthofen, Germany. Pedagogical University of Freiburg, 1-8.

Chen, S. Y and Ford, N. (1998) Modelling User Navigation Behaviours in a Hypermedia-Based Learning System: An Individual Differences Approach. International Journal of Knowledge Organization . 25(3) 67-78.

Chen, S. Y. (2002) A Cognitive Model for Non-linear Learning in Hypermedia Programmes . British Journal of Educational Technology. 33(4), 453-464

Chin D. (2001). Empirical evaluation of user models and user-adapted systems, User Modeling and User-Adapted Interaction, 11, 181-194.

De Bra, P., Brusilovsky, P., Houben, G.-J. (1999) Adaptive Hypermedia: From Systems to Framework. ACM Computing Surveys. 31(4es) (1999), Fu, L., Salvendy, G., and Turley, L. (2002) Effectiveness of user testing and heuristic evaluation as a function of performance classification. Behaviour & Information Technology. 21 (2002), 137-143

Höök K. (2000). Steps to take before intelligent user interfaces become real. Interacting with computers, 12, 409-426.

Kahn, M. J. and Prail, A. (1994) Formal usability inspections. In: Nielsen, Jakob ; Mack, Robert L(eds).: Usability Inspection Methods. 141-171.

Levi, M. D. and Conrad, F. G (1996) A Heuristic Evaluation of a World Wide Web Prototype. Interactions. 3 (1996) 50-61

Magoulas G.D., Papanikolaou K. and Grigoriadou M. (submitted) Adaptive web-based learning: accommodating individual differences through system's adaptation, British Journal of Educational Technology.

Nielsen, J. and Molich, R. (1990) Heuristic evaluation of user interfaces. Proceedings o] CHI'90 Conference on Human Factors in Computer Systems.249-256.

Nielsen, J. (1994a) Enhancing the explanatory power of usability heuristics. *Proc.of CHI'94 Conference.* 152-158.

Nielsen, J. (1994b) Heuristic Evaluation. Usability Inspection Methods. Wiley, New York.

Nielsen, J. (1995) Technology Transfer of Heuristic Evaluation and Usability Inspection. Paper presented at International Conference on Human-Computer Interaction.

Nielsen, J., and Mack, R. L. (1994) Usability inspection methods. New York, NY: John Wiley

Papanikolaou K., Grigoriadou M., Kornilakis H., and Magoulas G.D. (2003), Personalising the Interaction in a Web-based Educational Hypermedia System: the case of INSPIRE, User-Modeling and User-Adapted Interaction, forthcoming.

Paramythis, A., Totter, A., and Stephanidis, C., (2001). A Modular Approach to the evaluation of Adpative User Interfaces. In: S. Weibelzahl, D. Chin, G. Weber (eds.): Proceedings of the UM2001 Workshop on Empirical Evaluation of Adaptive Systems, Sonthofen, Germany. Freiburg: Pedagogical University of Freiburg, 9-24.

Vosniadou S. (1994). From cognitive theory to educational technology, Technology-based learning environments: psychological and educational foundations, Vosniadou S., et al. (eds.) NATO ASI Series F: Computer and System Sciences, 137, 11-18, Springer Verlag.

Weber, G. and Brusilovsky, P. (2001) ELM-ART: An adaptive versatile system for Web-based instruction. International Journal of Artificial Intelligence in Education 12 (4), Special Issue on Adaptive and Intelligent Web-based Educational Systems, 351-384.

Weibelzahl, S. and Lauer C. U.(2001). 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.

Weibelzahl, S. and Weber G. (2001). A database of empirical evaluations of adaptive systems. Proceedings of the Workshop "Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen" (ABIS 2001).

Weibelzahl, S. (2001) Evaluation of adaptive systems. In Proceedings of the Eighth International Conference, UM2001. (Lecture Notes in Computer Science LNAI 2109) 292-294.

Weibelzahl, S. and Weber, G. (2002). Advantages, opportunities, and limits of empirical evaluations: Evaluating adaptive systems. Künstliche Intelligenz, 3 (2).