HAL
open science

# An Approach for the Selection of Evaluation Methods for Interactive Adaptive Systems Using Analytic Hierarchy Process

Amira Dhouib, Abdelwaheb Trabelsi, Christophe Kolski, Neji Mahmoud

## To cite this version:

Amira Dhouib, Abdelwaheb Trabelsi, Christophe Kolski, Neji Mahmoud. An Approach for the Selection of Evaluation Methods for Interactive Adaptive Systems Using Analytic Hierarchy Process. 10th International Conference on Research Challenges in Information Science, RCIS 2016, Institute of Electrical and Electronics Engineers(IEEE), May 2016, Grenoble, France. pp.411-420, 10.1109/RCIS.2016.7549289 . hal-03350506

**HAL Id: hal-03350506**

**https://uphf.hal.science/hal-03350506v1**

Submitted on 14 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# An Approach for the Selection of Evaluation Methods for Interactive Adaptive Systems Using Analytic Hierarchy Process

Amira Dhouib
Miracl Laboratory
Faculty of Economics and Management Sciences
University of Sfax, B.P. 1088, Sfax 3000 Tunisia
amiradhouib@live.fr

Abdelwaheb Trabelsi
College of Computation and Informatics
Saudi Electronic University
Dammam, Saudi Arabia
abdelwaheb@gmail.com

Christophe Kolski
LAMIH-UMR CNRS 8201,
University of Valenciennes and Hainaut-Cambrésis,
Valenciennes, France
Christophe.Kolski@univ-valenciennes.fr

Mahmoud Neji
Miracl Laboratory
Faculty of Economics and Management Sciences
University of Sfax, B.P. 1088, Sfax 3000 Tunisia
mahmoud.neji@fsegs.rnu.tn

***Abstract*—The evaluation of interactive adaptive systems is the source of many difficulties for novice evaluators as well as for expert ones. Issues arise in the need to identify the appropriate evaluation methods to be used for a given evaluation context. This paper proposes an approach for the choice of evaluation methods for interactive adaptive systems with regard to their suitability depending on specific evaluation contexts. A multi-criteria decision making approach, namely the Analytic Hierarchy Process, is proposed for the selection methodology. The approach is based on a pair-wise comparison of a number of criteria that affect the choice of the appropriate evaluation methods. Four user-centred evaluation methods in conjunction with the layered evaluation for interactive adaptive systems are presented in the cases studies. These methods are: user-as-wizard, heuristic evaluation, focus group, and user test. The proposed approach is flexible in the sense that it can be used in the case of different evaluation methods. The results depend on the evaluation context where the methods will be used. Hence, we illustrate the proposed approach in the case of two evaluation contexts.**

***Keywords—interactive adaptive system; evaluation method; analytical hierarchy process; user-centred evaluation; layered evaluation***

## I. INTRODUCTION

The use of Interactive Adaptive Systems (IAS) has become increasingly important in recent years, with the growing application of these systems in many areas like education, museum, transport, e-commerce, etc. [1,2,3,4]. However, in practice, there are still many shortcomings and open questions about adaptive systems. Comprehensive and rigorous evaluation of interactive adaptive systems is important. It is essential not only to evaluate but also to ensure that the evaluation uses the suitable methods since an incorrect method can lead to wrong conclusions [5].

The evaluation of interactive adaptive systems is challenging due to the nature of adaptivity and the implications that adaptive systems have on interaction [6,7]. A variety of methods and approaches can be applied in order to evaluate these systems [8,9,10,11,12]. The problem of choice concerning evaluation methods is a source of many difficulties for the novice evaluators, and even for the expert ones. In fact, the diversity of evaluation methods involves a difficulty in the choice of the most appropriate ones for the evaluation of IAS depending on specific evaluation constraints. This raises the question, "Which evaluation method(s) is (are) the suitable for interactive adaptive systems in specific evaluation constraints and interaction contexts?" In order to address these issues, we propose in this paper a decision support approach for the choice of evaluation methods based on the Analytical Hierarchy Process (AHP). This multi-criteria decision making method has been applied to solve unstructured problems in a variety of decision-making situations, ranging from the simple personal decisions to the complex capital intensive decisions (For details see [13]).

Obviously, the task to decide whether to choose as appropriate evaluation methods for interactive adaptive systems in given evaluation contexts is not trivial [14]. The decision depends on multiple factors that are specific to the evaluation context. This paper aims to define the decision criteria that could be used to make a decision in the choice of suitable IAS evaluation methods depending on specific contexts. It illustrates also how the multi-criteria decision making approach AHP can be used to assist the decision-making process and then to guide the IAS evaluators.

One of the peculiarities that differentiate the evaluation of adaptive systems from that of the non-adaptive ones is the layered evaluation approach. The main idea behind this approach is to divide the adaptation into different layers, and evaluating them separately [15,16]. The different layers reflect

the various stages of the adaptation. Many authors recommend combining the layered evaluation and the User-Centred Evaluation (UCE) methods to get better evaluation results [17,18]. UCE methods have several advantages over other methods, for example: (1) they take the users explicitly into account, (2) they contribute to the quality of the adaptive systems, and (3) they allow detecting problems in the system functionalities [8,10]. Since we believe that the combination between UCE methods and the layered evaluation can become a key factor for a successful evaluation of interactive adaptive systems, in this paper, the focus will be on user-centred evaluation methods, and on relating these to the layered evaluation of IAS. In other words, the user-centred evaluation methods will be listed according to the different adaptation layers in which they can occur. For a feasibility study, only four evaluation methods will be considered (user-as-wizard [19], heuristic evaluation [20,21], focus group [22,23], and user test [24]). The proposed approach is flexible in the sense that it can be used in the case of other evaluation methods.

The remainder of this paper is organized as follows. We start with a review of the evaluation methods used for interactive adaptive systems; focusing on layered evaluation method and the four considered user-centred evaluation methods (user-as-wizard, heuristic evaluation, focus group, and user test) (Section 2).Then, we provide an overview of existing decision support systems in the field of the evaluation of regular (non-adaptive) systems (Section 3). Next, we provide a description of the proposed approach to support the decision making on evaluation methods selection, including the goal, the decision criteria and sub-criteria, the alternatives, and the AHP decision aid method (Section 4). Then, we present the decision support approach for the choice of appropriate evaluation methods based on the AHP method (Section 5). Following that, we illustrate two evaluation contexts of application in order to evaluate the feasibility of our approach (Section 6). Then, we discuss the obtained results (Section 7). Finally, we complete the paper with a conclusion and future work (Section 8).

## II. Evaluation Methods for Interactive Adaptive Systems

Our research question addresses the choice of appropriate evaluation methods, taking account of specific evaluation constraints (e.g., requirements resources, type of IAS, etc.). With a goal of feasibility study, the focus will be on four user-centred evaluation methods used in conjunction with the layered evaluation of IAS. The UCE methods will be listed then according to the different adaptation layers in which they can occur. The four UCE methods are: user-as-wizard, heuristic evaluation, focus group, and user test. These methods are thought to be representative. It should be noted that the proposed approach can be used in the case of different evaluation methods for interactive adaptive systems.

The following sub-sections present an overview of the layered evaluation and the four considered user-centred evaluation methods.

### *A. Layered Evaluation*

The important difference between the evaluation of adaptive and regular (non-adaptive) systems is that the evaluation of adaptive systems cannot consider the system as a whole. Many researchers recommend the use of the layered approach in assessing interactive adaptive systems [15,16,25]. This evaluation approach does not treat evaluation as a "monolithic" process but instead divide it into layers. Each layer of IAS reflects the stage/component of the adaptation; so it is evaluated independently of the others [15]. The layered evaluation helps in identifying the exact cause of the adaptation failure or any other error. The advantages of the layered evaluation are: (1) it provides insight into the success or failure of each stage of the adaptation, (2) it facilitates the improvements, and (3) it enhances the reusability of successful practices [16].

A series of layered evaluation approaches have been proposed in the IAS literature with different levels of granularity (layers). In each layer, different evaluations and analyses have to be taken into account. The layered evaluation approach suggested by [25] discerns two layers:

- Interaction assessment layer: This layer tests the validity of the conclusions drawn by the system concerning the characteristics of the user-computer interaction.
- Adaptation decision making layer: This layer evaluates the validity of the rules applied for the adaptation.

Another approach was presented by Weibelzahl [26]; this author identifies the following components/stages of adaptation:

- Evaluation of input data: This layer evaluates the reliability and external validity of the input data acquisition process, user behavior, and context (e.g., sensor data, click stream, etc.).
- Evaluation of the inference mechanism: This layer tests the validity of inference mechanism in different environments under real world conditions.
- Evaluation of the adaptation decision: The idea of the evaluation deals with determining whether adaptation decisions made are optimal.
- Evaluation of the total interaction: This layer evaluates both system and user under real world conditions.

A similar layered evaluation but with a greater degree of granularity was proposed by Paramythis *et al.* [15]; the following layers are distinguished:

- Collection of input data: The input data that an interactive system collects is predominantly derived from the user's interaction with it.
- Interpretation of the collected data: This layer evaluates the validity of the interpreted input data.
- Modelling the current state of the world: This layer compares the system-maintained dynamics model with the real-world entities to which they correspond.

- Deciding upon adaptation: At this layer, the adaptive system decides which adaptation theory/strategy to apply given the current user model.
- Applying adaptation: The adaptation decision can be applied in different ways (e.g., layouts, formulations).

In [27], a layered approach which would be applied for the evaluation of adaptive recommender systems is proposed. Two layers are identified:

- Evaluation of user modeling: Refers to the evaluation of the user modelling process, focusing mostly on whether the user characteristics are represented.
- Evaluation of adaptation decision making: Addresses the evaluation of the adaptation process, logic and results, focusing mostly on whether the personalization actions are valid and meaningful for the given state of the user model.

### B. *Relevant User-Centred Evaluation Methods: User-as-Wizard, Heuristic Evaluation, Focus Group, and User Test*

The user-centred evaluation methods for interactive adaptive systems have been recommended by several researchers, where all the evaluation phases can provide feedback to modify the knowledge base of the system itself [8,28]. The UCE methods help evaluators to detect the real problems encountered by users at the time of the execution of their task with the system [18]. Many user-centred evaluation methods have been identified in IAS literature (e.g., interviews, focus group, heuristic evaluation, think aloud protocols, expert review, parallel design, cognitive walkthrough, wizard of oz simulation, questionnaires, scenario-based design, task analysis, etc.) [8,10,11]. As already mentioned, for a feasibility study we focus on four UCE methods related to the adaptation layers of adaptive systems. The proposed approach can be used for different IAS evaluation methods. The selected UCE methods are: user-as-wizard, heuristic evaluation, focus group, and user test. The following sub-sections present the four evaluation methods.

*1) User-as-wizard:* In this method, participants play the role of the wizard and they are left completely free to perform the wizard's task without being given a script to follow [19]. The user-as-wizard method is inspired by the *wizard-of-Oz* and the *contextual design* methods [19]. This method consists of two stages: exploration stage and consolidation stage. In the first stage, the exploration one, participants take the role of the adaptive system, or most frequently of a functionality that corresponds to a particular layer. In the second stage, the consolidation one, a verification of the acceptability of the human performance is made in order to determine in what respects this performance can be improved. For this stage, it is recommended to use new participants, rather than reusing those of the exploration stage.

*2) Heuristic evaluation:* One or several expert evaluators examine a user interface using a set of criteria and look for problems that violate some of the general principles of good interface design [20]. Heuristics are general principles that can describe common properties of usable interfaces [29]. They help evaluators to discover sources of trouble of usability problems more easily. The most popular heuristics in usability testing are Nielsen's heuristics [29] consisting of ten broad guidelines based on a factor analysis of common usability problems. In [21], Magoulas and his colleagues proposed an integration of Nielsen's heuristics with layered evaluation for adaptive learning environments. For each level of adaptation, they modified a subset of the usability heuristics and added more detailed criteria for these heuristics.

*3) Focus group:* It is an informal technique structured as a discussion about specific topics moderated by a trained group leader [22]. It is used to collect user's attitudes, beliefs and desires during the requirement analysis phase and before system implementation [23]. The focus group method is considered to produce rich qualitative data about what users want and dislike. It is typically used to gather functional requirements, data requirements, usability requirements, and environmental requirements to be considered in the design of system adaptations, depending on the persons involved (e.g., users, evaluators).

*4) User test:* This evaluation method that can be carried out once the functionality corresponding to an evaluation layer has been implemented. In this method, users are given well-defined tasks to do. The user test method can be used to identify the users' performance (e.g., which adaptations they preferred, which confused them) and opinions [24]. The main difficulty of user test method is that it may be hard for participants to provide the kind of input required, necessitating the presence of special interactive facilities to support the process.

In order to contribute to the evaluation methods selection, only a few numbers of decision support systems and approaches are proposed in the literature. These decision support systems are proposed in the field of the regular (non-adaptive) systems and none of them are used in the case of interactive adaptive systems. A brief summary of these decision support systems is presented in the section below.

## III. A Survey of Existing Decision Support Systems in the Field of the Evaluation of Regular (Non-Adaptive) Systems

The scientific literature presents few decision support systems and approaches that can assist towards the selection of methods for the evaluation of regular (non-adaptive) systems [30,31,32,33]. For example, the planning aid proposed by [30] which offers advice in the form of heuristics about the selection of evaluation methods for the inexperienced evaluators. On the other hand, a decision support system, called Adhesion, was presented by [31]; it helps evaluators to identify the appropriate methods and criteria for the evaluation of human-computer interfaces, based on the fuzzy approach and the concept of expert's preferences. Another tool, called Usability Planner, was proposed by [32,33], to support the choice of user-centred design methods to be applied in a specific project, based on rules derived from ISO TR 16982 [34] and authors' experience. It helps to identify the particular methods to be selected according to the particular project stages and to estimate the

relative cost benefits of applying usability methods at different phases of interactive systems. Similarity with our approach is that both of them are used to support the choice of suitable methods depending on particular selected constraints. However, the particularity of our proposed approach is that it helps evaluators to choose the suitable evaluation methods dedicated to interactive adaptive systems, using a multi-criteria decision making method, namely the AHP.

The mentioned decision support systems were proposed to be used for regular (non-adaptive) systems. The existing studies did not report which evaluation methods to use for specific evaluation context of interactive adaptive systems. As no studies have addressed these issues before, our objective is to propose a decision support approach which will give a preliminary guidance for the evaluation of interactive adaptive systems. To our knowledge, such an approach does not have an equivalent in the evaluation of interactive adaptive systems field.

The proposed decision support approach for the choice of appropriate evaluation methods is described in detail in the next section.

## IV. Description of the Proposed Decision Support Approach

Choosing the appropriate evaluation methods is an essential decision for the evaluators; yet such a decision can be confounding with the numerous factors that need to be considered. In IAS literature, there is limited knowledge as to which evaluation methods are appropriate for different interaction contexts. IAS evaluators need to collect information present in different sources and to understand the suitability of each method in a particular evaluation context [14]. The objective of this work is to overcome this kind of weakness by proposing a decision support approach for the choice of suitable evaluation methods for interactive adaptive systems. Its main objective is to propose the appropriate methods which are usable for the evaluation of IAS in specific evaluation contexts. The principles on which it is based are described below (Fig. 1).

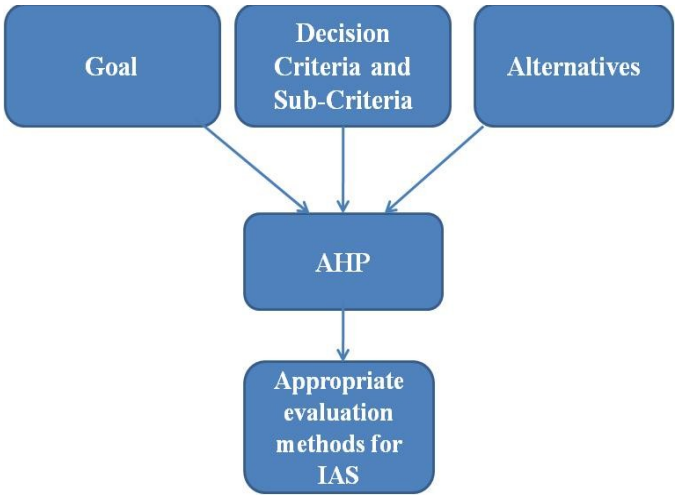


Fig. 1. Decision process for the choice of appropriate evaluation methods for interactive adaptive systems

The following sub-sections discuss the components of the above decision process for the choice of IAS evaluation methods.

### A. Goal

As already mentioned, the goal of the proposed decision support approach is to support the choice of the appropriate evaluation methods for IAS depending on specific evaluation constraints and interaction contexts.

### B. Decision Criteria and Sub-Criteria Affecting the Evaluation Methods Choice

The choice of the appropriate set of evaluation methods depends partially on a number of factors which can affect the evaluation process of adaptive systems. The factors considered in the decision making process, are related to: temporal, financial and human resources, location, evaluation phases, type of IAS, types of adaptation, adaptation layers, complexity of task, and user characteristics (i.e. number, expertise with the IAS and age).

Table 1 presents the decision criteria and sub-criteria and their corresponding explanations.

TABLE I. Description of Criteria and Sub-Criteria Affecting the Evaluation Methods Choice

| Decision Criteria and Sub-Criteria | Explanations |
|---|---|
| Temporal resources | Determines the time available for conducting the evaluation of interactive adaptive systems. |
| Financial resources | Determines the budget available for conducting the evaluation of interactive adaptive systems. |
| Human resources | The individuals involved in the evaluation process [35]. The evaluation of interactive adaptive systems can either involve evaluators, users, and/or other (e.g., designer). |
| Location | Determines the location where the evaluation is carried out, either a laboratory setting or real-life setting [8]. |
| Evaluation phases | According to [9], we distinguish three life-cycle stages in which an evaluation method can occur: (1) the *requirement phase*, which occurs before any system implementation, (2) the *preliminary evaluation phase*, which occurs during the system development, and (3) the *final evaluation phase*, which occurs at the end of the IAS implementation. |
| Type of IAS | Interactive adaptive systems can be classified according to their categories. Many types of adaptive systems are identified in the literature (e.g., adaptive hypermedia system, adaptive e-learning systems, adaptive recommender system, adaptive information retrieval system, etc.) [36]. |
| Type of adaptation | Three levels of adaptivity can be identified in IAS: (1) *user selection*: the system initiates the action and presents the user with the preferred alternative. The user has then to specify the action, (2) *user approval*: |

| | | |
|---|---|---|
| | | the system initiates the action and notifies the user about the selected alternative. The user has then to approve the system's choice or to choose another alternative, and (3) *fully adaptive*: the system initiates the action and did not presents the user with other alternatives [37]. |
| Adaptation layers | | The layer which needs to be considered for the evaluation of IAS. According to Paramythis *et al.* [15] an evaluation method can be assigned to one or more of the following: (1) collection of input data (CID), (2) interpretation of the collected data (ID), (3) modelling of the current state of the world (MW), (4) deciding upon adaptation (DA), and (5) applying adaptation (AA). The evaluation methods can also be used for the evaluation of adaptation as whole. |
| Complexity of task | | Certain IAS evaluation methods are applicable for simple tasks. Other methods are applicable for complex tasks [38]. |
| User charactestrics | Number | Number of users involved in the IAS evaluation process. |
| | Expertise with IAS | The interactive adaptive system can be used by various types of users who differ in terms of level of expertise with the system. This can influence the interaction with the IAS and needs to be considered when evaluating it [38]. |
| | Age | The evaluation might consider the properties of IAS used by various types of users differing in age [38]. |

As in the case of the regular (non-adaptive) systems a number of criteria are to be considered while choosing the appropriate evaluation methods (e.g., location, temporal and financial resources, phases of evaluation, etc.). However, in the case of interactive adaptive systems, additional criteria can influence the interaction with the adaptive systems and should be considered in the decision making process (e.g., type of adaptation, type of IAS). Considering these criteria, there must be a selection between evaluation methods. In this study, the four selected UCE methods for adaptive systems are the alternatives. The latter can be defined as the objects, between which the choice will be made. In the next sub-section, we present a classification of the considered evaluation methods.

### *C. Alternatives: Classification of Evaluation Methods for Interactive Adaptive Systems*

There is a wide variety of methods that contribute to the evaluation of interactive adaptive systems [9,11,12]. Each evaluation method can be defined by its particular aspects. For a feasibility study, the focus will be on four UCE methods, which represent the alternatives in this study. These evaluation methods are: user-as-wizard, heuristic evaluation, focus group, and user test (See Section II.B). In this paper, we classify UCE methods on the basis of phases of evaluation, quality factors assessed in each layer, etc. This classification is intended to capture the characteristics of the considered evaluation methods.

Table 2 presents a classification of the considered evaluation methods for IAS according to the following factors:

- The adaptation layer: The layer which the method is particularly suitable for [15].
- The inspected quality factors: The types of evaluation measures the methods are most suitable to evaluate in each layer (e.g., appropriateness of adaptation, comprehensiveness, precision) [15].
- The phases of evaluation: Three life-cycle stages in which an evaluation method can occur [9]: (1) the requirement phase, (2) the preliminary evaluation phase, (3) the final evaluation phase.
- The stakeholders: Evaluation methods differ in terms of the persons involved in the evaluation process [35].
- The advantages and disadvantages: The main advantages and disadvantages deriving from the application of each method in the evaluation of interactive adaptive systems [12,15].

The four UCE methods are often selected for the simple reason that they support the different evaluation phases and adaptation layers of IAS. The preliminary evaluation phase is represented by three methods which are: user-as-wizard, heuristic evaluation, and focus group. The requirement phase is represented by focus group, and the final phase is represented by user test. These UCE methods seem then to be appropriate for making a decision based on the characteristics of evaluation methods.

In order to determine the appropriate evaluation methods depending on specific evaluation settings, we use the multi-criteria decision making method AHP due to two reasons. First, it allows making qualitative and quantitative analyses in the same decision-making methodology and provides a useful way to evaluate consistency of evaluation measures, which reduces inconsistencies in the decision-making procedure [39]. The second reason is its simplicity since there is no need to build a complex expert system with the decision maker's knowledge embedded.

The following sub-section describes the AHP aid method.

TABLE II. CLASSIFICATION OF EVALUATION METHODS CONSIDERED IN THIS STUDY

| Evaluation Methods for IAS | Adaptation Layers | Quality Factors Assessed | Phases of Evaluation | Stakeholders | Advantages and Disadvantages |
|---|---|---|---|---|---|
| User-as-Wizard | MW* | Validity of interpretations or inferences, predictability, scrutability, conciseness, comprehensiveness, precision, sensitivity | Preliminary evaluation phase | Users or evaluators | (+) Can even be done before the layer itself has been implemented<br>(+) May inspire design of the layer<br>(–) Requires task that humans understand<br>(–) Some tasks are inherently more difficult for humans than for computers |
| | DA* | Necessity of adaptation, appropriateness of adaptation, subjective acceptance of adaptation, predictability, scrutability, breadth of experience | | | |
| | AA* | Usability, timeliness, unobtrusiveness, controllability, acceptance by user, predictability, breadth of experience | | | |
| Heuristic Evaluation | CID* | Accuracy, latency, sampling rate | Preliminary evaluation phase | Evaluators | (+) Quick and inexpensive<br>(+) Making prediction about the interface design without involving users<br>(–) Evaluators are not real users<br>(–) Need to choose the appropriate heuristics |
| | ID* | Validity of interpretations, privacy, controllability, predictability, scrutability, transparency | | | |
| | MW* | Validity of interpretations or inferences, privacy, scrutability, conciseness, predictability, comprehensiveness, precision, sensitivity, transparency, controllability | | | |
| | DA* | Necessity of adaptation, appropriateness of adaptation, subjective acceptance of adaptation, predictability, scrutability, breadth of experience, privacy, transparency, controllability | | | |
| | AA* | Usability, unobtrusiveness, privacy, timeliness, breadth of experience, acceptance by user, predictability, transparency, controllability | | | |
| | Whole | Effectiveness, efficiency, usability, criteria specific for system's objectives | | | |
| Focus Group | MW* | Validity of interpretations or inferences, scrutability, predictability, conciseness, comprehensiveness, precision, sensitivity, privacy, transparency, controllability | Requirement evaluation phase, preliminary evaluation phase | Users, evaluators, designers | (+) Can generate large amounts of data quickly<br>(+) Can discuss events happening over long time span<br>(–) Can only cover a few topics (or it will time-consuming)<br>(–) Depends on good moderator |
| | DA* | Necessity of adaptation, appropriateness of adaptation, subjective acceptance of adaptation, predictability, scrutability, breadth of experience, privacy, transparency, controllability | | | |
| | AA* | Usability, timeliness, unobtrusiveness, controllability, acceptance by user, predictability, breadth of experience, privacy, transparency | | | |
| User Test | DA* | Necessity of adaptation, appropriateness of adaptation, subjective acceptance of adaptation, predictability, scrutability, breadth of experience, privacy, transparency, controllability | Final evaluation phase | Users, evaluators | (+) Can provide objective performance measures<br>(+) Can be quite natural for users<br>(–) Requires tasks that humans understand<br>(–) Can be time-consuming |
| | AA* | Usability, timeliness, unobtrusiveness, controllability, acceptance by user, predictability, breadth of experience, privacy, transparency | | | |
| | Whole | Effectiveness, efficiency, usability, criteria specific for system's objectives | | | |

* CID = Collect Input Data, ID = Interpret Data, MW = Model the current state of the World, DA = Decide upon Adaptation, AA = Apply Adaptation

### *D. Analytic Hierarchy Process*

The Analytic Hierarchy Process (AHP) proposed by Saaty [40], is a multiple criteria decision making method applied to uncertain decision problems with multiple criteria. It can be used to solve complex decision problems. Analytic hierarchy process is one of the most widely used multiple criteria decision-making method [41]. Many works have been published based on AHP. They include applications of AHP in different domains such as choosing best alternatives [42], prioritization [43], and benchmarking [44], etc. One advantage of AHP is that it is simple because there is no need of building a complex expert system with the decision maker's knowledge embedded in it. AHP also allows making qualitative and quantitative analyses by the same decision-making

methodology [39]. The AHP method is based on six essential steps [39]: (1) defining the unstructured problem and stating the goal clearly, (2) decomposing the problem into a hierarchy tree of different decision elements: goals, criteria, sub-criteria and alternatives, (3) computing the weights for the different criteria through a pair-wise comparison; the result of the pair-wise comparison can be summarized in an (N_N) decision matrix A, as in (1), where N is the number of evaluation criteria considered. Every element of the matrix A represents the quotient of weights of the criteria. The diagonal is always 1 since Cij = 1 (since the criteria, i (row) are being compared to themselves, j (column)) and the lower triangular matrix is filled using Equation (2).

$$A=\begin{bmatrix} 1 & C_{12} & \dots\dots & C_{1N} \\ C_{21} & 1 & \dots\dots & C_{2N} \\ \dots\dots & \dots\dots & 1 & \dots\dots \\ C_{N1} & C_{N2} & \dots\dots & 1 \end{bmatrix} \quad (1)$$

$$C_{ij} = \frac{1}{C_{ji}} \quad (2)$$

The relative importance between two criteria is measured according to Saaty's discrete nine-value scale method shown in Table 3, The three other steps are: (4) determining the priority weights of alternatives with respect to each criterion (comparison of various alternatives with respect to the decision criteria for selection), (5) enumerating the overall priority weights for all the alternatives, (6) analyzing and checking the *Consistency Ratio* (CR) in order to validate and determine the acceptance of the weights. On the basis of Saaty's empirical suggestion [39] that a CR <0.1 is acceptable, it is concluded that the foregoing pair-wise comparisons to obtain attribute weights are reasonably consistent. In contrast, if the CR value is larger than the acceptable value, the matrix results are inconsistent and the judgments should be reviewed.

For more information on how to use the AHP aid method see [39].

TABLE III. THE PAIR-WISE COMBINATION SCALE

| Value | Description |
|---|---|
| 1 | Equal importance |
| 3 | Moderate importance |
| 5 | Strong importance |
| 7 | Very strong importance |
| 9 | Extreme importance |
| 2,4,6,8 | Intermediate values between two adjacent judgments |

## V. SUPPORTING THE CHOICE OF EVALUATION METHODS FOR INTERACTIVE ADAPTIVE SYSTEMS BASED ON THE AHP

Fig. 2 illustrates the hierarchy of decision for the choice of evaluation methods for a specific evaluation context which has four different levels. The overall goal of the decision problem is that the "Choice of appropriate evaluation methods for interactive adaptive systems" is put at the top level of the hierarchy (level 1). The middle level in the hierarchy explains the group of criteria to take into consideration (level 2); the user characteristics criterion was broken down into three sub-criteria (level 3). The last level of the hierarchy's decision shows the alternatives which are the evaluation methods for interactive adaptive systems (level 4). In this study, the alternatives are: user-as-wizard, heuristic evaluation, focus group, and user test.

In the next section, we test the proposed decision support approach in two different evaluation contexts. The goal is to find the suitable UCE methods in the case of specific evaluation contexts.

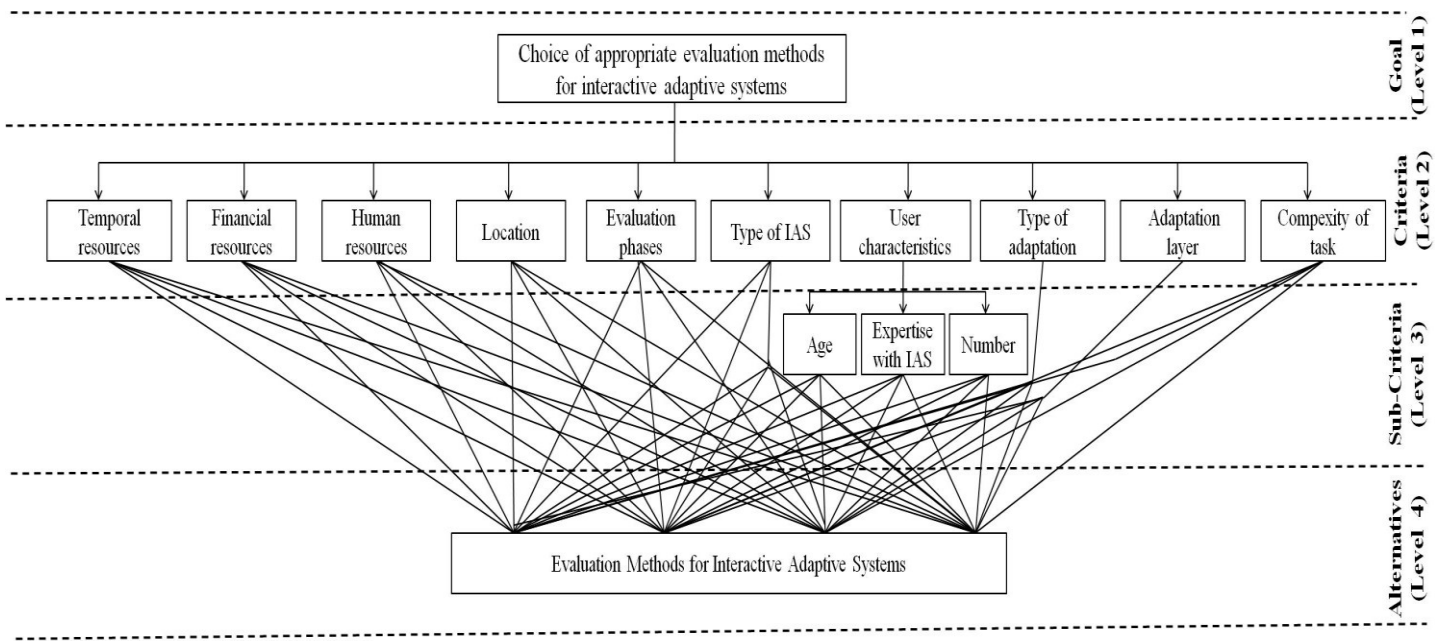


Fig. 2. AHP hierarchy for the choice of appropriate evaluation methods for interactive adaptive system

## VI. Cases Studies of AHP on Adaptive Hypermedia Systems

In order to illustrate the practical effectiveness of the proposed approach, we demonstrate the use of the AHP aid method by comparing the four considered UCE methods. We illustrate the proposed approach in case of two evaluation contexts. Different results are proposed according to the different evaluation constraints. The relative importance of criteria depends on the context in which the evaluation method will be used.

### *A. First Evaluation Context*

In this study, we compare the four selected evaluation methods in terms of two evaluation contexts. The first context describes an adaptive hypermedia system which offers personalized information adapted to the preferences of users. The evaluator needs to decide whether to choose as appropriate methods for the usability evaluation of the given adaptive system. In this evaluation context, a range of criteria are taken into account in making the decision: the evaluation methods can occur in the preliminary evaluation phase of the development process with the presence of users and medium temporal resources. The evaluation methods are used to evaluate the whole adaptive hypermedia system.

Table 4 shows the scale ranking for the pair-wise comparison that has been obtained from the evaluator. Associated to each cell, there is a row criterion, call it X, and a column criterion, call it Y. The evaluator supposed to compare each criterion with others according to their perception of the importance of each one. Hence, the research question is: "How much more important is X than Y in choosing the appropriate evaluation methods to use?" These values can vary from an evaluation method to another according to evaluation constraints. Below, we will explain the choice of some weights presented in Table 4. Let us consider the phases of evaluation and the human resources criteria; we give 1/2 value to the phases of evaluation compared to the human resources, which means that the human resources is considered twice more important than phases of evaluation criterion. This is the same reason for choosing 1/2 as weight for human resources/adaptation layer and the type of IAS/adaptation layer. A decision criterion compared to itself is assigned the value 1. The main diagonal elements of the pair-wise comparison matrix are all 1.

TABLE IV. Comparison Matrix of the First Evaluation Context

| | TR* | PE* | HR* | T-IAS* | AL* | Weight |
|---|---|---|---|---|---|---|
| TR* | 1 | 1/3 | 1/4 | 1/3 | 1/3 | 0.07 |
| PE * | 3 | 1 | 1/2 | 1/2 | 1/2 | 0.147 |
| HR* | 4 | 2 | 1 | 1/2 | 1/2 | 0.205 |
| T-IAS* | 3 | 2 | 2 | 1 | 1/2 | 0.248 |
| AL* | 3 | 2 | 2 | 2 | 1 | 0.33 |

*TR= Temporal Resource, PE= Phases of Evaluation, HR= Human Resources, T-IAS= Type of IAS, AL= Adaptation Layers

The value of weight of temporal resource criterion (0.07) presented in Table 4, is obtained from the average value of pair-wise comparison matrix row with normalized criteria matrix, the first row with a value of 0.35 (sum of the normalized matrix criteria of the first row) divided by the number of criteria that is five so obtained the results of 0.07. The same way done also on phases of evaluation, human resources, type of IAS, and adaptation layers criteria to determine the normalized matrix for criteria by means of matrix values divided by the number of criteria.

Then, the evaluator has to compare all the alternatives with respect to each criterion. Each matrix compares each criterion with each alternative. For example, the second matrix (Table 5) presents a comparison of temporal resources criterion with the alternatives (user-as-wizard, heuristic evaluation, focus group, and user test).

TABLE V. Evaluation of Alternatives with Respect to Temporal Resources Criterion

| Temporal Resources | User-as-Wizard | Heuristic Evaluation | Focus Group | User Test | Weight |
|---|---|---|---|---|---|
| User-as-Wizard | 1 | 1/3 | 1/2 | 2 | 0.17 |
| Heuristic Evaluation | 3 | 1 | 2 | 3 | 0.45 |
| Focus Group | 2 | 1/2 | 1 | 2 | 0.26 |
| User Test | 1/2 | 1/3 | 1/2 | 1 | 0.12 |

According to the temporal resources criterion, the most suitable evaluation method found is heuristic evaluation (0.45), and respectively focus group (0.26), user-as-wizard (0.17), and user test (0.12). After entered the judgments of temporal resources criterion, the evaluator has to enter the judgments for the rest of criteria.

AHP uses a *Consistency Ratio* to validate and determine the acceptance of the weights. The CR is acceptable, if it does not exceed 0.10. If it is more, the judgment matrix is inconsistent [40]. The consistency ratio resulting from our measure is 0.049 which is considered acceptable.

The comparative analysis resulting from AHP evaluation is depicted in Table 6. In conclusion, the result shows that heuristic evaluation is the appropriate evaluation method (0.278) for the evaluation of the adaptive hypermedia system in the given constraints. The evaluation results put user test (0.258), focus group (0.248), and user-as-wizard (0.216) in the second, third and fourth positions, respectively.

TABLE VI. Weighting Coefficient for Alternatives

| | TR* | PE* | HR* | T-IAS* | AL* | Weight |
|---|---|---|---|---|---|---|
| User-as-Wizard | 0.17 | 0.308 | 0.2 | 0.25 | 0.167 | 0.216 |
| Heuristic Evaluation | 0.45 | 0.308 | 0.142 | 0.25 | 0.333 | 0.278 |
| Focus Group | 0.26 | 0.308 | 0.329 | 0.25 | 0.167 | 0.248 |
| User Test | 0.12 | 0.076 | 0.329 | 0.25 | 0.333 | 0.258 |

*TR= Temporal Resource, PE= Phases of Evaluation, HR= Human Resources, T-IAS= Type of IAS, AL= Adaptation Layers

### *B. Second Evaluation Context*

We take here an example of another adaptive hypermedia system. In this evaluation context, a range of criteria are taken into account: the evaluation methods can occur in the

requirement phase with the presence of users and evaluators and high temporal resources. In this context, the methods are used to evaluate the applying adaptation layer of the adaptive hypermedia system.

Table 7 presents the pair-wise comparison relative to this evaluation context. The evaluator is supposed to compare each criterion with others according to his/her perception of the importance of each one.

TABLE VII. COMPARISON MATRIX OF THE SECOND EVALUATION CONTEXT

| | AL* | TR* | HR* | T-IAS* | PE* | Weight |
|---|---|---|---|---|---|---|
| AL* | 1 | 3 | 2 | 1/2 | 1/2 | 0.193 |
| TR* | 1/3 | 1 | 1/3 | 1/3 | 1/3 | 0.075 |
| HR* | 1/2 | 3 | 1 | 1/2 | 1/2 | 0.149 |
| T-IAS* | 2 | 3 | 2 | 1 | 1/2 | 0.25 |
| PE* | 2 | 3 | 2 | 2 | 1 | 0.333 |

*AL= Adaptation Layers, TR= Temporal Resources, HR= Human Resources, T-IAS= Type of IAS, PE= Phases of Evaluation

The obtained consistency ratio in this example is equal to 0.047 which is considered acceptable. The final matrix presented in Table 8 shows the decision which should be taken according to this evaluation context. With this study, it is deduced that the focus group (0.348) was the best evaluation method for the given constraints, followed respectively by the user test (0.235), user-as-wizard (0.22), and heuristic evaluation (0.197). This result was obtained by using the AHP method.

TABLE VIII. WEIGHTING COEFFICIENT FOR ALTERNATIVES

| | AL* | TR* | HR* | T-IAS* | PE* | Weight |
|---|---|---|---|---|---|---|
| User-as-Wizard | 0.25 | 0.412 | 0.167 | 0.25 | 0.16 | 0.22 |
| Heuristic Evaluation | 0.25 | 0.108 | 0.167 | 0.25 | 0.16 | 0.197 |
| Focus Group | 0.25 | 0.187 | 0.333 | 0.25 | 0.52 | 0.348 |
| User Test | 0.25 | 0.293 | 0.333 | 0.25 | 0.16 | 0.235 |

*AL= Adaptation Layers, TR= Temporal Resources, HR= Human Resources, T-IAS= Type of IAS, PE= Phases of Evaluation

## VII. DISCUSSION

In this paper, we used AHP method to determine the suitable evaluation methods for a given evaluation context. In the first context, where the evaluator has to choose the appropriate methods to evaluate the whole adaptive hypermedia system, heuristic evaluation is found to be the most appropriate evaluation method addressing the given situation, while user test is the second suitable method, focus group is the third method, and user-as-wizard is the worst evaluation method. According to the obtained results in the second evaluation context, the focus group was the best, while the user test is the second; the user-as-wizard and heuristic evaluation are the less suitable methods.

The evaluation methods proposed are different in the two evaluation contexts; this depends on the decision criteria, the evaluator's perception of importance of each criterion, and the alternatives considered in the decision making. In fact, the cases studies show that different evaluation methods are suitable depending on the given evaluation constraints, although these methods may be not appropriate for other contexts. It should be noted that the choice of the appropriate evaluation methods for IAS is strongly linked to the context within which it is applied.

As already mentioned, for a feasibility study, only four evaluation methods are considered as alternatives despite the fact that the approach can be used in the case of different evaluation methods for IAS. It should be noted that, a larger number of evaluation methods requires a high number of judgments in the comparison matrices. Although every single evaluation is very simple, the task of comparison may become time consuming.

The main objective of the proposed approach is to give a preliminary guidance for IAS evaluators to support the choice of appropriate methods depending on specific evaluation contexts. The selection of appropriate evaluation methods depends, naturally, on decision criteria and alternatives considered in decision making. In this study, certain decision criteria are determined: temporal, financial and human resources, location, evaluation phases, type of IAS, type of adaptation, complexity of task, user characteristics, and adaptation layers. The user characteristics criterion is broken down into three sub-criteria, which are: age, number, expertise with the IAS.

## VIII. CONCLUSION AND FUTURE WORK

A major problem which evaluators are confronted with is the choice of suitable evaluation methods for interactive adaptive systems depending on specific evaluation constraints. An approach for choosing the most appropriate evaluation methods for IAS has been proposed based on the principles of multi-criteria decision making. The proposed approach presents a popular approach of multi-criteria decision making namely AHP. The approach is illustrated in the context of adaptive hypermedia systems in order to decide which of the four considered UCE methods we need to focus on. The used methods are: user-as-wizard, heuristic evaluation, focus group, and user test. The decision support approach is flexible in the sense that it can be applied in the case of different evaluation methods for adaptive systems. The novelty of this approach lies in the possibility to assist semi-automatically the choice of IAS evaluation methods to be adopted in each evaluation context. Since, the evaluator has to give the relative importance of each element in the hierarchy in order to use the analytic hierarchy process method.

Future directions of this research will deal with applying the evaluation in other fields of adaptive systems. Since we have tested the proposed approach only in the hypermedia context, it would be also interesting to generalize the approach to other fields of application. In addition, only four methods have been used to validate and illustrate the practical effectiveness of the proposed approach. It would be interesting to extend the proposed decision support approach to other evaluation methods. We plan also to enhance the proposed decision method. A decision support system to validate and test the proposed approach will be at the center of ongoing research.

## REFERENCES


[1] M. Alshammari, R. Anane, and R. Hendley, "Adaptivity in e-learning systems," The 8th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS 2014), pp. 79-86, Birmingham, United Kingdom, 2014.

[2] D. Goren-Bar, I. Graziola, C. Rocchi, F. Pianesi, O. Stock, and M. Zancanaro, "Designing and redesigning an affective interface for an adaptive museum guide," Affective Computing and Intelligent Interaction, vol. 3784, pp. 939-946, 2005.

[3] C. Kolski, "Human-computer interactions in transport," ISTE Ltd and John Wiley & Sons Inc., ISBN 978-1848212794, 2011.

[4] A. Goy, L. Ardissono, and G. Petrone, "Personalization in e-commerce applications," In Brusilovsky, P., Kobsa, A., Nejdl,W. (eds.), The Adaptive Web: Methods and Strategies of Web Personalization, Lecture Notes in Computer Science, vol. 4321, Springer-Verlag, Berlin Heidelberg New York, 2006.

[5] M. De Jong, and P. Schellens, "Reader-focused text evaluation: an overview of goals and methods," Journal of Business and Technical Communication, vol. 11, pp. 402-432, 1997.

[6] S. Weibelzahl, and G. Weber, "Advantages, opportunities and limits of empirical evaluations: evaluating adaptive systems," Künstliche Intelligenz, vol. 16, pp. 17-20, 2002.

[7] F. Del Missier, and F. Ricci, "Understanding recommender systems: experimental evaluation challenges," Proceedings of the Second Workshop on Empirical Evaluation of Adaptive Systems, held at the 9th International Conference, pp. 31-40, 2003.

[8] L. Van Velsen, T. Vander Geest, R. Klaasen, and M. Steehounder, "User-centered evaluation of adaptive and adaptable systems: a literature review," The Knowledge Engineering Review, vol. 23, pp. 261-281, 2008.

[9] C. Gena, "Methods and techniques for the evaluation of user-adaptive systems," The Knowledge Engineering Review, vol. 20, pp. 1-37, 2005.

[10] C. Gena, "A user-centered approach for adaptive systems evaluation," in Proc. of the Fifth Workshop on User-Centred Design and Evaluation of Adaptive Systems held in conjunction with the 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH2006), pp. 3-12, Dublin, Ireland, June 2006.

[11] C. Mulwa, S. Lawless, M. Sharp, and V. Wade, "The evaluation of adaptive and user-adaptive systems: a review," International Journal of Knowledge and Web Intelligence (IJKWI), vol. 2, pp. 20, 2011.

[12] C. Gena, and S. Weibelzahl, "Usability engineering for the adaptive web," In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.), The Adaptive Web: Methods and Strategies of Web Personalization, pp. 720-762, Springer, Berlin, 2007.

[13] International Journal of the Analytic Hierarchy Process, http://www.ijahp.org/index.php/IJAHP/index.

[14] A. Dhouib, A. Trablesi, C. Kolski, and M.Neji, "A framework for supporting the choice of usability evaluation methods for interactive adaptive systems," 15th International Conference on Intelligent Systems Design and Applications, 14-16 December, Marrakesh, Maroc, 2015.

[15] A. Paramythis, S. Weibelzahl, and J. Masthoff, "Layered evaluation of interactive adaptive systems: framework and formative methods," User Modeling and User- Adapted Interaction, vol. 20, pp. 383-453, 2010.

[16] P. Brusilovsky, C. Karagiannidis, and D.G. Sampson, "The benefits of layered evaluation of adaptive applications and services," In S. Weibelzahl, D. N. Chin, & G. Weber (Eds.), Proceedings of Empirical Evaluation of Adaptive Systems workshop associated with UM2001, Sonthofen, Germany, pp. 1-8, 2001.

[17] C. Mulwa., S. Lawless, M. Sharp, and V. Wade, "The evaluation of adaptive and user-adaptive systems: a review," International Journal of Knowledge and Web Intelligence (IJKWI), pp. 20, 2011.

[18] S. Lawless, C. Mulwa, and A. O'Connor, "A proposal for the evaluation of adaptive personalised information retrieval," The 2nd International Workshop on Contextual Information Access, Seeking and Retrieval Evaluation, Milton Keynes, UK, pp. 4, 2010.

[19] J. Masthoff, "The user as wizard: a method for early involvement in the design and evaluation of adaptive systems," In: 5thWorkshop on User-Centred Design and Evaluation of Adaptive Systems at AH'06, pp. 460-469, Dublin, Ireland, 2006.

[20] J. Nielsen, and R.Molich, "Heuristic evaluation of user interfaces," In: Proceedings of CHI '90, Seattle, Washington, pp. 249-256, 1990.

[21] G.D. Magoulas, S.Y. Chen, and K.A Papanikolaou, "Integrating layered and heuristic evaluation for adaptive learning environments," In: 2nd Workshop on Empirical Evaluation of Adaptive Systems at UM2003, Johnstown, PA, pp. 5-14, 2003.

[22] T. Mahatody, M. Sagar, and C. Kolski, "State of the art on the cognitive walkthrough method, its variants and evolutions," International Journal of Human-Computer Interaction, vol. 26, pp. 741-785, 2010.

[23] R. Krueger, and M. Casey, "Focus groups: a practical guide for applied research," 4th ed., Sage Publications, Los Angeles, 2009.

[24] J.S. Dumas, and B.A. Loring, "Moderating usability tests: principles and practices for interacting," Morgan Kaufmann, San Francisco, 2008.

[25] C. Karagiannidis, and D. Sampson, "Layered evaluation of adaptive applications and services," In: 1st International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Trento, Italy. LNCS, vol. 1892, pp. 343-346, Springer, Berlin, 2000.

[26] S. Weibelzahl, "Evaluation of adaptive systems," In: 8th International Conference on User Modeling. LNCS, vol. 2109, pp. 292-294. Springer, Berlin, 2001.

[27] N. Manouselis, C. Karagiannidis, and D.G. Sampson, "Layered evaluation for data discovery and recommendation systems: an initial set of principles," IEEE 14th International Conference on Advanced Learning Technologies (ICALT), pp. 518-519, 2014.

[28] A. Díaz, A. García, and P. Gervás, "User-centred versus system-centred evaluation of a personalization system," Information Processing & Management, vol. 44, pp. 1293-1307, 2008.

[29] J. Nielsen, "Heuristic evaluation. usability inspection methods," New York: John Wiley & Sons, pp. 25-64,1994.

[30] I. Denley, and J. Long, "A planning aid for human factors evaluation practice," Behaviour and Information Technology, vol. 16, pp. 203-219, 1997.

[31] A. Nendjo ella, C. Kolski, F.Wawak, C. Jacques, and P. Yim, "An approach of computer-aided choice of UI evaluation criteria and methods," In Vanderdonckt and A. Puerta (Eds.), Computer-Aided Design of User Interfaces II, Kluwer Academic Publishers, Dordrecht, pp. 319-328, 1999.

[32] X. Ferre, N. Bevan, and T.A. Escobar, "UCD method selection with usability planner," In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI '10). ACM, 2010.

[33] X. Ferré, and N. Bevan, "Usability planner: a tool to support the process of selecting usability methods," in Human-Computer Interaction–INTERACT 2011, ed: Springer, pp. 652-655, 2011.

[34] ISO TR 16982, "Usability methods supporting human centred design," 2002.

[35] S. Balbo, "Software tools for evaluating the usability of user interfaces," Proc. of the 6th International Conference on HCI, Pacifico Yokohama (Japan), pp. 337-342, 1995.

[36] C. Mulwa, and V. Wade, "A web-based evaluation framework for supporting novice and expert evaluators of adaptive e-learning systems," The International Conference on E-Technologies and Business on the Web (EBW201z3), Thailand, 2013.

[37] T. Lavie, and J. Meyer, "Benefits and costs of adaptive user interfaces," Int. J. Hum.-Comput. Stud. Vol. 68, pp. 508-524. 2010.

[38] T. Lavie, J. Meyer, K. Beugler, and J. Coughlin, "The evaluation of in-vehicle adaptive systems," User Modeling: Work on the EAS, pp. 9-18, 2005.

[39] T.L. Saaty, "Decision making with the analytic hierarchy process," Int. J. Services Sciences 1, pp. 83-98, 2008.

[40] T.L. Saaty, "The analytic hierarchy process," McGraw-Hill International, 1980.

[41] S.V. Omkarprasad, and S. Kumar, "Analytic hierarchy process: an overview of applications," European Journal of Operational Research, vol. 169, pp. 1-29. 2006.

[42] E.W.T. Ngai, “Selection of web sites for online advertising using AHP,” Information and Management, vol. 40, pp. 233-242, 2003.

[43] T. Reichel, and G. Rünger, “Prioritization of product requirements using the analytic hierarchy process,” In Proc. of the 14th International Conference on Enterprise Information Systems, ICEIS 2012, vol. 2, SciTePress, pp. 70-76, 2012.

[44] E.H. Forman, and S.I. Gass, “The analytic hierarchy process: an exposition,” Operations Research, vol. 49, pp. 469-486, 2001.