SUMMATIVE EVALUATION FOR THE USABILITY OF INTELLIGENT TUTORING SYSTEM (SPATH)

Shahliza Abd Halim, 1 Nor Bahiah Hj Ahmad, 2 Siti Mariyam Hj Shamsuddin, 3 Norreen Haron
Faculty of Science Computer and Information Systems
University of Technology Malaysia, 81310 Skudai
shahliza@utm.my, 1 bahiah@utm.my, 2 mariyam@utm.my, 3 norreen.haron@gmail.com

ABSTRACT
The various combination of multidisciplinary areas in Intelligent Tutoring System (ITS) contributes to various methods of evaluation for ITS and therefore the suitable method of evaluation must be chosen wisely. This paper discusses the evaluation of an adaptive hypermedia learning system called SPATH. SPATH provides adaptation of learning content by personalizing the learning material structure based on student learning style adapted from student's personality factor, Myers-Briggs Type Indicator, and by the student's knowledge acquisition level. In this research, summative evaluation has been used where we addressed the educational impact of the system on students and its practical acceptability in terms of usability. For the usability study, we used questionnaire to measure three key factors of usability such as, learnability, efficiency and satisfaction. Preliminary results of the usability study revealed that this system has a high percentage in learnability and satisfaction factor.

INTRODUCTION
Most of the early research in adaptive educational hypermedia was inspired by the area of Intelligent Tutoring Systems (ITS). A combination of ITS and learning materials organized as hypermedia was the beginning for the research on Adaptive Hypermedia System (Brusilovsky,2000). Due to the close relationships between ITS and AHS, the method of evaluation for ITS is considered as a suitable evaluation for AHS. Among the advantage of evaluation is that it provides an opportunity to learn from mistakes and is capable of improving the ITS life-span as well as their usability (Opperman et al., 1999).

Adaptive Hypermedia Learning System (AHLS) has been identified as an effective strategy for solving many learning problems involved in large hypermedia, such as cognitive overload and user disorientation (Brusilovsky,2003). The idea of adaptive hypermedia is to adapt the course content accessed by a particular user to his/her characteristics. Most adaptive educational hypermedia system researches focuses on adapting to user features like user’s goals/tasks, knowledge, background, hyperspace experience, preference and interests (Brusilovsky,2000). However, a web-based educational system must also include information about student learning styles to adapt optimally instructional materials to the student (Brusilovsky,2003; Alomyam, 2004; Ford & Chen, 2000; Magoules et al., 2003).

This research has developed an adaptive hypermedia learning system called SPATH for learning Data Structures at Faculty of Science Computer and Information System in Universiti Teknologi Malaysia (UTM). We have explored the computational intelligence technique that effectively able to adapt the learning material based on the student learning style. In this case, we combined the student’s personality factor, Myers-Briggs Type Indicator (MBTI) and Fuzzy Logic technique to produce a dynamic course adaptation which will present the appropriate structure of the learning material to the student (Haron & Salim, 2005). As for the classification of students based on their level of knowledge acquisition we used supervised Kohonen network for classification purpose (Yusob et al., 2004).

The system has been tested and implemented on Computer Science students. This paper discusses the evaluation of SPATH in order to measure the efficiency, learnability and the satisfaction of the system usability. We propose a reliable method for the usability study of the system. We further describe the method and process done while performing the study and also provide the illustrations of the preliminary results. Finally, the results achieved are discussed followed by the conclusion and the future work of the research.
EVALUATION METHODS FOR ITS

Evaluation for ITS is not an easy task as ITS is the combination of multidisciplinary areas including expert system, Computer Based Instruction, education, psychology and also Human Computer Interaction (Opperman et al., 1999; Mark & Greer, 1993; Silius & Tervakari, 2003). Both (Opperman et al., 1999) and (Mark & Greer, 1993) highlighted that only few agreed upon standards within the ITS community to guide investigators who wish to evaluate ITS given the diversity of the evaluation methods.

There are two types of evaluation for ITS, formative and summative evaluation. Formative evaluation mainly occurs during design and early development of a project. It frequently addresses the question of relationship between the architecture of ITS and its behavior. On the other hand, summative evaluation is concerned with the evaluation of a completed system and making of formal claims about the system. It answers the question regarding the educational impact of an ITS on students. However, these types of classification are still too broad where a lot of methods can fall in either one of these classes.

(Opperman et al., 1999) stated that various methods have been further classified, so that the method could be differentiated from others on a scale between external evaluation (considering the whole system) and internal evaluation (testing a component of the system). In addition a method could be classified along a dimension consisting of exploratory research versus experimental research. Though the proposed classification provide a simple yet robust way to select evaluation methods, the classification needs future work to add other dimension of formative and summative evaluation to the classification chart.

Another research has been done on evaluation solely on the usefulness of web based learning environment (Silius & Tervakari, 2003). The research takes into consideration the multidisciplinary evaluation framework for evaluating a web based learning system. The framework combines two main issues, usability and utility issues where utility is broken down into two parts, pedagogical usability and added value. (Silius & Tervakari, 2003) showed the importance of usability remains the same regardless of how much web based learning is used in the teaching as a whole. As for pedagogical usability, its importance gradually increases as the focus of teaching shifted from traditional teaching into more on web based teaching.

In this research, we use summative evaluation where we address the educational impact of this system on students. We also use the multidisciplinary evaluation framework where we focus on the usability study of this system as (Silius & Tervakari, 2003) clearly shows that the importance of usability is consistent regardless of the focus of teaching either traditional or web based teaching.

USABILITY STUDY FOR ITS

Usability is a quantitative and qualitative measurement of the design of user interface grouped into five key factors: learnability, efficiency, memorability, errors and satisfaction (Nielsen, 1993). The setting of the usability study can vary. A usability laboratory can be used for a controlled experiment. A workplace test can be used to test the user in their normal work environment such as at their desk during a routine work day. There is also web-based usability testing referred to as remote usability testing where the user and experimenter are not physically located in close proximity of each other. The different categories of usability tests consist of performance measurement, thinking aloud protocol, coaching method, retrospective testing, constructive interaction, and questionnaires (Nielsen, 1993; Lazar, 2001).

Performance measurement takes place when quantitative measures are taken during the testing such as the number of tasks completed successfully by the user, length of time to complete the test tasks, number of errors, and time spent recovering from errors. Thinking aloud protocol exists when users vocalize their thoughts and therefore share their positive and negative interpretations of different website features. The coaching method enables the users to ask questions and receive answers which give researchers insight into the type of help documentation or better technology design needed. Questionnaires are also a form of testing as it provides an opportunity to gather more usability feedback from a user after a testing session (Carsen & Patterson, 2005).
In this study, questionnaires were used as a tool to gather feedback from the participants and the testing phase was conducted in a controlled environment setting. Three out of the five key factors, learnability, efficiency and satisfaction of the students were the focus of the study. Learnability of the students is measured based on the ease of use of the students when working towards completing the task specified for them. Efficiency looks at how productive the students once having learned the software. Satisfaction which is the third factor is to measure the students level of pleasure while using the system.

ADAPTIVE HYPERMEDIA SYSTEM (SPATH)
There are three main components in SPATH, adaptive engine, domain model and user profile model. Adaptive engine interface enable the user to interact with the system. User will be presented with the learning material and navigation path adapted based on the learning style and information stored in the user profile model. Adaptive engine generate features of adaptation and a tree structure to present the learning material structure.

User profile model stores the information about learners. The information is extracted from both explicit and implicit user profile. The explicit information is the information that the learner gave while filling in the registration form and the questionnaire when they used the system for the first time. The implicit information is the information that the system collects while the user interact with the system. Domain model comprise two subcomponents, dynamic curriculum sequencing and adaptive navigation path. Using Fuzzy logic, the student learning style and the sequence of learning material structure based on their learning style has been determined. In this research, we have identified four types of student learning style, Introvert-Sensor, Extrovert-Sensor, Introvert-Intuition and Extrovert-Intuition based on personality factor MBTI. Table 1 shows the sequence of the learning structure based on the learning style identified.

Table 1: Example of 4 situations of the learning material structure based on the learning style

<table>
<thead>
<tr>
<th>Learning Style</th>
<th>Structure of learning material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introvert-Sensor</td>
<td>Example</td>
</tr>
<tr>
<td></td>
<td>Theory</td>
</tr>
<tr>
<td></td>
<td>Activities</td>
</tr>
<tr>
<td></td>
<td>Exercise</td>
</tr>
<tr>
<td>Extrovert-Sensor</td>
<td>Activities</td>
</tr>
<tr>
<td></td>
<td>Exercise</td>
</tr>
<tr>
<td></td>
<td>Example</td>
</tr>
<tr>
<td></td>
<td>Theory</td>
</tr>
<tr>
<td>Introvert-Intuitive</td>
<td>Theory</td>
</tr>
<tr>
<td></td>
<td>Example</td>
</tr>
<tr>
<td></td>
<td>Exercise</td>
</tr>
<tr>
<td></td>
<td>Activities</td>
</tr>
<tr>
<td>Extrovert-Intuitive</td>
<td>Exercise</td>
</tr>
<tr>
<td></td>
<td>Activities</td>
</tr>
<tr>
<td></td>
<td>Theory</td>
</tr>
<tr>
<td></td>
<td>Example</td>
</tr>
</tbody>
</table>

The knowledge level of the student will be updated once the student logs off the system. In this situation Kohonen program will be executed and be given the latest information on the student activities while learning on-line, the score for doing assessment and the interaction history as input vector. Adaptive navigation path component further will assist the student while navigating through the system.

METHODS AND PROCESSES OF EVALUATION
Summative evaluation conducted in this study was tested on 44 Computer Science students in Faculty of Science Computer and Information System, Universiti Teknologi Malaysia. The learning material for the participants to study, the task and the processes of the usability study are discussed in the following subsections.
Learning Material
The learning material used in this study is based on the topic Sorting Techniques. The learning material is presented in Malay language. The topics involved are Bubble Sort, Selection Sort, Insertion Sort, Merge Sort and Quick Sort. In order to handle different types of learning styles, the learning material for this topic is structured into theory, exercise, example and activity (Ahmad et al., 2005). The theory part consists of explanation and pictures describing the Sorting technique. The exercises tested the students knowledge on the Sorting concepts. After the students finished the exercises, they were given their score and hints for the answer of each questions. The example shows the simulation of different types of the Sorting process. Students can view the simulated Sorting activity based on the execution of the algorithm code line by line. In the activity session, students can generate data to be sorted randomly and view the simulation of the Sorting technique.

The structure of the learning material is adapted based on the student learning style. Different category of learner will get different tree structure. Figure 1 shows the sequence of the learning material based on the Extrovert-Intuitive students. For each learning material, links are provided for each Sorting topic.

![Figure 1: The tree structure for learning sorting concept in support of Extrovert-Intuitive students](image)

Task
The task given to the participants was to study Sorting Techniques provided by the system. As part of the task, the following steps would be the processes that each participant would go through during the usability study:

- Listen to brief explanations about the system.
- Read the task instructions.
- Read short and simple user manual if they need further guidance about the system.
- Fill in the on-line questionnaire based on MBTI in order to determine the student’s personality factor.
- Study the learning materials according to the task given.
- Complete the assessment on-line after completing the study session.
− Fill in questionnaires on the learning satisfaction.
− Fill in questionnaire on the usability of the system.

Processes
The usability study is carried out in a laboratory in a controlled environment. The students were required to complete their learning within two hours or until they have finished completely their learning. Two facilitators were in charge in case the students need explanations or help. A desktop with Internet connection was provided for each participant to complete the study. The process of the study is as follows:

1. For the study, the students were divided into two groups. The first group is required to learn based on the sequence of the topic provided in the system. The second group are allowed to learn freely without following the sequence.
2. The students were given a short and simple user manual to guide them while using the system
3. Before the student can use the system, he/she has to register and fill in an online questionnaire. The information collected during the registration were used to gain some personal information such as the user’s name, student’s id, cumulative point average and programming knowledge status in order to initialize the student model.
4. On-line questionnaire based on MBTI is used to determine the student’s personality factor value either as Introvert-Sensor, Extrovert-Sensor, Introvert-Intuition or Extrovert-Intuition. Figure 2 shows the sample of the on-line questionnaire. Data collected on the questionnaire were used to determine the learning style of the students using Fuzzy Logic.
5. The students started learning and they can chose the learning material structure given to them either sequentially or freely.
6. After the students finished learning, they have to do the assessment on-line. The assessment is provided in order to measure the knowledge level of the students on the topic.
7. Upon the completion of learning and assessment, participants were required to fill up two questionnaires on learning satisfaction and usability of the system. The second questionnaire is adapted with slight changes from (Granic et al., 2004) with close ended and open ended questions concerning on user subjective satisfaction.
8. Analysis of the data collected during the study consists of data from on-line learning style, students learning satisfaction and usability of the system questionnaire. The results are discussed in the following sections.

RESULTS
The results elaborated in this section are mainly from the analysis of data collected from the two questionnaires based on the three key factors in usability study the learnability factors, efficiency factors and satisfaction factors. Apart from the three key factors, the students learning styles are determined based on the on-line questionnaire given in the system.

Learning Style Categorization
Based on the MBTI personality factor, the students are categorized into four groups of learning styles. Figure 3 shows the number of students for each learning styles where 21 students are in Extrovert-Sensor group, 8 students are grouped in Introvert-Sensor, 4 students are in the Extrovert-Intuitive group and 11 students are in the Introvert-Intuitive group.
Figure 2: Interface of the on-line questionnaire for personality factor based on MBTI

Figure 3: Learning style of the participants

Learnability Factor

The learnability factor is measured quantitatively based on students’ ease in completing the task of learning using the system. The learnability factor is measured based on these three factors:

i) The ease in learning based on personalization of the students learning style using the MBTI personality factor.
ii) The ease in learning based on the learning structure given
iii) The ease in learning based on the navigational path
Based on the questionnaire, 42 students were either strongly agreed or agreed on the effectiveness in learning based on MBTI learning style while only 2 students disagree in the learning style approach. The high learnability factor of this system is further proved by the number of students prefer the to learn based on the learning structure sequence given. The last element which contributes to the ease of using the system is the adaptive navigational paths provided by the system. 6 of the respondents strongly agree and another 30 respondents agree that the navigational paths provide ease of using the system. Only 8 of the respondents disagreed that annotated navigational paths provide ease of using this system. Table 2 shows the three features used to measure learnability factors of the system.

Table 2: Three features to measure learnability factor in using the system

<table>
<thead>
<tr>
<th>Feature</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectiveness of learning</td>
<td>4</td>
<td>38</td>
<td>2</td>
</tr>
<tr>
<td>Preferable learning structure sequence</td>
<td>7</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td>Ease of using adaptive navigational path</td>
<td>6</td>
<td>30</td>
<td>8</td>
</tr>
</tbody>
</table>

Efficiency Factor

Efficiency factor of this system is measured by the productiveness of the students in terms of the score from the assessment they have completed in the system. The assessment has been done on-line by the students after they have completed their study using the system. Figure 8 shows that the number of the students with score between 80 to 100 is quite high, 28 students while 11 and 2 students has either scored between 40 and 79 and between 0 and 39 respectively. The number of students with their respective assessment score is as shown in Table 3.

Table 3: Feature to measure efficiency factor in using the system

<table>
<thead>
<tr>
<th>Assessment Score</th>
<th>No. of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-39 marks</td>
<td>11</td>
</tr>
<tr>
<td>40-79 marks</td>
<td>5</td>
</tr>
<tr>
<td>80-100 marks</td>
<td>28</td>
</tr>
</tbody>
</table>

Satisfaction Factor

The final key attributes measured is the satisfaction factor of the user while using the system. The students satisfaction were determined based on attributes such as the familiarity and understandability of the terms used in the system, the easiness to understand the learning contents, the suitability of colors used in the system, the usefulness of the interactivity in the system and the overall satisfaction the user experienced when they use the system. Table 4 shows the number of students and their learning satisfaction in using the system.

Table 4: Feature to measure satisfaction factor in using the system

<table>
<thead>
<tr>
<th>Participants learning satisfaction</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants learning satisfaction</td>
<td>4</td>
<td>38</td>
<td>2</td>
</tr>
</tbody>
</table>

The usability questionnaire have two open ended questions for the participants to make suggestions to enhance the application functionality and to express participants feeling while using the system. Some excerpts of the feedback from the participants are:

1. The system is interesting because I can clearly see how the sort function working.
2. Interesting especially using the flash application but the yellow color for the background is giving me pressure.
3. Good, this is a new technique to improve the student in learning Data Structure.
4. Interesting because it is easy to understand the subject by using this system.
5. A little confusing at first but at the end I am amazed. It is easier to learn according to the learning style.

DISCUSSION
The results showed that the system has achieved high percentage of learnability factor. Most of the students (61.4%) have followed the sequence given by the system, although they have been given the task of learning freely regardless of the sequence of learning material. This result showed that sequence of learning material structure suits their learning style and thus helps them in their learning.

The second key attributes in the usability study is the efficiency factor. Efficiency is measured from the productiveness of the participants or in this study the assessment score of the students after learning using this system. Although only 11.4% of the students scored between 80 and 100, it is an acceptable figure considering students were expected to understand both the theory and algorithms of each sorting technique within two hours and also due to the detailed coverage of the assessment questions which really test the students in terms of their understanding. The performance of the students can be improved if the web based usability study is adopted instead of controlled experiment. With web based usability study the participants can learn freely without time constraint (Cartens & Patterson, 2005).

The last key attributes, the satisfaction factor is measured based on the level of pleasure the participants experience while using this system. In this study, the results has showed high satisfaction factor whereby majority of the participants agree on the pleasurable attributes the system offer such as the understandable terms, the learning material structure and the interactivity in the system. Furthermore, the subjective feedback given by the participant demonstrate that learning material structure as in example and activity structures really give them more understanding of the subject thus giving them satisfaction in using the system.

From this research point of view, the summative evaluation which addresses the educational impact of this system to student can be answered by the learnability factor which shows the suitability of the learning style in personalizing the learning material for the students. Another factor contributing to the educational impact of the system to the user is the learning material structure itself such as example and activity structures really provide the students satisfaction in learning using this system.

Even though the results of this study shows positive outcome in the usage of this system but there are also several weaknesses we encounter in conducting the usability study. The drawbacks are:

i) There are no comparison between students who have already learned data structure and students who are still learning this subject. If this type of comparison can be done, more interesting result can be analyzed.

ii) In this research there is also no comparison done between the usages of this system with other similar system. In future, we decide to do a comparison study on the usage of this system with INSPIRE system (Grigoriadou et al., 2001).

iii) There are comments from the students particularly on the choice of color of the system interface, which some of them thought as glaring and also mismatched. Other students were annoyed with the continuous system's automatic refresh where they have to reselect again the tree structure in order for all the learning material to be displayed again.

CONCLUSION AND FUTURE WORK
This paper has discussed the evaluation of an adaptive hypermedia system by implementing the usability study. Three key factors in the usability study has been measured. The method and processes of the usability study has also been elaborated. From the preliminary results and discussion, it can be concluded that SPATH has high percentage for learnability and satisfaction factor. These factors contributes to the answering of the question asked in the summative evaluation: "What is the educational impact of this system to user?". For future
work of the study, we intend to implement web usage mining in order to measure the efficiency of the learning approach given to the student based on their learning style.

REFERENCES
Granic, A., Glavinic, V. & Stankov, S. (2004). Usability evaluation methodology for web-based educational systems. 8th ERCIM Workshop User Interfaces For All. 28-29 June
Haron, N. & Salim, N. (2005). Integrating personality factor (MBTI) and Fuzzy logic techniques in individualizing learning material of adaptive hypermedia learning system”. Konvensyen Tenologi Pendidikan Ke 18. 16-19 September, Marang, Terengganu.