

User Testing of a System Prototype for Adaptive English Learning using Perturbation Learner Model

Nguyen Viet Dung, Suziah Sulaiman, Mohd Fadzil Hassan

Computer and Information Sciences Department,
Universiti Teknologi PETRONAS
31750 Tronoh, Perak, Malaysia

Keywords:

Adaptive learning
English learning
Perturbation learner model
Usability testing
User-centered testing

ABSTRACT

This paper presents a study to evaluate a system prototype that supports adaptive English learning. The prototype was previously developed based on a conceptual framework using perturbation learner model but no formal evaluation was conducted. The evaluation study consists of user testing on the system's functionality and usability aspects. The findings suggest conformity of the learning content from the teacher's perspective. Similarly, the students viewed the system as easy to understand, easy to navigate and easy to recognize.

*Copyright © 2013 Information Systems International Conference.
All rights reserved.*

Corresponding Author:

Nguyen Viet Dung
Computer and Information Sciences Department
Universiti Teknologi PETRONAS
Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia
Email: vietdung8688@gmail.com

1. BACKGROUND

Earlier research has proved that adaptive learning is an advanced educational method applicable to a variety of educational and training models to enhance the effectiveness and efficiency in teaching and learning. The idea behind adaptive learning is agreed to be an absolute utilization of explicit learner models – created and constantly updated through a process called user modeling – capable of managing users' information and automatically adapting the presentation of learning content and material according to the users' responses and progress when interacting with the learning system

This paper presents an evaluation study of a system prototype to aid the learning of English as a second language in Vietnam. The prototype was previously developed by the authors based on a conceptual framework using perturbation learner model to support adaptive learning. Perturbation learner model was chosen as it supports feature-based user modeling and manages the learners' errors or misconceptions [1, 2, 3, 4]. It provides a fault model for the purpose of providing the learners with constant and appropriate explanations, hints, guidance, and annotations in order for them to gradually and firmly erase their mal-knowledge throughout their learning process [5]. Even though the conceptual framework has been integrated into a system prototype, no evaluation has taken place to determine how well the perturbation model adopted supports adaptive learning. This issue becomes the central focus of this paper.

2. OVERVIEW OF TESTING AND EVALUATION APPROACHES

Testing and evaluation are very important to any systems before being delivered to the end-users [6]. There are several approaches available to achieve the most satisfactory and reliable results namely empirical evaluation, layered evaluation, heuristic evaluation, and, most typically, user-centered evaluation [7]. According to [8], [9], and [10], empirical evaluation is necessary to gauge the usability of a system by applying techniques in real-world scenarios, and observation in controlled experiments. Empirical evaluation is expected to uncover certain types of errors in the recommended systems which would remain otherwise undiscovered. The key to good empirical evaluation is approved to be the proper design and execution of the experiments in a way that the particular factors to be tested can be easily separated from other confounding factors [11]. Empirical studies are good at identifying design errors and false assumptions but time-

consuming and require a great number of human resources; thus, usually employed in large-scale research projects only [9].

Layered evaluation, on the other hand, refers to the suggestions that the systems' adaptation mechanism is to be decomposed into separate layers in order to be evaluated effectively [12]. Several adaptation layers are described as "a collection of input data, interpretation of the collected data, modeling of the current state of the world, deciding upon adaptation, and applying adaptation" [13]. Layered evaluation is exhaustive but relatively complex and requires expert knowledge to be conducted properly [10].

Heuristic evaluation refers to a usability evaluation method in which an analyst looks for usability problems by checking the user interface against a set of predefined heuristics or principles [15]. The use of heuristics ensures that the entire system can be evaluated in-depth and specific problems can be discovered at an early stage before a running prototype is released [11, 16]. By this way, the detection and diagnosis of potential usability problems can be improved – without indications to how they are to be fixed, though. Heuristic evaluation requires experienced evaluators.

User-centered evaluation, finally, refers to the assessment of a system's usability as dedicated to its intended end-users i.e. to what extent a system facilitates its users in fulfilling their primary tasks [14]. The system is required to take into account as much useful information about the users as possible (learner's features and statistics) in order to yield relevant and valuable testing and evaluation results [11, 14].

In relation to the aim of the study where the focus is on determining the usefulness of the conceptual framework in adaptive learning, empirical and layered evaluations turn out to be not relevant; instead, user-centered approach holds a number of benefits in terms of resource savings, guarantee of system's functionality completeness, minimized repair efforts, and improved user satisfaction [11]. In this regard, user-centered evaluation is more suitable as the main evaluation approach for this particular research, supported with the use of several heuristics in an informal manner chiefly for functionality checking.

3. THE EVALUATION STUDY

The objectives of this study are to evaluate the (i) system functionality and, (ii) usability of the system prototype. The first objective involves testing the accessibility and learning materials of the prototype while the second is about testing the usability of the prototype for its effectiveness, learnability, flexibility (adaptivity), efficiency, errors, memorability, attitude and satisfaction.

3.1. Participants

20 sixth-form students from a secondary school in Thai Binh province – Vietnam were selected and supervised by Ms. Nguyen Dieu Linh, a young English teacher (two-year experience) graduating from Hanoi University and majoring in English language. The testing process took place in the late 2nd semester – of a two-semester academic year – and lasted for three weeks, ending just before the students' final examination.

3.2. Test Set-up

The system was set up and monitored in a computer-based classroom with local network. Ms. Nguyen assisted in this arrangement and volunteered in the first trial to make sure the databases were working correctly, and the website could be accessed from any computers in the same classroom network. The students were briefed on how the system works. They were authorized to enter the classroom for unsupervised learning whenever they felt at ease during working hours. This is to address the limitation of the local network as the system was inaccessible outside the classroom. For the variance in the students' timetables – the students were randomly selected from a few different classes for sampling purpose.

3.3. Method

During the three weeks of testing, the students were made free but encouraged to optimize the use of the learning system as much as possible. They were allowed to skip their English lesson in class to focus on their learning with the system and two compulsory tutorials involving all the 20 students were scheduled weekly for two hours per slot for the purpose of intensive study. Ms. Nguyen was kept available all the time for active and opportune support and played the leading part in keeping track of the students' activities and performances. The results were finalized in the form of recorded databases, together with which, a specialized evaluation of the quality and appropriateness of the learning content and material used for the system prototype from the teacher's perspectives was provided. Finally, a survey on the accessibility and usability of the system from the students' perspectives was conducted.

4. RESULTS AND DISCUSSIONS

A formal user testing was conducted to elicit the students' responses and observe their performances as well as getting the teacher's feedbacks, and opinions. The results of the user-system interaction activities were collected in the form of recorded databases – since they reliably reflect how the students actually got themselves involved in the adaptive learning process, followed by a survey on their subjective feelings and opinions as a method of verifying the reliability of these implicitly recorded testing results.

4.1. Student's Learning Activities

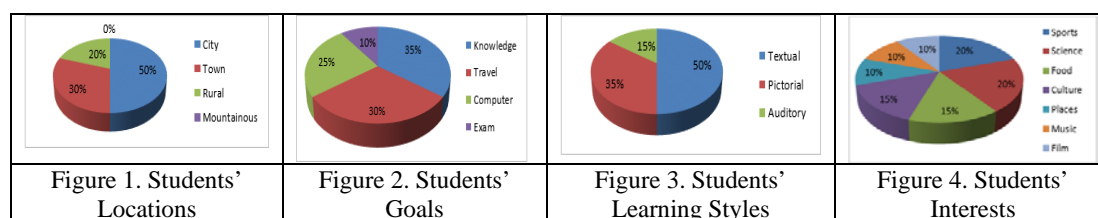
Table 1 shows all the 20 participating students' learning preferences – a sum-up of tables Background and Feature in Learner Info Database:

id	yob	gender	pob	location	level	internet	goal	interest	style	time
8	2000	Female	Hung Yen	city	6	limited	exam	food	pictorial	3
9	2000	Male	Thai Binh	city	6	full	computer	sports	textual	4
10	2000	Female	Nam Dinh	town	6	full	computer	music	auditory	3
11	2000	Female	Thai Binh	rural	6	full	travel	culture	pictorial	4
12	2000	Male	Thai Binh	rural	6	full	exam	science	textual	3
13	2000	Female	Thai Binh	rural	6	limited	knowledge	culture	textual	3
14	2000	Male	Thai Binh	city	6	limited	knowledge	science	textual	3
15	2000	Male	Thai Binh	city	6	full	travel	places	pictorial	2
16	2000	Female	Thai Binh	town	6	full	knowledge	sports	textual	3
17	2001	Female	Thai Binh	city	6	full	knowledge	science	textual	2
18	2000	Female	Thai Binh	town	6	limited	computer	culture	auditory	3
19	2000	Male	Thai Binh	city	6	full	knowledge	film	textual	3
20	2000	Male	Thai Binh	city	6	full	travel	music	pictorial	2
21	2000	Male	Thai Binh	city	6	limited	travel	food	pictorial	3
22	2000	Female	Thai Binh	city	6	full	knowledge	places	pictorial	4
23	2000	Male	Thai Binh	rural	6	limited	knowledge	science	textual	3
24	2000	Female	Thai Binh	town	6	full	computer	sports	pictorial	4
25	1999	Male	Hai Phong	town	6	limited	travel	sports	textual	4
26	2000	Female	Thai Binh	city	6	limited	computer	food	auditory	2
27	2000	Male	Thai Binh	town	6	full	travel	film	textual	3

id	test	posPerf	negPerf	level	active	download
8	8	6	2	2	14	6
9	10	8	2	3	19	9
10	9	6	3	1	17	5
11	11	7	4	1	18	7
12	11	9	2	3	19	7
13	7	5	2	1	13	2
14	6	5	1	2	13	8
15	8	4	4	0	16	4
16	16	13	3	5	23	13
17	11	8	3	2	16	4
18	5	5	0	2	11	7
19	10	10	0	5	14	8
20	11	6	5	0	15	4
21	6	3	3	0	10	4
22	14	9	5	2	21	6
23	6	2	4	0	13	4
24	8	8	0	4	14	7
25	5	4	1	1	11	3
26	7	6	1	2	15	6
27	9	4	5	0	17	4

All students were in Grade 6 at the time of the test, hence their year of birth is supposed to be 2000 (12 years old) – the two cases of 2001 and 1999 do not actually affect as sometimes students go to school one year earlier or later or the date when they were born is close to the year of 2000 (late 1999 or early 2001). There were 10 males and 10 females, 17 of whom are from Thai Binh province where the secondary school is located while the other 3 are from the neighbor provinces or city – Hung Yen, Nam Dinh, Hai Phong.

From Figures 1-4, there is a big diversity in the students' learning preferences. In particular, half of them live in the city while respectively 30% and 20% reside in small towns and the countryside. Since the testing environment is confined to the Red River Delta – Northeastern Vietnam, no students are reported to be from mountainous regions. With regard to goal, interests, and learning style, all the provided options hold a certain proportion of responses. For goal, most of the students showed their concern for gaining knowledge, travelling, and using a computer (only 10% insisted on passing an exam). Relatively equal percentages are also recorded for each of the interests' options (ranging from 10-20%), with sports and science earning the most interest. As for learning style, textual holds the most part (50%), followed by pictorial (35%) and auditory (15%). It can be said that all the provided options for goal, interests, and learning style are relevant to the students' real concerns. Table 2 summarizes their activities and performances on that basis.



From Table 2 the students experienced different accesses to the adaptive e-learning system – as introduced in the previous section: most of them got access to the website between 10 and 19 times inclusive, with only two seeming to have more free time or interest with over 20 indications of activeness. However, the frequency of access is not directly proportional to the number of tests performed or library files downloaded, implying that the students might have spent more time on the learning modules, not felt interested in the provided reference material, or logged out of the website several times without doing the

post-tests. Based on the statistics, a total of 178 pre-tests and post-tests were completely attempted, 71.91% of which had positive results (scoring over 70% of total marks). Students ID 16 and 22 were those ones who performed the most tests, 16 (13 positive) and 14 (9 positive) respectively; whereas, students ID 18 and 25 each performed only 5 tests, respectively achieving 5 and 4 positive results.

The seemingly most effective and efficient learner was noted to be student ID 19, performing 10 tests after only 14 times of access, all of which were marked positively. This learner, together with students ID 18 and 24, were reported without negative performances. Students ID 20, 22, and 27, on the contrary, scored the most negative results, the last of whom, along with student ID 23, even had a record of more negative than positive. Besides students ID 15 and 21 who had a balance of positive and negative performances, the other 16 learners were found to have completed the testing with a positive difference.

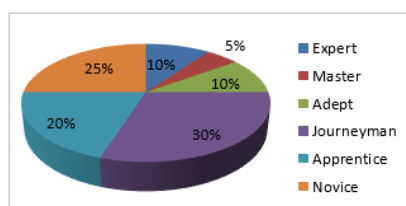


Figure 5. Students' Levels of Performance

Figure 5 compares the students' final levels of performance, denoted by corresponding status. Up to 75% of the learners ended the learning process with the lower levels of performance, one-third of whom – a quarter of all, got level 0 (Novice); the other two-thirds – half of all, respectively achieved level 1 (20%) and 2 (30%). Of the other quarter, 10% are Adept (level 3), 5% are Master (level 4), and 10% (2 students) were satisfactorily able to accomplish the highest level (Expert).

The results indicate that despite a rather high percentage of positive test performances, majority of students were unable to reach a higher performance level due to the mistakes they made during the learning process. This is still acceptable considering the small number of tests performed (an average of almost 9 per student – while it is required to score at least 6 positive results without a negative one to only reach level 3). The students may have had more chances to improve their performance if they have better access to the e-learning system or the testing process had lasted longer. In total, there were 118 registered library file downloads, 44 of which are from those students with higher levels of performance (25%), laying a possibility that the more library files the learners referred to, the better performances they would make.

4.2. Student's Survey Results

Two set of questionnaires were prepared to seek for the learners' feedbacks about the accessibility and usability of the prototype, respectively. As mentioned earlier, the website was only accessible through the local network in the computer-based classroom; thus, the purpose of the first questionnaire was not to assess the website's accessibility all over the Internet but rather to measure its potential to be widely accessed if implemented online – as compared to the limitations of classroom-based implementation. The type of questions asked from the users range from queries pertaining to availability of Internet access from home to students' willingness to use the system if available online. Three possible answers were expected from the students i.e. "yes", "no" or "not sure".

The findings reveal that most students have Internet access from their home (14 over 20), 12 of whom frequently go online (Table 1). However, not many of them are aware on the existence of e-learning systems. Among the nine cases of awareness, only 4 had experience in learning English online. Up to 11 learners claimed that it is not a good way to learn a foreign language online (2 were unsure), still as high as 15 of them agreed upon the prototype's usefulness for learning English online, revealing that their opinions were somehow affected through their interaction with the learning system during the testing period. In terms of accessibility, 13 learners express their difficulties in accessing the website except for the two compulsory tutorials they were assisted to attend weekly, as compared to a number of 16 showing their interest in becoming a frequent learner in case the system is made available over the Internet. This is to reassure that the utilization of classroom-based systems is very limited [17], and the system stands a high potential of getting widely used once it is put online. Table 3 presents the results for evaluating the prototype's usability.

The second set of questionnaire was prepared to measure the usability of the system from the students' own perspectives i.e. getting an idea of what they thought about the system prototype after some time interacting with it. The students were expected to respond to all questions with a rating point between 1 and 5 inclusive, respectively signifying the lowest and highest degree of agreement.

As highlighted in Table 3, three-fourths of the questions acquired more than 75% of average to high agreements (rated 3-5 points). 19 students agreed that the system is easy to understand, easy to navigate, and easy to recognize, 17 of whom typically offered high (16) and very high (1) agreements. Only 17 learners (10 with average point) approved of the appropriateness of the learning content and material delivered to them, still with two disagreeing and the other one even expressing strong disapproval. This is to indicate that even though the adaptation mechanism did work relatively properly, there still exist some problems that must be carefully reviewed in order to regain or reinforce learners' satisfaction.

Table 3. Evaluation on the Usability of the Prototype

Survey Question	Rating Point				
	1	2	3	4	5
Is the system easy to understand, navigate, and recognize?	0	1	2	16	1
Do the learning content and material delivered reflect your learning preferences?	1	2	10	7	0
Do the tests provided conform to what you learn in the learning modules?	0	0	8	10	2
Does the system help you to recognize, diagnose, and recover from any mistakes?	0	3	11	6	0
Is there a diversity and plenty of learning content and material provided?	4	11	5	0	0
Do the quizzes and reference material help your learning performances?	0	1	9	8	2
Does the system help you to improve your basic English skills, your in-class performances, and somehow achieve the goal you specified?	2	4	12	2	0
Overall, do you think the learning system is useful?	0	4	13	3	0

Surprisingly, the learners showed their high appreciation to the conformity of the prepared tests and the way their mistakes were highlighted and explained. Even though the learning content and material may not have been adequate to all, the accordingly arranged tests were quite reasonable and the students were able to recognize and recover from the mistakes they made with the support of the system's error handling utility. Besides, the provision of ordinary quizzes and library files were also highly appreciated although the majority of learners agreed that more diversifications of such facilities may improve their learning.

The only thing that most of the students found unsatisfactory was the shortage and lack of diversity of the learning content and material – as high as 15 graded this 1 or 2, which might also be the leading reason why several of them claimed that they did not learn much from the testing process (given 1 and 2 points in turn by 2 and 4 learners) i.e. their basic English skills and in-class performances have not been improved and their specified goal of using the system has not been fulfilled. In brief, this limitation has already been predicted during the system prototype development process owing to the shortage of time and human resources while a huge collection of learning content and material is said to be a must for the success of a general learning system, let alone adaptive counterparts which are more functionality complex and resource consuming. Despite this drawback, the system prototype was overall evaluated as being acceptable, promising to become a useful tool for English learning.

4.3. Teacher's Evaluation

The teacher's overall evaluation in terms of the quantity, quality, and appropriateness of the learning content and material prepared for the system prototype was sought after. All the elements of the lessons provided – vocabulary, grammar, reading, and communication – were recommended by Ms. Nguyen (the teacher in charge) before the prototype was created. These are the fundamental elements of an English lesson conveyed to the students in Vietnamese secondary schools. Similarly, the review exercises, quizzes, and tests were suggested to be in the form of multiple-choice questions to reduce the learners' potential input errors as well as to agree with the general form of English testing currently being maintained in Vietnam. The teacher was given a set of questions in which she has to give a rating point from 1 (very poor) until 5 (very good).

The findings reveal that the teacher highly agreed upon the conformity and appropriateness of the learning content and material prepared for the students (4 points), mainly because the material was selectively refined from the official resources for Vietnamese grade-6 students (via textbooks, reference books, or online). Despite this conformity, the learning material was commented as neither not so rich in content nor quantity, with only 2 points, due to the difficulties in the content preparation process in terms of time and human resources. The types of exercises prepared for revision or tests were said as not diverse as there should be more exercises to test the students' understanding particularly on vocabulary and grammar.

Nevertheless, the teacher still approved on the way the learning content and material were delivered to the students – giving 3 points – and so as to the mechanism of error handling. This signifies an important point that despite the limited quantity and low diversity of the learning content and material prepared, the way they got presented to the learners as well as the way potential errors got dealt with was still acceptable;

in other words, the prototype's adaptation mechanism can be said to have functioned properly, to some extent, and has been successful in managing both the learner and the fault models (perturbation model).

Apart from the learning content and material, the teacher expressed high approval to the provision of simple quizzes and reference material with a view to diversifying the students' learning approaches. According to her, reference material plays a major part in encouraging the students' self-study which helps them to obtain additional knowledge that might not be covered in the learning programs, while practicing ordinary quizzes is a relatively good way for them to test and improve their outstanding performances. This finding indicates that the proper uses of additional utilities (quizzes and library) might have a positive effect when combined with adaptive learning, acting as a trade-off to the entirety of the learning system. Overall, the teacher claimed that she is satisfied with the use of the prototype to support adaptive learning.

5. CONCLUSION

In summary the system prototype tested has received satisfactory results. Even though the development process was not fully successful in terms of addressing the limitation of time and effort required for the collection and refinement of learning content and material – which made the prototype less adaptive, the databases have been reasonably managed to ensure proper adaptations, contributing to the accomplishment of the functions of content management and learner activity tracking.

REFERENCES

- [1] L. Nguyen and P. Do, "Learner model in adaptive learning," *World Academy of Science, Engineering and Technology*, vol. 45, no. 70, pp. 395-400, Sep. 2008.
- [2] P. Dillenbourg and J. Self, "A framework for learner modeling," *Interactive Learning Environments*, vol. 2, no. 2, pp. 111-137, Jun. 1992.
- [3] R. Sunitha, R. Shyamala, and G. Aghila, "Towards learner model sharing among heterogeneous e-learning environments," *International Journal on Computer Science and Engineering*, vol. 3, no. 5, pp. 2034-2040, 2011.
- [4] J. Sánchez and N. Baloián, "Modeling 3D interactive environments for learners with visual disabilities," in *Proceedings of the 10th International Conference on Computers Helping People with Special Needs*, Linz, Austria, 2006.
- [5] S. Somyürek, "Student modeling: Recognizing the individual needs of users in e-learning environments," *International Journal of Human Sciences*, vol. 6, no. 2, pp. 429-450, 2009.
- [6] T. Ley, B. Kump, A. Maas, N. Maiden, and D. Albert, "Evaluating the adaptation of a learning system before the prototype is ready: A paper-based lab study," in *Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization*, Trento, Italy, 2009.
- [7] C. Mulwa, S. Lawless, M. Sharp, and V. Wade, "The evaluation of adaptive and personalised information retrieval systems: A review," *International Journal of Knowledge and Web Intelligence*, vol. 2, no. 2/3, pp. 138-156, 2011.
- [8] S. Weibelzahl, "Evaluation of adaptive systems," in *Proceedings of the 8th International Conference on User Modeling (UM) 2001*, LNAI 2109, M. Bauer, P.J. Gmytrasiewicz, and J. Vassileva, Eds. Springer-Verlag Berlin Heidelberg, 2001, pp. 292-294.
- [9] D. Chin, "Empirical evaluation of user models and user-adapted systems," *User Modeling and User-Adapted Interaction*, vol. 11, no. 1-2, pp. 181-194, 2001.
- [10] C. Gena, "Methods and techniques for the evaluation of user-adaptive systems," *The Knowledge Engineering Review*, vol. 20, no. 1, pp. 1-37, Mar. 2005.
- [11] C. Mulwa, S. Lawless, M. Sharp, and V. Wade, "The evaluation of adaptive and personalised information retrieval systems: A review," *International Journal of Knowledge and Web Intelligence*, vol. 2, no. 2/3, pp. 138-156, 2011.
- [12] 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, no. 5, pp. 383-453, Dec. 2010.
- [13] A. Paramythis and S. Weibelzahl, "A decomposition model for the layered evaluation of interactive adaptive systems," in *Proceedings of the 10th International Conference on User Modeling*, Edinburgh, UK, 2005.
- [14] L.V. Velsen, T.V.D. Geest, R. Klaassen, and M. Steehouder, "User-centered evaluation of adaptive and adaptable systems: A literature review," *The Knowledge Engineering Review*, vol. 23, no. 3, pp. 261-281, 2008.
- [15] D. Lavery, G. Cockton, and M. Atkinson, "Heuristic evaluation: Usability evaluation materials," University of Glasgow, Scotland, 1996.
- [16] J. Nielsen and R. Molich, "Heuristics evaluation of user interfaces," in *Proceedings of the ACM CHI 90 Human Factors in Computing Systems Conference 1990*, Seattle, USA, 1990.
- [17] A. Paramythis and S. Loidl-Reisinger, "Adaptive learning environments and elearning standards," in *2nd European Conference on Elearning*, Glasgow, Scotland, 2003.

BIBLIOGRAPHY OF AUTHORS

	<p>Nguyen Viet Dung is a research student in the Department of Computer and Information Sciences, at Universiti Teknologi PETRONAS (UTP). Suziah Sulaiman is a senior lecturer in the department while Mohd Fadzil Hassan is the Dean, Center for Graduate Studies at UTP. Both of them are the research advisors for Nguyen Viet Dung.</p>
--	---