Malaysian Smart School Courseware Usability study: The Effectiveness of Analytical Evaluation Technique Compared to Empirical Study

Azizah Jaafar
Department of Information Science
Faculty Technology and Information Science
Universiti Kebangsaan Malaysia
43600 UKM, Selangor
MALAYSIA
aj@ftsm.ukm.my

Abstract: - Usability evaluation methods are used to evaluate the interaction of the human with the computer for the purpose of identifying aspects of interaction that could be improved t increase usability. This research was carried out to determine the effectiveness of analytical evaluation technique that was designed, developed and named as Jalan Rentasan Kognitif (JRK) evaluation technique. This evaluation technique was conducted on thirty-seven surrogate users that consist of courseware developers, teachers and university students using the Malaysian Smart School Mathematics courseware. The usability result was then compared statistically with the empirical usability evaluation called Task Analysis Exploratory through Observation using video camera (TAEO). The respondents of the empirical method were secondary school students who used the same courseware as one of the learning aids in school. Usability indicators used in the study were usability problems and users' satisfaction. The study showed that the effectiveness of the two techniques was quite similar if they were used independently during the evaluation process. The usability indicators obtained in the study could be used as references in the usability comparisons of other educational courseware.

Key-Words: - Usability Evaluation Technique, Task Analysis, Surrogate user, Usability problems.

1 Introduction

Human Computer Interaction (HCI) is a study that involved interaction among users, computer and tasks [6]. It also [9] related to the understanding how users use computer system to perform specific task in specific environment until the best system can be built to satisfy users requirements. Interaction between human who performed the task on the computer in specific environment consists of four elements, (i) User, (ii) Specific tasks, (iii) Specific environment and (iv) Computer System. The main element of the computer system is interface. The interface works as a medium that enables users to communicate physically and cognitively with the computer system [9]. Even though the interactions between human and computer through the system interface involve user's physical actions towards input devices in order to produce output, but the interactions are influenced by a lot of factors that could not been seen by the users [6].

2 User and Usability Concept

Usability is a term that could usually be found in the discussions of User Centered Design (UCD) or in rules and principle of design. Usability and Usability Engineering (UE) have taken place in software engineering discipline started end of this decade. Initially, usability was based on user friendliness concept. This concept was very popular and it had been used to explain either a system was easy or difficult to be used. This definition was given by Oxford dictionary and had been argued by HCI experts. Norman and Draper [10] specifically mentioned that they do not want this concept being used in User-Center Design concept that has been introduced by them.

Usability has been defined at earlier stage based on user friendliness concept (user can easily use any particular software. Conversely, ISO DIS 9241-11 defines usability as "… *the effectiveness, efficiency and satisfaction with which specified users could achieve specified goals in particular environment*" [7]. Effectiveness in this context refers to the extent to which goal can be achieved. Effectiveness measures by the ratio percentage between the achievement and goal. Efficiency meanwhile refers to the amount of effort required to accomplish a goal. Users can be said more efficient if they can accomplish a specific goal with less amount of effort as compared to the expected effort. Satisfaction refers to the level of comfort that the users feel when using a product and how acceptable the product is to users as a means of

achieving their goals [7]. ISO definition highlighted that software usability was depending on users, goals to be achieved while using the software and environments where the system being used. ISO definition clearly explains that users are the main important element in usability. User's element influenced by four main criteria which are user's experience, user's domain knowledge, user's background and age.

3 Usability Evaluation

Usability evaluation has been accepted as a tool that can provide answers regarding whether the system was designed and developed according to the user's requirements [14]. The evaluation process consists of several task activities being implemented in order to support design [9]. The evaluation conducted in the early phase of design is to establish the right design. The main usability problems and problems related to the features and functional problems were being given more emphasis in this phase [2]. Once the design can be accepted and the goal of the design is clearly understood, the focus will be switch to the details and precise usability problem in order to design a complete interface that is free from usability problem

3.1. Evaluation Technique

A few decades ago, various evaluation techniques have been developed to evaluate the interface design of a system. Some of the techniques were suitable to be used at the earlier stage of development process and the rest were suitable to be used once the system has being implemented. There are many types of evaluation techniques and they are different in the aspects of: (i) Based on users or experts (ii) Numbers of training needed before the evaluation process and (iii) Emphasis given to the user's task and usability problems and many more [8 [16]; [6]; [3]; [5] used total number of usability problems in their research as a usability indicator evaluate and compare the usability evaluation techniques performance. Meanwhile, Sweeny, Maguire dan Shackel [15] has developed HCI evaluation framework and suggested that time to complete the specific task as one of the usability evaluation indicator. Grissom and Perlman [4] used time as a usability evaluation indicator when they compared the computer software. In this research study, the total number of usability problems is used as the main usability indicator in comparing the effectiveness of the two types of usability evaluation techniques. They are the analytical evaluation technique, called Jalan Rentasan Kognitif and the empirical evaluation technique, called Pemerhatian Analisis Tugasan [1]. Jalan Rentasan Kognitif (JRK) is an evaluation technique that was developed based on Usage Centered Design and Cognitive Walkthrough evaluation technique [16].

3.2 Research Questions

Several research questions are developed to achieve research objectives. The research questions are as follows:

i. Is the effectiveness of the developed JRK evaluation technique is as the same as an empirical evaluation technique (PAT)?

ii. How are the perception levels of both users (the teachers and students) towards the capability of the evaluated courseware?

iii. Are there differences in the perception levels of the surrogate users and students toward usage of the courseware? If any, are their experiences in using the courseware affect the perception levels?

3.3 Construct, Concept and Variables

Several hypotheses are constructed to answer the above research questions. The hypotheses a constructed based on the research conceptual framework and the highly usable variables as shown in Figure 1.

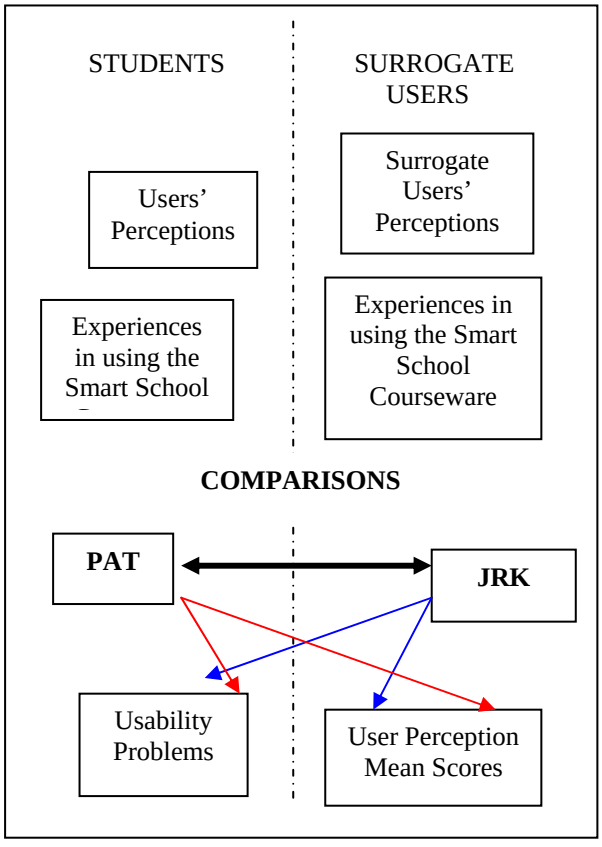

Figure 1 Research Conceptual Framework and Highly Usable Variables

The hypotheses constructed are as follows:
i. The performance/usability problem means found by the students is equal to the performance/usability problem means found by the surrogate users.
ii. The total mean score of perceptions obtained from the users are equal to the total m score of perceptions obtained from the surrogate users.
iii. The total mean score of perceptions obtained from the users and the surrogate users are equal at each experience level in using the evaluated courseware.

4 Methodology

The aim of this research was to identify the effectiveness of the analytical usability evaluation technique (JRK) compared to the empirical usability evaluation technique (PAT). Based on the research conceptual framework as shown in Figure 1.0, a methodology called RKPM was developed and the methodology was shown in figure 2 of the following page. The courseware that has been chosen as a case study was the Smart School Courseware that was installed in 92 pilot schools since 2002. The independent variables were then analyzed statistically to check either it will or will not affect the dependent variables significantly. The main usability indicators used are: (i) Usability problems (ii) Users perception means scores.

4.1 Research Techniques

Several research techniques involved in the research methodology and they are as below:
i. Observation: Observation is related to the tasks performed by users. This technique was named PAT (Task Observations Analysis). It was designed based on Usage Centered Design. Consists of forms that have sequence actions listed to verify steps being taken by users through experiments method in the lab. Observations were also conducted through video camera to support manual observations accurateness.
ii. Exploration. Exploration performed by the surrogate users assisted by a research tools named Jalan Rentasan Kognitif (JRK) through an evaluation technique called JRK Evaluation Technique.
iii. Questionnaire/Survey: The questionnaires are designed with 5 rated guided responses. Cronbach Alpha analysis was used to test the reliability of the instrument developed for this study. The questionnaires had been given to users and surrogate users.

4.2 Samples and Population

Generally, task assignments conducted in both of the

evaluation techniques represent ease of learning criteria
that would specifically show the effectiveness and efficiency of the evaluated courseware in the criteria. The measurement of user's performance would refer to usability indicators. Usability indicators that were used in this research were usability problems
[16] and user perceptions [11]; [4].

4.3 Jalan Rentasan Kognitif (JRK)

JRK evaluation process involves several analysts that act as surrogate users. The analysts comprise software developers, teachers and research assistant. They predicted what would induce the real users to solve problems, formulate current goal, and choose suitable actions based on the feedbacks of the system from the given actions.

Later, the analysts re-evaluate previous goal, do some modification and develop new goal for the next actions. The strengths and accurateness of this process, depends on the analysts' interpretation that were completed based on real user background and profile (Wharton 1994). JRK evaluation technique had been developed based on Cognitive Walkthrough. However, this technique had gone through several modification processes. In addition, the procedure has been made easy and simple to be used without wide knowledge requirement in HCI. Surrogate users that involved in the evaluation process were 37 people that comprise 18 software developers, and the rest were teachers, university students and researchers that involved in software development process. All of the respondents involved were not being paid and they were given their cooperation to help in using the evaluation technique in this research. Surrogate user samples were represented in the Table 1.

Table 1 Students and Surrogate Users Based on Location of the Regions

ocation	**Schools lents)**	**Surrogate Users**	
		eachers, esearch sistants & niversity	ware Developers

		Students	
h	2 (48)	3	2
ral	3 (72)	10	13
h	2 (48)	2	1
Coast	2 (48)	2	2
Malaysia	2 (48)	2	0
l Samples	11 (264)	19	18

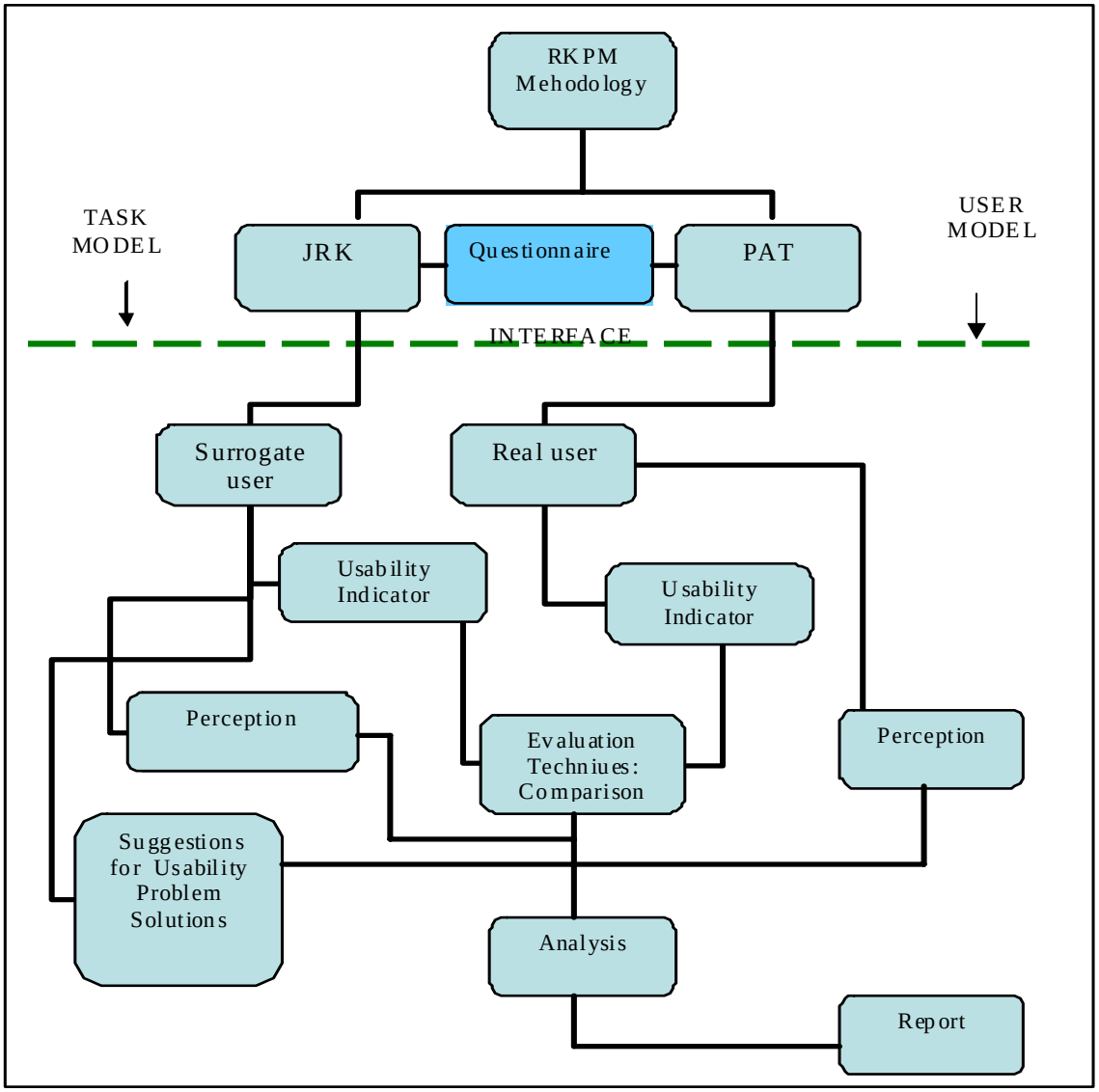

Figure 2 RKPM Mehodology

4.4 PAT Evaluation Technique

Population samples for PAT evaluation technique consist of form 1 students and teachers of the Smart Schools (as a case study). Since Malaysia is quite a sizeable country, the cluster sampling method has been used based on location of schools and the location of the work places for the surrogate users (Refer to Table 1).

There were two groups of research sample that involved for the both techniques. The first group sample consists of 24 students that represent each school from 12 chosen Smart School. The chosen schools were from five locations category that had been discussed previously. Clustering method being used in selection of 24 students based on class performance (Excellent, Good, intermediate and weak). The samples would

explore and complete the predetermined tasks analysis based on the specified modules in the courseware The second groups were the analysts or the surrogate users. They went through the same modules and complete them with the assumptions that they were the real users. Questionnaires were distributed to both samples before and after evaluation process.

5 Result

Results of this research were divided into two parts as below:

i. JRK and PAT Comparison Analysis

Independent sample t test was carried out to compare means of the two groups of evaluation techniques. Based on Lavene test, the variances for both samples are assumed not equal and it was found that the differences between the usability problem means of the students and the surrogate were not significant with high p value. Therefore, the null hypothesis was accepted and this showed that the usability problems obtained in the JRK evaluation process was as equal as the usability problems obtained in PAT empirical evaluation process. This also means that the effectiveness of the JRK evaluation technique was as equal as the effectiveness of PAT, one of the empirical evaluation techniques.

ii. Perception

Following the introduction of the courseware to the Smart School Pilot Project, it was found that only 12% of the 609 students claimed to have frequent use of the Smart School courseware in school. Most students (73%) however revealed that they seldom used the courseware. Whilst, 11% claimed to had limited exposure to the courseware and the total of 3% of these students had never used the courseware in school. The result of the study is shown in fig. 3

As explained earlier, a study was carried out at selected schools' samples that were located in 5 different regions and Fig. 4 indicated that the southern region with 25% and East Malaysia with 19% were ahead of the other regions in term of usage frequency of the courseware during lessons in their respective schools.

The differences in the usage frequency in school regions as shown in fig 4 were addressed by a hypothesis test. The test indicated that the usage frequency of the Smart School courseware was significantly affected by the schools' regions with p value <0.0005.

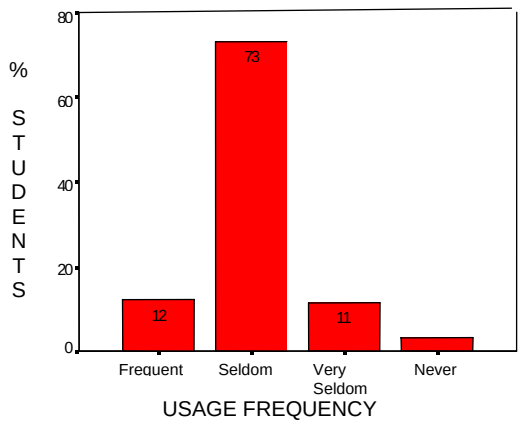

Fig. 3 Frequency Usage of the Smart School Courseware

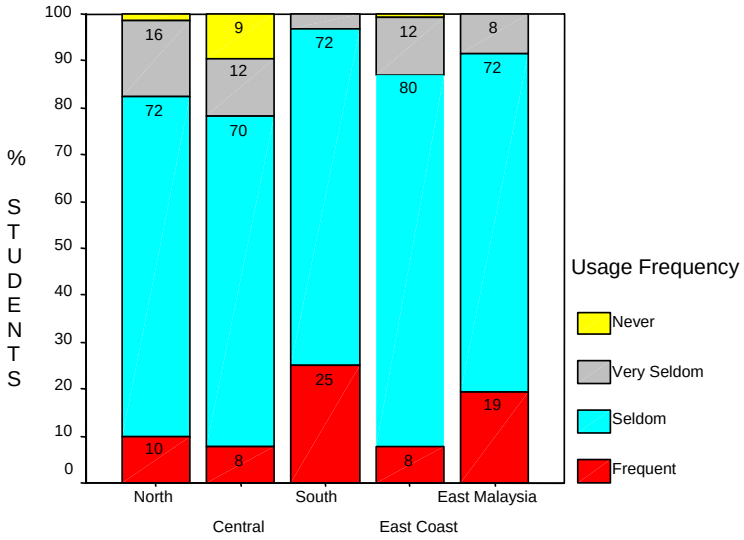

Fig 4 Usage Frequency of the Courseware on the Regions Basis

This section also covered the perceptions of the teachers and the students toward overall capability and potential of the Smart School courseware especially in assisting the teachers and the students in teaching and learning processes. User's opinions were also obtained regarding the ability of the evaluated courseware to draw users and teachers interest to use it frequently, enhance their understanding in learning and the possibility of the courseware to replace conventional teaching and learning materials in the near future. The mean score for students' opinion regarding the perceptions was 6.0 (good). Meanwhile, the mean score for the surrogate users was 3.8 (medium weak). This means that the mean scores of the perceptions of these two group categories were quite distinct.

The distinction of both scores was then tested statistically using ANOVA. The result showed that there was a significant different between means score of the perceptions of the user and the user surrogate with the p value was less than 0.05. This means that the real user thought that the Smart School courseware was attractive, highly interactive, and usable and it has a potential to replace the conventional teaching and learning materials. Unfortunately, the surrogate user especially the teachers had thought otherwise. Therefore, the second

hypothesis was rejected. In the case of experience levels factor of both users in using the courseware, they were found to have no significant impact on the means score of perceptions.

6 Conclusion

In conclusion, the effectiveness of analytical usability evaluation technique called 'Jalan Rentasan Kognitif (JRK)' is equivalent with empirical evaluation technique. Consequently, JRK can be used independently. JRK is an evaluation technique that is performed by surrogate user. It could be used iteratively and at any stages of development processes with less cost. It could be an answer for the most objectives of the user interface system designs that is to make user requirements achievable. Usability problems and user perceptions obtained from this research can be used as usability indicators for courseware/software applications for students and teachers. This result could also be used as a guideline to evaluate effectiveness and efficiency of courseware interface usability for the mathematic subject. Meanwhile, the perceptions of the students and teachers obtained from this research provide ideas, guidance and alternative opportunity for the teachers at Smart School to vary their teaching and learning techniques in order to improve students' performance in any particular subjects.

References

[1] Azizah Jaafar, Developing and Designing Tool and Methodology for an Evaluation of Education Courseware. PhD Thesis. Bangi:Universiti Kebangsaan Malaysia. 2005.

[2] Azizah Jaafar & Cassaigne, N. Cognitive walkthrough: Improving educational software usability*: Proceedings of the International Conference on Computers in Education /SchoolNet2001*, Seoul, Korea, 2001. v. 1, pp. 472-480.

[3] Cuomo, D.L.& Bowen, C.D. Stages of user activity model as a basis for user-system interface evaluations. *Proceedings of Human Factors Society 36th Annual Meeting. Innovations for Interactions*. 1992. pp.. 554-559.

[4] Grissom, S.B. & Perlman, G. StEP(3D): A standardized evalaution plans for three dimensional techniques. *Int J. Human-Computer Studies*. 1995. 43 (12): pp. 15-41.

[5] Jeffries, R. Usability problems report: Helping evaluators communicate effectively with developers. In. J. Nielson & R. L Mack (ed.). *Usability Inspection Methods.* 1994. pp. 121-129. New York: Wiley.

[6] Johnson, Peter. *Human-computer interaction: Psychology, task analysis and software engineering.* Berkshire: McGrawHill. 1992.

[7] Jordan, P. W. *An Introduction to Usability.* London: Taylor & Francis. 1998.

[8] Mack, R. L. & Nielsen, Jacob. Executive summary. In. J. Nielson & R. L Mack (ed.). *Usability Inspection Methods.* pp. 132-149. New York: Wiley. 1994.

[9] Nielsen, J. Heuristic evaluation. In. J. Nielson & R. L Mack (eds). *Usability Inspection Methods.* pp 3-7. New York: Wiley. 1994.

[10] Norman, D.A. Cognitive engineering. In. Donald A. Norman & Stephen W. Draper (ed.). *User Centered System Design: New Perspectives in Human-Computer Interaction*. pp. 289-299. New Jersey: Erlbaum. 1986.

[11] Novara, F. Usability evaluation and feedback to designers – An experimental study. In. Shackel B (ed.). *Human-Computer Interaction: INTERAT*. 1987. pp. 337 - 340. Armsterdam: Elsevier.

[12] Preece, J. *A guide to usability: Human factors in computing*. Wokingham: Adddison-Wesley. 1993.

[13[11] Rieman, J., Franzke, M. & Redmiles, D. Usability evaluation with the Cognitive Walkthrough. *Proceedings of Human Factors in Computing System.* 1995. pp 213-218.

[14] Rubin, J. *Handbook of usability testing: How to plan, design and conduct*. Canada: John Wiley & sons. 1994.

[15] Sweeney, M., McGuire, M. & Shackel, B. Evaluating user-computer interaction: a framework. *Int. J. Man-Machine Studies*. 1993. v. 38: pp. 689-711. New York: Springer.

[16] Wharton, C., Rieman, J., Lewis, C. & Polson, P. The Cognitive Walkthrough Method: A Practitioner's Guide. In J. Nielson & R. L Mack (eds.). *Usability Inspection Methods.* pp. 77-89. New York: Wiley. 1994.