Expert Evaluator Decision Tree Tool to Support Heuristics Educational Game (HEG) Evaluation

[1]Shirindokht Farhady, Amirfarhad Farhadi, [1]Dr . Azizah Jaafar, and [1]Dr . Riza Sulaiman

[1]Faculty of Information Science and Technology, Department of Visual Informatics, University Kebangsaan Malaysia (UKM), Malaysia

Emails: shirin_d_f@yahoo.com; amirfarhadfarhadi@gmail.com

Abstract: A great potential of educational computer games in learning process has prepared a good opportunity for teachers to apply "playful elements" in learning environments. As a result educational computer games are widely used as tools to entertain, instruct, motivate and develop skills. The effectiveness of a computer game in sense of help players to learn has been matter of game education industry for years. One of the best methods to assess game usability is game heuristics. Heuristics is an evaluation method that was developed and improved tremendously from general interface evaluation to specific evaluation such as computer game application. This study aimed to provide an easy to use, yet valid method to best evaluate usability of a game, it focuses on two techniques based on two types of evaluators; the expert evaluators and real users. A decision tree, which is based on two levels of evaluation, is the outcome of this study with which, game players with the medium understanding of game can benefit from. Here first a collection of game evaluation criteria are collected from literature and then has been verified to be valid and related by experts.

[Farhady Sh, Farhady A.F., Jaafar A, Sulaiman, R. **Expert Evaluator Decision Tree Tool to Support Heuristics Educational Game (Heg) Evaluation**. *Life Sci J* 2013;10(4s):569-575] (ISSN:1097-8135). http://www.lifesciencesite.com. 87

Key word: evaluation, computer game, heuristics, decision tree

1. Introduction

Nowadays, a rising number of teachers have been trying to motivate their student to learn through assimilating educational games within teaching process (Roblyer, 2006). In learning environments a great potential of educational computer games in learning process has prepared a good opportunity for teachers to apply "playful elements". Educational computer games and other types of computer based applications have been extensively used in learning process; either directly or indirectly (Omar & Jaafar, 2008). As a supporting tool, they provide a proper environment for teaching and learning (Omar & Jaafar 2008), while they engage students' motivation during learning process (Virvou & Katsionis, 2008).

Although, considering the existence of all kinds of educational computer games, it is wondered if any of them has the ability of actually improving student learning. Many researches show the advantages of educational computer games. Percival and Ellington (1993) believe that educational computer games could provide a user, a virtual world to experience a real world, and they can reinforce teaching process by illustrating. Betz (1995) stated that computer games will improve learning power through visualisation, experimentation, and creativity which will make learning more fun (Kafai, 2001). In the works of Gee (2004) and Koster (2005), they frequently mentioned that via educational computer games the process of learning transformation will be facilitated. Other factors also namely, improving strategic thinking, the practice of logic, memory, problem solving, and critical thinking skills are stressed by Leung and Yu (2007). Wangenheim and Thiry (2008) declared that the educational computer game prepares user learning environment in real situation and at the same time creates self confidence among the users.

Dempsey et al. (1996), in his study proved that educational computer games could entertain, instruct, change attitude and develop skills. Therefore, evaluation of educational computer games is a crucial phase in order to assess the positive effects of such tools (Wangenheim & Thiry 2008; Seifer & Holmes 2009). However, lack of the measurement in this case still remains an issue (Papaloukas & Xenos, 2008; Wangenheim & Thriy, 2008; Lin, 2009; Sharda, 2007). The existing evaluation methods have been developed to be applied for evaluating computer software, though, same methods are mostly employed to evaluate computer games without any special adjustment such as Nielsen's heuristics (Federoff, 2002). Therefore, it obviously needs to develop specific evaluation methods for computer games.

Meanwhile, determining the expert level is a crucial function while there is not any benchmark for such processing. Even though, enough literature could be obtained about expert evaluation but information regarding expert levels, which can be determined by several criteria such as work experience are really scarce.

As abovementioned factors the following are the short summaries for problem statement of this research:

- There are a lot of educational games in market; still, the easy tools which users want to use is not available, so tools cannot provide a good feedback for them about the educational games which are going be purchased.
- Expert's evaluators (i.e. heuristic evaluation) require a tool that can provide them information massive but simple calculation that assist them to do make decision.
- The existing evaluation methods have been developed normally to be applied for evaluating computer game, however, there is no available specific method yet to be employed to evaluate educational computer games (Gonzalez &Masip, 2009).

The significance of this research is identifying the main characteristics of educational computer games that allow the design and development a simple decision tree based on expert evaluators. The level of learning is crucial when a game is categorised as an educational material. In addition, although there are online tools to help users with finding and selection of educational game the writer could not find any available on-line evaluation system for educational game which can extend its domain for collecting information. It also can be accessible anytime and anywhere. This system is based-on expert heuristic evaluation and the research will create a general on-line evaluation for educational computer games.

2. Literature Review

In order to examine the product, Evaluation process can be conducted in two methods (Zaidatun, 2009), Formative and Summative. Formative Evaluation is a process of examining a product during the producing steps and Summative Evaluation refers to the testing of the product after it is produced. (Bhola,1990), explains the formative evaluation as "a method of judging the worth of a program / product while the program activities / products, are forming or happening formative evaluation focuses on the process. "And Summative Evaluation is "a method of judging the worth of a program / product at the end of the program activities / products. The focus is on the outcome."

The statement is supported by Western Michigan University (2008), and instructional assessment resources (2010), that define formative evaluation as a research carried out throughout the development phase of an educational program, in order to provide feedback to improve the program, whereas, summative evaluation has been defined as a research at the end of the educational program to consider the good points of such program to help the process of decision making for the program expansion, continuation, adoption or termination.

Looking at the results of others can be a good starting point for evaluation. Designing a meta-study, considering the first impressions of people, self-reflections and checking books, newspapers and databases for reviews are among the options in this method.

While Zaidatun (2009), has defined four methods for evaluation, others have different ideas. Traditional Laboratory-Based Usability Evaluation and Remote Evaluation (Kodiyalam 2003), Heuristic Evaluation (HE) and Structured Expert Evaluation Method (SEEM) (Bekker et al. 2009), heuristic evaluation and user testing(Tan et al. 2008), inquiry and experimental methods (Papaloukas & Xenos 2008), and heuristic evaluation and analytic hierarchy (Delice & Gungor, 2009), have been compared differently.

Quality assurance (QA) staffs in a game company usually test the game mechanics, which include the way that a player is able to move in the game environment (jump, run, drive, etc.). These people will make sure that there are no bugs in all the games that are shipped. Such mechanics are created by animators and implemented into the game engine by the programmers and then, level designers put them into the environment of the game. These three processes as mentioned comprise the game mechanics.

Game play involves all the challenges and issues that one player faces in order to try and win the game. Game play is defined by Crawford (1982), as pace and cognitive effort.

Based on the genre of the game (First person, role-playing, first person shooter, adventure) and the platform (personal computer, console, coin-operated machine) all the aspects mentioned are different. For example, adventure game used to mostly be played on computers but now they are appearing in consoles as well. Using controllers' buttons are not very popular among adventure gamers. In this way, the usability of a game is the same as the other software and therefore, usability only can be evaluated while considering the context.

In such case, Nielsen (2003), who has been called as a leader of web usability defines this concept as a "quality attribute that assesses how easy user interfaces are to use". The term "usability" illustrates different approaches to improve ease-of-use throughout the design process. It can be described as a capability to be used easily, effectively, and

satisfactorily (Shackel 1991). To support that, Dillon (2001), claims that "Usability is a measure of interface quality that refers to the effectiveness, efficiency and satisfaction with which users can perform tasks with a tool."

Moreover, as Gaffney (1999), stresses usability is a technique to make sure that the anticipated users of a systems can conduct the tasks intended. Effectively, satisfactorily and resourcefully, Rafidah and Jaafar (2008), explain a usable system as:

> *"A usable system enables users to achieve their task and goals quickly, easily, effectively and the users satisfied with the outcomes."*

Therefore, effectiveness, efficiency, and satisfaction are the three basic usability metrics. Effectiveness is to what extent specific goals of a system are achievable (Park & Lim, 1999). Erik and Morten (2000), explain it as the accuracy and completion of specified goals achieved by user, while, "quality of solution" and "error rates" are its indicators. A system has effectiveness while it can provide information effectively to achieve certain goals (Jeng, 2005). In this case, the number of correct answers can measure the level of effectiveness. Shortly, effectiveness refers to the accomplishment of a goal by user (Papaloukas & Xenos, 2008).

Decision trees (DTs) represent one of the most important and popular solutions to the problem of classification. They have been shown to have excellent performance in the field of data mining and machine learning (Alhammady, 2006). Decision trees are a simple, but powerful form of multiple variable analysis. They provide unique capabilities to supplement, complement, and substitute for

- Traditional statistical forms of analysis (such as multiple linear regression)
- A variety of data mining tools and techniques (such as neural networks)
- Recently developed multidimensional forms of reporting and analysis found in the field of business intelligence (SAS Customer Support. 2012).

Decision trees attempt to find a strong relationship between input values and target values in a group of observations that form a data set. When a set of input values is identified as having a strong relationship to a target value, then all of these values are grouped in a bin that becomes a branch on the decision tree. These groupings are determined by the observed form of the relationship between the bin values and the target. AHP (Analytic Hierarchy process is the most common way of constructing a decision tree using multiple criteria and weighted analysis conducted to weigh the criteria and form a weighted hierarchical decision tree (Kalton,1983).

This research intend to create and organize a weighted desicion tree evualator for educational game. In order to provide that, first a list of most related criteria which are related to the scope of educational game is needed, which was done through intensive literature review. After gathering the criteria, the first draft of the educational computer game evaluation criteria was drawn.

In this framework, six categories will form the final evaluation index for educational games. The list of these factors more or less exists in the literature of game evaluation but not much investigation has been done in educational games.

Considering all the facts, writer has not just concentrated on the factors, which were found in the educational game, literature and have tried to expand it to the new factor taken from game evaluation. The effectiveness of new and exist categories and sub-factors will be investigated through this research, and as it mentioned earlier the effect of each factor and consequently sub-factors in the final index would be analysed and investigated based on the experts' views.

3. Methodology and data collection

As every other research in the area of information technology, this study is based on the process of proposing approach and then investigating the effect of the proposed approach using quantitative analysis that will contain data collection among certain groups of experts whose job expertise are related to game industry or more preciously educational game experts. The results of this investigation will be the main material for discussion and the important results would be highlighted. In order to design a highly related, yet simple questionnaire which best fits within the requirements of this study, a deep literature review among the available resources is conducted to best fulfil the requirements and criteria which are needed to be investigated for the whole study life cycle. In the initial planning phase, the objective of the project along with the scope and the high level methodology is extracted and defined so the project will find its path through literature review. As Teijlingen and Hundley (2002), stress pilot study amounts to a smaller version of a full-size study which can be seen a crucial element of a good study. To support this, Gay and Mills (2009), defined it as a dress rehearsal while it follows every procedure exactly as planned to identify unanticipated problems. In this study, before conducting the actual research, experts were chosen for the pilot study through the questionnaire. Pilot study was conducted to identify whether or not the research instruments can be proper in the actual

study. The same research procedure was done to conduct the pilot study.

Although random sampling techniques prepare the best opportunity to get unbiased samples, researchers cannot always use random sampling due to practical constraints (Gay & Mills, 2009). The chosen method to select the sample for this study is convenience sampling that is non-random approach. According to Gay & Mills (2009), it is a process of involving whoever happens to be available in the sample.

Therefore, in order to do the evaluation process, the questionnaire was sent to about 993 experts in the area of programming with focus on game programmers around the world. In order to do so, first a list of these experts' names were extracted from Linkedin.com website and subsequently an email was sent to them containing the link for the online survey along with a cover letter about the purpose of the research and some guidelines on how to fill in the questionnaire. From this list, unfortunately a small number of them were willing to participate in the survey; thus the researcher received only 20 responses. These responses were regarded as the basis for evaluating the relevancy of the selected criteria in the framework for the educational game evaluation and based on these results the first framework was evolved to second framework.

4. Results

In order to compare the effect and importance of each of the main categories from the respondents' point of view, the weighted average for each of them was calculated which is shown in Table 1.

Table 1. Main Categories Weighted Average

Main Category	Weighted Average (0-5)
Game Interface (GI)	4.04
Game Mechanics (GM)	3.49
Game Content (GC)	3.64
Game Playability(GPL)	3.64
Game Pedagogical (GPD)	3.93
Game Feedback and Immersion (GFI)	4.08

For better understanding, the results are shown in Figure 1 as well.

As Figure 1 shows, the most important categories are Game Feedback and Immersion (GFI) and Game Interface (GI) and the least important category is Game Mechanics (GM).

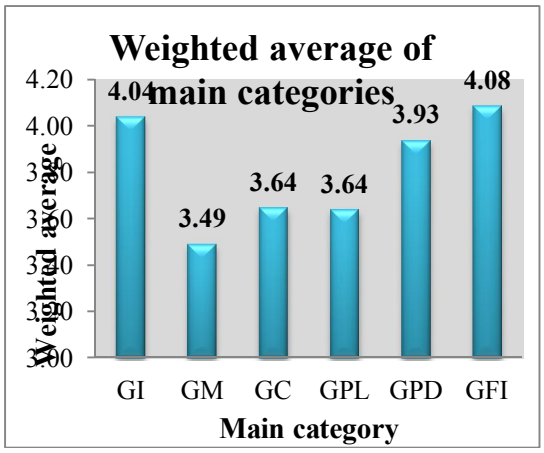

Figure 1. Main categories' weighted average

After finding the most and least effective categories from the respondents' point of view, it is needed to investigate the effect of such criteria for each category. The device that the player interacts with the game is called the game interface. The criteria considered in this research for GI are:

- The controls, which can be customized and are based on the standards setting of the industry.
- Controls, which are intuitive, have similarities with natural method of doing such actions.
- The interface ought to be non- intrusive to the possible extent.
- At any point desired, the players can know about their scores and status in the game.
- In order to shorten the learning curve, follow the standards set by the gamin community.
- Interfaces should be consistent in colour, control, dialog design and typography.
- The layers of the menu of an interface must be minimum.
- The meaningful feedback can be provided through using sounds.
- Readers are not expected to read manuals.
- Layers should be able to save games in different states.
- Art should be used in a way to its function.

In order to determine the importance of each of the above criteria, the weighted average of the respondents' answers for each attribute was calculated and shown in Figure 2.

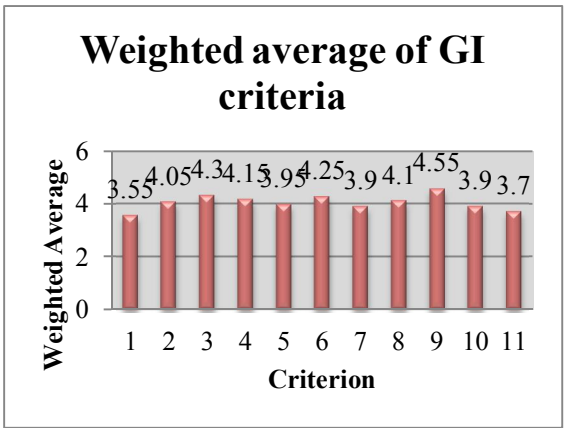

Figure 2. Weighted Averages of Game Interface (GI) Criteria

As Figure 2 shows the most weighted average is for the 9th criterion, which do not look forward to the user reading a long manual as an necessary method of learning, and the least weighted average is for the 1st one, which is the customizable controls based on industry standard settings.

In order to find out the level of agreement among different respondents about effectiveness of each criterion, the standard deviation for the responses to each question is calculated and shown in Table 2.

Table 2. Standard Deviation for Responses to each Question in GI Category

Question No.	1	2	3	4	5	6	7	8	9	10	11
Standard Deviation	0.89	0.76	0.92	0.88	0.89	0.72	0.85	0.85	0.76	1.25	1.03

As Table 2 illustrates the standard deviation for almost all of the questions is below 1, which shows the responses where close to each other, and the respondents had almost the same opinion about the questioned criterion. Only question 8, which is about using sound to provide meaningful feedback, has the standard deviation of 1.25.

In order to evaluate a game, it should be registered in the system. The information fields, which are needed to register a game, include: name, type, company, environment, OS compatibility, hardware compatibility and description. Figure 4.19 shows the game registration page in the website.

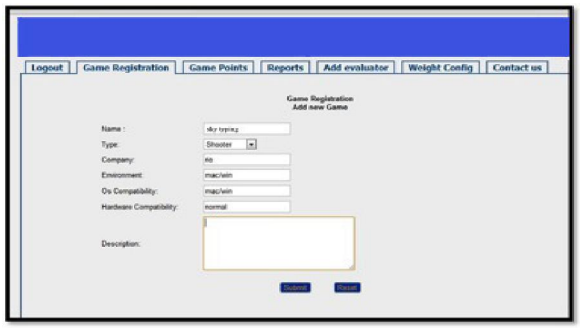

Figure 3. Game Registration Page

The "Game points" page is designed based on the categories and subcategories of the developed decision tree in this research. The user is asked to choose the game they want to assess and evaluate the particular game in each of the criteria and score it out of 5. These scores are used for the purpose of the overall evaluation of the game and further for comparing different educational games with each other. Figures 4.20, 4.21 and 4.22 are snapshots of the game points page.

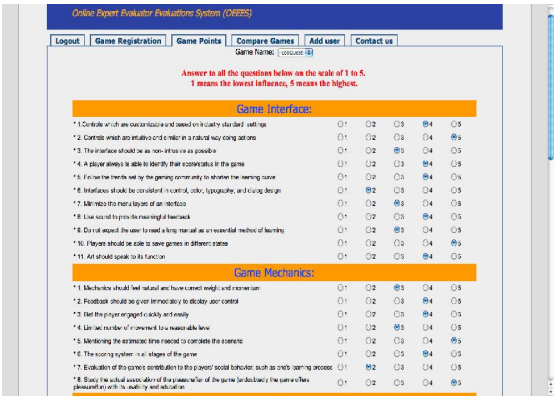

Figure 4. Game Points Page 1

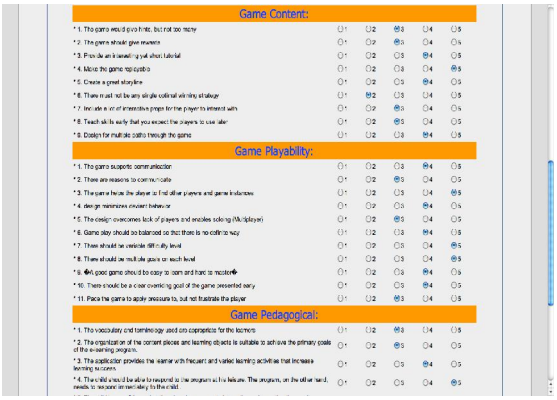

Figure 5. Game Points Page 2

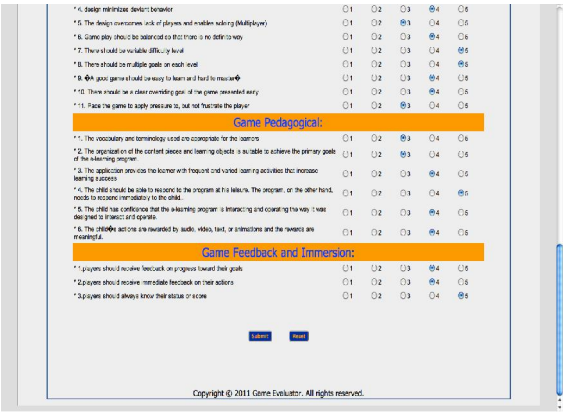

Figure 6. Game Points Page 3

The pilot study of developed website was done with knowledgeable potential users of the website. This pilot study aimed at identifying the possible drawbacks of the site in terms of usage, usefulness, interaction and user-friendliness. For this purpose a group of 6 potential users were chosen and were asked to fill in the questionnaire about the functionalities of the developed website. These people were mostly university lecturers, system engineer, and PhD students. They had at least 2 years of experience in the field of game industry. They described their experience in this field, which can be categorized as following:

- 2D and 3D game designer using flash and 3D studio max.
- Image processing for game development.
- Design and implement educational gaming.
- 3D MAX and other software for developing games specialist.
- Software Design and Testing specialist.
- Social Media and knowledge sharing application consultant related to the education.

The overall feedback on the developed website shows that it can be a useful tool for evaluating educational games and provides good results for the user. The results of the pilot study are discussed in next parts.

5. Conclusion

There were some problems and challenges, which the researcher faced during the study. One of them was lack of quantitative researches in the field of evaluating educational games since most of the researches in this field were qualitative and descriptive. The other problem was in accessing the experts in the field of the educational games for evaluating the proposed decision tree and defining the weight and importance of each criterion. The researcher had to send enormous amount of emails in order to get responses from the experts.

The main objective of this study was to design a decision tree and a decision tree for evaluating educational games. This decision tree should be comprised of different categories and subcategories with the weights assigned to each of them, which show the importance and weight of each criterion. These weights then are used to find out the score of further evaluated educational games.

It should be mentioned here that this research based its criteria selection on previous studies (e.g.Federoff 2002; Jegers 2007; Korhonen & Koivisto 2007; Melone 1980; Melone 1982; Papaloukas & Xenos 2008; Shelley 2001; Sweetser & Wyeth 2005). Although none of these studies weighted evaluation on the metrics which makes the result of this paper useful as they are both based on previous studies in their criteria selection and an improvement comparing to previous ones because of weighted decision tree that is created on experts' opinion which is presented in two step questionnaire assessment. To the best of the researcher's knowledge, there is similar study in the area of educational game evaluation to date which provide weighted decision tree as an outcome, hence this decision tree could be the basis of further studies.

References

1. Bekker, M.M., Baauw E. & Barendreqt W. 2008. A comparison of two analytical evaluation methods for educational computer games for young children, Cognition Technology and Work. *Journal of Cognition, Technology & Work* 10(2): 129-140.
2. Betz, J. A. 1995. Computer Games: Increase Learning in an Interactive Multidisciplinary Environment. *Journal of Educational Technology Systems*24(2): 195-205.
3. Bhola, 1990. The Design and Evaluation of a Virtual World-Based Learning Environment: Information Security Learning Using Second Life.*Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology,* pp. 22-27.
4. Crawford, C. 1982. *The Art of Computer Game Design*. Washington State University, Vancouver: Mcgraw-Hill Osborne Media. http://pdf.textfiles.com/books/cgd-crawford.pdf.
5. Delice, E. K. & Gungor, Z. 2009. The usability analysis with heuristic evaluation and analytic hierarchy process. *International Journal of Industrial Ergonomics*39: 934-939.
6. Dempsey, J. V., Lucassen, B. A., Haynes, L. L. & Casey, M. S. 1996. *Instructional applications of*

computer games.the Annual Meeting of the American Educational Research Association. New York: ERIC Document Reproduction Service.
7. Federoff, M. A. 2002. Heuristics and Usability Guidelines For the Creation and Evaluation of Fun in Video Games. MA thesis. Indiana University.
8. Gee, J. P. 2004. *Situated Language and Learning: A Critique of Traditional Schooling*. London: Routledge.
9. González, M. & Masip, L. 2009. Quantitative analysis in a heuristic evaluation experiment. *Advances in Engineering Software*40(12): 1271-1278.
10. Kafai, Y. 2001. The educational potential of electronic games: from games-to-teach to games-to-learn,*Conference on playing by the rules: the cultural policy challenges of video games. Chicago, Illinois*, http://culturalpolicy.uchicago.edu/papers/2001-video-games/kafai.html.
11. Kodiyalam, N. 2003. Remote Usability Evaluation Tool,Master of thesis Computer Science and Applications Blacksburg Virginia Polytechnic Institute and State University.
12. Koster, R. 2005. *A Theory of Fun for Game Design* Ed: CA, USA.Paraglyph Press.
13. Leung, W. L. & Yu, K. M. 2007. Development of online game-based learning for TRIZ, *Technologies for e-learning and digital entertainment Springer*: 925-935.
14. Lin, H.F. 2009. An application of fuzzy AHP for evaluating course website quality,*Computers & Education* 54: 877–888.
15. Nielsen, J. 1993. *Usability Engineering*. Boston, MA, USA, Academic Press.
16. Omar H. M. & Jaafar A.2008. Playbility Heuristics Evaluation (PHE) Approach for Malaysian Educational Games, Proceeding of International symposium on information technology: 1-7.
17. Papaloukas, S. & Xenos, M. 2008. Usability and education of games through combined assessment methods,*1st international conference on pervasive Technologies Related to Assistive Environments*.
18. Percival, F.& H. Ellington 1993. *Handbook of educational technology*. London, Kogan Page.
19. Roblyer, M. D. 2006. Integrating educational technology into teaching. Upper Saddle River Ed: NJ, Merrill Prentice Hall.
20. Seifer, S. D. & S. Holmes 2009. Tools and methods for evaluating service-learning in higher education. National Service-Learning Clearinghouse, http://www.servicelearning.org/instant_info/fact_sheets/he_facts/tools_methods/. [22 Apr 2012].
21. Sharda, N. K. 2007. Designing, Using and Evaluating Educational Games: Challenges, Some Solutions and Future Research, from http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-386/p08.pdf. [1 Mar 2012].
22. Tan, W., Liu, D.& Bishu, R.2009. Web evaluation: Heuristic evaluation vs. user testing. *International Journal of Industrial Ergonomics* 39 (4): 621–627.
23. Virvou, M. & Katsionis, G. 2008. On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE. *Computers and Education*50(1): 154 - 178.
24. Wangenheim, C. G. V.& Thiry, M. 2008. Empirical evaluation of an educational game on software measurement. *Empir Software Eng 14*: 418–452.
25. Zaidatun, T. 2009. *Web Evaluation*, Educational Multimedia. http://eprints.utm.my/11887/1/Pembangunan_Laman_Web_Untuk_Solid_Geometry_III_Bagi_Tingkatan_3_Menggunakan_Teori_Pembelajaran_Kontekstual.pdf. [3 Feb 2012].

2/20/2013