The Evaluation of an Online Virtual Game Environment (SimSafety) using HOU's Software Quality Laboratory

Michalis Xenos, Spyridon Papaloukas, and Nektarios Kostaras

***Abstract*—This paper presents the formative evaluation of an early version of SimSafety (an online virtual game environment) based on ISO9126 that was conducted in a Hellenic Open University specialized Software Quality Laboratory. The participants of the evaluation were children and parents that were exposed to Internet dangers safely through playing games. The paper presents the results of the evaluation, the changes in the environment based on these results and the evaluation procedure per se (goals, setting, equipment, methods, participants, findings).**

***Index Terms*—Online Game Environment, Software Quality Evaluation, ISO9126.**

I. Introduction

THIS PAPER presents the evaluation of an online virtual game environment using a specialized laboratory for software quality evaluation. This online environment has been developed under the framework of the EU Lifelong Learning Programme and the project "*SimSafety – Flight Simulator" for Internet Safety*" from a consortium of seven partners from six countries: Greece, Cyprus, Portugal, Romania, Finland and United Kingdom. Further details for the project can be found at: http://www.simsafety.eu.

Although SimSafety is not just a videogame, in some parts of this paper the term videogame is used, complying to the international literature, SimSafety is an *online virtual game environment* which supports: i) role-playing game scenarios, offered in real-time to ii) dynamic groups of online users, iii) exposed to (penalized) safety traps, which should be recognized and overcome through (accredited) appropriate game actions. Therefore SimSafety allow socializing and playing videogames, but these games are specifically designed towards a purpose: to expose children to Internet risks, but safely.

The evaluation of the SimSafety has been conducted in the Software Quality Laboratory of the Hellenic Open University (http:/quality.eap.gr) and the goal of the evaluation was to identify weaknesses of the SimSafety based on the project's quality plan. The evaluation was formative, i.e. it was performed in an early version of the SimSafety allowing the identification and correction of all the recorded problems and involving a number of primary users of the environment (children and parents).

The rest of this paper is organized as follows: section II is a brief literature review on usability of videogames and videogames and learning. Section III presents the evaluation framework and goals, the software quality laboratory, the methods and tools used for the evaluation and the outcome of the evaluation. Section IV discusses results from the evaluation, as well as the changes in the final version of the environment that were stimulated by the evaluation, while section V summarizes the conclusions of this paper.

II. Literature Review

A. Usability of videogames

In this paper, *game usability* is addressed as the degree to which a player is able to learn, control, understand, be intrigued and enjoy a game. This definition is accepted on an early informal survey of usability problems cited in critical game reviews and on playability heuristics [1]-[3]. Some usability issues seen in games are similar to those seen in other application areas, such as the need to design visual consistency and readability. Game's usability is highly related to how easily the player is able to start playing the game and to discover –and eventually use– all the features and functions available in the game. There is not a specific model that has been "adopted" by the scientific community for the measurement of games' usability. In an effort to define usability in relation to a game the ISO standard of usability ISO 9241-11 is used, in which three basic usability metrics are described: *effectiveness*, *efficiency* and *satisfaction* [4]. Effectiveness is linked with accuracy and total achievement of the goal set by the user; efficiency relates to the resources spent for the attainment of the objectives, and satisfaction affiliates with the users' mood.

Software Quality Research Group (http://quality.eap.gr) has developed a specialized laboratory as well as methods and tools used to measure software usability, some of which were used in this survey. However, games also have usability considerations that are not seen in other areas. For example, user errors are usually undesirable in other domains, but are expected in many games since they are designed to challenge users and to force them to develop new skills so that they can achieve in-game objectives [2].

The videogames like all other software types have an

interface that should provide effective and efficient means of interaction between the user and the videogame; but of course when studying the playability of a videogame, which is an indispensable part of usability, it is obvious that the three usability metrics do not have the same influence on the videogame.

The Microsoft Playtest Group has developed a sophisticated approach to videogame evaluation and this includes the important components of fun, ease of use, challenge and pace [5] and their methods include RITE - Rapid Iterative Testing and Evaluation [6]. The present research accepts the fact that the aspects affecting a game's usability are associated with the element of anatomy-structure of the game that they affect. The structure of the videogame consists of the game's interface, mechanism and game play. The game's interface is the device through which the user interacts with the game; the mechanism is the combination of animation and programming, whilst game play refers to the procedure that the user goes through so as to reach the game's objective. All three different areas are functional, must be satisfied and should be analyzed and evaluated during the development of the videogame. The applied methodology of this research analyzes all factors in the context of usability.

B. Videogames and learning

The science of education apart from developing various theories has also attempted to isolate methods through which the learning procedure results can be improved. A number of these methods apply to videogames [7]. Although games had received an initial criticism focusing on their negative effects (addiction, unsocial behavior, violence), contemporary research has revealed far more positive aspects of games, such as their educational value [8], their aid in socializing [9], development of favorable attitudes to social issues such as the conservation of the environment and improvements in the classroom climate and interrelationships [10] or in enhancing logical thinking and decision making [11].

Regarding the group between the ages 9-13 which was the target group of the presented case, games can be influential in both the emotional and intellectual development of adolescents. Surprisingly, he goes so far as to assert that the game players usually have a higher intellectual level than non-players belonging to the same peer group. There are studies [12] which from the past decade foresee that games of social simulation (like the Sims and the Second Life are today) and of historical simulation and strategy will be accepted and used in the future as important knowledge tools. It is worth mentioning that recent studies [13], [14] have associated the knowledge a player can acquire by a game with its rank of usability.

From all the aforementioned references, it can be deduced that, for learning, videogames are of unquestionable importance, and can be used, in fact have already begun to be used, at different academic levels. In addition to stimulating motivation, video games are considered very useful in acquiring practical skills, as well as increasing perception and developing skills in problem-solving, strategy assessment, media and tools organization and obtaining intelligent answers. Of all the games available, simulators stand out for their enormous educational potential. In addition, researchers advocate use of videogames in the classroom to expose the values they (videogames) disseminate; that is to say, to develop a critical attitude towards certain undesirable behaviors (e.g., violence) or values (e.g., sexism). Research has yet to prove that video games are intellectually harmful; on the contrary, many studies defend their great importance in the development of intellectual abilities.

The user often tests new ways of handling the obstacles that emerge and has to try new techniques or invent tricks to get past them. In a similar manner, the theory of self-efficacy (i.e. if you believe, you will manage to succeed; you will try much harder) is related to games, since all games have points, levels, items that the user has to collect and, therefore, is motivated to continue playing and trying constantly for a better result. The theory of learning through a specific target (i.e. you learn better if you work on a well set target) is related to a specified objective that all games have. Additionally, for the theory of cooperation based learning, studies in classrooms have proved that team work improves significantly the learning process results. There has been an improvement of 50% in comparison to individual learning process [7].

Furthermore, apart from the knowledge that can be acquired through playing, it is possible to outline a series of procedural objectives that videogames can help to fulfill. Among them, it is important to emphasize the following: Reading (some games promote book reading, such as the Lord of the Rings), logical thinking, observation, spatiality and geography (during games such as Civilization IV that involve spatial representations), problem-solving and decision-making, strategic planning.

The time we spend on the game is also a very important factor that enhances the learning process. It is evident that we are able to learn a subject much better if we spend more time studying it. Statistic analysis proves that an average American gamer spends an average of 6.8 hours per week on videogames (http://www.theesa.com). Respectively in the 2004 American Freshman: National Norms Survey by UCLA's Higher Research Institute shows that pupils in secondary education spend five to eight hours per week to do their homework. If pupils had access on their daily homework through educational games then they would double the time on their homework [15].

III. The Setting, The Infrastructure, The Methods and The Participants

A. ISO9126-based evaluation

Based on the project's quality plan [16] the software quality evaluation was based on the ISO9126 standard [17]. The ISO9126 describes the abstract term 'quality' using six factors: *functionality*, *reliability*, *usability*, *efficiency*, *maintainability* and *portability*. The factors are associated to a

number of criteria in a hierarchical manner and finally to a set of metrics.

Since maintainability and portability were assured due to the use of the OpenSimulator platform emphasis was placed on the four ISO9126 factors that affect directly the users, i.e. functionality, reliability, usability and efficiency. The OpenSimulator is a 3D Application Server, able to host massive multiplayer on-line 3D environments. This platform bears many similarities to the popular Second Life environment, since it is a project inspired by Second Life (in an attempt to inverse-engineer and deploy its functionality under open-source terms). Despite being at an early stage of development (hasn't reached Beta version yet), the platform proves to be quite stable and robust, even when serving many concurrent users [18], therefore maintainability and portability issues were resolved due to the opening to the open-source community.

B. The software quality laboratory and the evaluation methods

Experimental usability evaluation methods take place in laboratories or in users' locations using mobile equipment and tools. Such a laboratory (Software Quality Laboratory) has been developed and used by the Software Quality Research Group of Hellenic Open University and used in this evaluation case. A schematic of the laboratory is shown in figure 1.

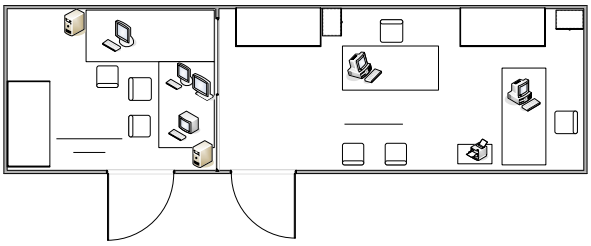

Fig. 1. The Software Quality Laboratory of Hellenic Open University.

The Software Quality Laboratory consists of one room designated as the *testing room* (on the right in figure 1) and a second room designated as an *observation and control room* (on the left in figure 1). The observation room is separated by the testing room by a *one-way mirror* so that the members of the quality evaluation team can watch the game players whilst the players cannot see them. The maximum number of individuals inside the testing room is two children (accompanied by their parents) and –in some cases– a member of the assessment team, depending on the case of the evaluation. All other members of the quality evaluation team are seated inside the observation room viewing the proceedings.

The equipment installed in this laboratory consists of one roof-mounted video camera that records the player's behavior, such as hand movements, facial expression and general body language throughout the experiment. A microphone is also placed near where the players are positioned to record their utterances. Another microphone is installed in the observation room to record evaluators' comments on the behavior of the player. Video from the camera as well as the real time image of the participant's monitor are both directed to video monitors into the observation room where they are recorded for further evaluation. Since the members of the quality evaluation team are not in the same room with the players this eliminates almost entirely any possible biasing effects due to inadvertent non-verbal communications or mannerisms. The test room is structured in such a way that gives the player the feel of a normal office room and not a laboratory. Finally the camera (roof-mounted) and the microphone are placed in such a way that are not easily observable, despite the fact that the player is informed of their existence before the beginning of the experiment.

For the experiments conducted with the participation of children and parents, special recording tools, such as *Usability Logger* [19] developed by Software Quality Research Group, were used in order to record all players' actions. These tools combine recordings from player's camera and microphone (required in order to participate in the survey) as well as recordings of all users' actions in the screen (screen captures, mouse movements, keystrokes, and idle time). The use of such tools was essential since some games within the SimSafety environment required the participation of 5 children, therefore some children played the game outside of the laboratory (but within Hellenic Open University premises, in a room nearby and under the supervision of one of the evaluators).

Regarding the methods used some limitation enforced by the fact that the users were children of ages among 9 to 13 years. Therefore evaluation methods applicable to adults, such as the Thinking Aloud Protocol, or Co-discovery could not be used. The methods used were: *Actions Logging*, *Experts Observation* and a final *Interview* including children and parents.

Fig. 2. View from the recordings inside the observation room.

Actions Logging (or User Logging) is an evaluation method that includes recording of all user's activities by the use of special equipment (which in this case was 2 cameras, 2 microphones and the aforementioned specialized logging software). Experts Observation added the evaluators'

comments to the data collected by various sources (in this case the experts data were recorded as verbal notes and combined to the roof camera recordings). All these data were digitally archived, synchronized and analyzed using Observer XT. Finally the analysis results were compared to the findings from the interviews.

Figure 2 shows the real-time recordings inside the observation room, while the evaluator is commenting. The screen on the left shows the video from the roof camera which is combined with the audio recordings of the evaluator's comments. The screen on the right shows the video captured from the PC in which the child is playing the game, combined with the audio captured inside the test room (in this case the dialogue between the child and his mother, also shown in the left screen).

C. *The participants and the data gathered*

Six children and five parents (one parent was with two children) had participated in the evaluation experiment. In order to set up the environment in 6 PCs and offer technical support to the participants, three technicians (members of RACTI, the partner that developed the environment) were present during the evaluation. Finally four evaluators participated in the evaluation. Two evaluators were sited inside the observation room; another one was inside the room where the rest of the children played the game; another evaluator was available outside the test room to aid in case some advice regarding the evaluation was required. In the evaluation also participated one member of the HOU team responsible for capturing the evaluation process using a digital photo camera (producing the photos presented in this paper).

Fig. 3. Preparations before the evaluation.

Figure 3, shows the final preparations before initiating the evaluation process. Both doors of the observation room and the test room were open just to capture the moment (the doors are closed during the evaluation). In the left side of figure 3 two evaluators are setting up the recording software and devices, while in the right side a parent and a child are waiting to start playing the videogame. It must be noted that six users (in this case children and parents) are considered sufficient in order to participate in such experiment, but in this case some videogames within the SimSafety environment require a minimum number of five participants, therefore it was essential to increase the number of participants above the numbers required for the aforementioned evaluation methods.

A number of particularities related to this evaluation procedure were recorded. Usually switching users from the other room into the test room is easy when it involves adults, but in some cases the children did not want to stop the game! On the other hand, adults sometimes feel peculiar inside the test room, once they are informed that they are monitored and all their actions are recorded. Children of course forgot all about the observation and the evaluation and focused directly on the game.

The evaluation procedure lasted just over 3½ hours of actual evaluation recordings (the overall process lasted quite more including the breaks, the initial orientation, the discussion and switching –sometimes unwilling users– in and out of the test room). The overall time to analyze the archived data was approximately 18 hours.

IV. The Evaluation Results and the Changes in the Game Environment

The analysis of the recorded data of the evaluation of this initial (alpha version) of SimSafety aided to identify a number of problems. Most problems were under the category usability but this was expected and it was in accordance to the literature review of section II. Table I summarizes the major problems that were detected; the symbols used are *F* for functionality, *E* for efficiency, *R* for reliability and *U* for usability.

TABLE I
Major Problems Reported During The Evaluation

Category	Detection method	Short Description
F	Observation	The goal of the game is not always clear to the children.
F	Observation and interviews	Some videogames perceived as quite 'adult' from the children.
E	Data logging	Delays in the avatar during walking through the SimSafety areas.
R	Observation and data logging	System crashed.
U	Data logging	Children had more fun exploring the environment, rather than playing the videogames.
U	Observation and interviews	After exploring the area and playing all the games there is nothing else to do.
U	Data logging	The lack of a visible scoring system reduces competition and fun.
U	Interviews	Children asked for more items to use in the SimSafety area.
U	Data logging and interviews	The stating requirements of some videogames are not clear to the children.

Apart from these problems some minor problems were also reported that either are not significant and are immediately solved on the following version, or they are depended on the

limitations of the 3D environment used. The most typical problem of the second category was the lack of avatars that look like children, since it was not supported in this version.

Regarding the functionality related problems, the fact that the goal of the game was not always clear to the children it was addressed by adding detailed instructions in certain areas and by creating an introductory place, which was the starting place for all first-time visitors. It is also true that some games (especially role-playing games inside the environment) require detailed instructions and still cannot be played effectively by children. On the other hand, these games were kept, since they are very successful when one of the roles is being played by an adult making provocative questions and initiating interesting discussions, as it was proven in evaluations and pilot usage after this one.

Both problems related to efficiency and reliability were related totally to the alpha version of the platform used and solved entirely in the following versions (even from the beta version).

Regarding the usability issues, it is true that children when joining the environment had more fun exploring the environment, rather than playing the videogames. Data logging and observation showed that they prefer exploring, walking or flying around, teleporting and not playing the videogames immediately. It is also true (and shown in data logging and mentioned in the interviews) that after exploring the area and playing all the games there is nothing else to do. The solution to both these problems was to introduce items related to internet safety (such as a piece of paper with a login and a password on it) that children could use freely (i.e. not related to a specific videogame) while exploring the environment. The use of a 'Report Centre' where a game player can declare a lost and found object or report that he has been bullied aided towards learning while exploring and interacting. Furthermore, the use of bots (virtual people that interact with the children) added more fun to the game and reduced this problem. Clear evidence that the problem was solved is the fact that children that played the game once, they kept logging into the environment.

The problem of scoring was also solved, adding points into any 'proper' action related to Internet Safety and, as discussed previously, more items were added into the environment. Of course not all the items asked by the children (such as bikes, cars, etc.) but only the ones related to the goal of the game, still with some exceptions in order to increase the essential fun element of the game. Finally, regarding the 'visibility of game status' and the starting requirements of each game this was resolved by adding instructions and mechanisms to inform the children about the number of players registered in each game and the number of players that are required to start playing.

V. Conclusion and Future Goals

Evaluation is a dynamic process. The one presented in this paper was the first, preformed in a suitable setting and adding to identify and solve problems in the alpha version. Further evaluations and pilot runs in schools proved that all the reported problems were addressed and the results regarding children satisfaction were encouraging and promising in relation to game sustainability.

Acknowledgment

The authors would like to thank the children and parents participated in the SimSafety Evaluation Experiments, as well as all the consortium members (technicians and researchers) that aided in the evaluation experiments.

References

[1] M. Federoff, Heuristics and Usability Guidelines for the Creation and Evaluation of Fun in Video Games, Thesis at the University Graduate School of Indiana University, 2002.

[2] D. Pinelle, N. Wong, T. Stach, *Heuristic Evaluation for Games: Usability Principles for Video Game Design*, Proceedings, Game Zone on CHI 2008 April 5-10, 2008 Florence, Italy.

[3] H. Desurvire, M. Caplan, J.A. Toth, *Using Heuristics to Evaluate the Playability of Games*, in Proc. Conference on Human Factors in Computing Systems, New York: ACM Press, 2004.

[4] ISO 9241 Part 1-17: *Ergonomic Requirements for Office Work with visual display terminals*, Standard 2008.

[5] R.J. Pagulayan, K. Steury, B. Fulton, R.L. Romero, Designing for fun: User-testing case studies, In *Funology: From Usability to Enjoyment*, M.A. Blythe, K. Overbeeke, A.F. Monk, P.C. Wright, (Eds.), Kluwer Academic Publishers, Dordrecht, Boston, London, 2002, pp. 137–150.

[6] M.C. Medlock, D. Wixon, M. Terrano, R.L. Romero, B. Fulton, Using the RITE method to improve products; a definition and a case study, *Proc. of the Usability Professionals Association Conference*, B., 2002.

[7] D. Johnson, J. Maruyama, R. Johnson, M. Nelson, A. Scon, Effects of Cooperative, Competitive and Individualistic Goal structures on Achievement: A meta-analysis, *Phychological Bulletin*, Vol. 89, 1981, pp. 47-62.

[8] M. Aguilera, A. Mendiz, Videogames and Education, *ACM Computers in Entertainment*, Vol. 1, No. 1, October 2003.

[9] D. Garvey, "Simulation: A Catalogue of Judgements, Findings, and Hunches", in P.Tansey, ed, *Educational Aspects of Simulation*, London, McGraw-Hill, 1971, pp. 204-227.

[10] W. R. Heitzman, *Educational games and simulations*, Washington: National Education Association, 1974.

[11] H. Ellington, F. Percival, *Educating 'through' science using multi-disciplinary simulation games: Programmed Learning and Educational Technology*, 1977, pp. 117–126.

[12] E. E. Gabriel, Que faire avec les jeux video?, *Hachette Education*, Paris., 1994, pp. 160.

[13] S. Papaloukas, M. Xenos, *Usability and Education of Games through Combined Assessment Methods*, 1st ACM International Conference on PErvasive Technologies Related to Assistive Environments, PETRA-2008, Athens, Greece, July 15-18.

[14] M. Xenos, S. Papaloukas, N. Kostaras, *Games' Usability and Learning – The Civilization IV Paradigm*, IADIS GET2009, Game and Entertainment Technologies 2009, Algarve, Portugal, 17 - 19 June 2009, pp. 3-10.

[15] J. M. Mayo, For Science and Engineering Education, *Communications of ACM,* Games, Vol. 50, No. 7, 2007, pp. 31-35.

[16] D5.1 *Quality Management Plan*, Deliverable of the SimSafety Project, Available at: http://www.simsafety.eu, May 2009.

[17] ISO/IEC 9126, *Software Product Evaluation – Quality Characteristics and Guidelines for the User*, International Organization for Standardization, Geneva, 2001.

[18] D. Kalaitzis, E. Valeontis, V. Delis, M. Fountana, Experiences from Developing Online VR Environments: The 'SimSafety' Case Study, *Proc. of Social Applications for Lifelong Learning*, Patra, Greece, 4-5 November 2010.

[19] N. Kostaras, M. Xenos, C. Koutsogiannis, Usability Evaluation of Software Applications with the Use of Usability Logger, *Proc. of the 8th Joint Conference on Knowledge-Based Software Engineering*, JCKBSE 08. Pireaus, Greece, 2008, pp. 395-404.