GAMES’ USABILITY AND LEARNING – THE CIVILIZATION IV PARADIGM

Michalis Xenos*, Spyros Papaloukas, Nektarios Kostaras
Hellenic Open University, School of Sciences & Technology, Computer Science Dept.
23 Saxtouri Str., Patras, Greece, GR-26222

ABSTRACT
This paper presents the use of a well recognized and accredited game (i.e. Civilization IV) as a paradigm in order to measure how usability of a game is related to knowledge acquisition during game play. Using a specialized laboratory equipped with monitoring devices and tools we measured the perceived game usability of 50 users with various levels of gaming experience. These results compared with knowledge acquisition during a determined game period have revealed that knowledge acquisition is highly depended on game usability. The paper briefly presents the methods used for usability evaluation and knowledge acquisition and the usability assessment laboratory. It also presents the data from the sessions involving 50 users and the discussion regarding games’ educational value in relation to their usability, as well our final conclusions.

KEYWORDS
Software Games, Game Usability, Educational Games, Collaborative Games, Usability Laboratory, Usability Assessment Methods.

1. INTRODUCTION
Probably every computer user has played a computer game at least once, while nowadays computer games (games from hereinafter) are part of everyday life of many computer users and especially young people. Millions of young people –and many not so young– spend hours daily involved in games, therefore games play a vital part in contemporary culture.

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 (Aguilera & Mendiz, 2003), their aid in socializing (Garvey, 1971), development of favourable attitudes to social issues such as the conservation of the environment and improvements in the classroom climate and interrelationships (Heitzman, 1974) or in enhancing logical thinking and decision making (Ellington, 1977). According to Estallo (1994), games can be instrumental 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 (Gabriel, 1994) 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 (like Civilization) will be accepted and used in the future as important knowledge tools. Nowadays, scientists accepting the contribution of a game to the learning process, are developing models like GAeL-Game Accelerated eLearning (Fladen & Blashki, 2005), that utilizes game principles in a learning environment.

In this paper we use Civilization IV (http://www.civiv.com) as a paradigm in order to measure the degree of knowledge acquisition during game play. Civilization IV was selected primarily since it is not an educational game, therefore the players do not aim to acquire knowledge, but while playing they receive historical and geographical information. Our aim was to measure how players retain this information they

* Tel. +30 2610 367405, Fax. +30 2610 367416, E-mail: xenos@eap.gr
receive while playing the game and the correlation of knowledge acquisition with perceived game usability. In other words, how game usability affects knowledge acquisition during game-play.

Usability of a game is defined as the degree that a player can learn how to play and understand a game (Federoff, 2002; Desurvire, 2004). 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. Software Quality Research Group (SQRG, 2009) has developed a specialized laboratory as well as methods and tools used to measure software usability, some of which were used in this survey. Players’ involvement included 10 players participating in the game using the laboratory premises (recorded data of over 20 hours from multiple sources), 15 players playing the game in their own places (using remote logging and recording tools) and 25 players in typical software laboratories (using similar recording and logging tools).

The rest of this paper is organized as follows: section 2 is a brief literature review on games, education and usability evaluation methods. Section 3 presents the Civilization IV game, the software usability laboratory and the methods and tools used to measure perceived usability and knowledge acquisition. Section 4 discusses results from the various experiments involving 50 users, while section 5 summarizes the conclusions of this paper.

2. GAMES, EDUCATION AND USABILITY

During a game the user’s brain produces dopamine (Koepp et al, 1998). The dopaminergic transmission can affect the learning process and concentration; enhance behaviour and is critical for sensorimotor functions in the striatum (Robbins & Everitt, 1992; Schultz et al, 1993). Apart of the medical indication of improved learning capabilities during a game, educational theories (Johnson et al, 1981) are highly related to games. More precisely the theory of experiential learning (i.e. if you do it you learn it) is one of the key elements of a game providing that the users have to navigate themselves in the game’s world, follow a game scenario but also make crucial decisions which affect the outcome of the game. In many cases they are made to repeat the same action several times in order to succeed. Also the theory of inquiry-based learning (i.e. what will happen if I do that?) is relevant to games as they give the user the chance of unlimited navigation round the game’s world. 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 (Johnson et al, 1981).

Furthermore, apart from the knowledge that can be acquired through playing, it is possible to outline a series of procedural objectives that video games can help to fulfill (Aguilera & Mendiz, 2003). 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 two latest being important in strategy games such as Civilization IV).

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 (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 (Mayo, 2007).

Studies on games have numerous extensions not all of which are relevant to the present research that focuses on measuring the usability of videogames. 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 we refer to ISO 9241 standard of usability (ISO 9241, 2005) in which the three basic usability metrics are described: effectiveness, efficiency and satisfaction.
The effectiveness is linked with accuracy and total achievement of the goal set by the user, while efficiency relates to the resources spent for the attainment of the objectives and satisfaction affiliates with the users’ mood. Frokjaer et al (2000) believe that these three factors have to be considered as separate, independent usability factors. The games like all other software types have an interface that should provide effective and efficient means of interaction between the user and the game; but of course when studying the playability of a game, which is an indispensable part of usability, it is obvious that the three usability metrics do not have the same influence on the game. Clanton (1998) suggests a way of combining various usability forms into three areas that are relevant to the structure of the game; these are 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 game.

The development of heuristics in order to evaluate videogame’s usability has been a very popular among game manufacturers and begun with Malone’s (1980) influential heuristics for designing instructional video games. Since then several researchers have attempted to identify the nature of successful games in the form of heuristics; a great number of these attempts, however, failed to produce a solid one like Malone’s, who developed specific game prototypes using different elements of his theory and tested them with children (Malone, 1982). Instead, recent heuristics tend to be based on designer experience (Rouse, 2001), interviews with members of the game industry (Federoff, 2002), or simply drawn from existing literature (Desurvire et al., 2004). In most cases, the heuristics are not actually applied or tested in any way, but are offered as a kind of instinctual knowledge ultimately derived from the game industry (Barr et al., 2007).

3. CASE DESCRIPTION

As aforementioned, this study involved 50 users that all played Civilization IV. From these users 10 played the game monitored on the software usability laboratory, while 15 using their own computers in their own places and 25 on computer laboratories. In this section we briefly present Civilization IV and the methods, tools and equipment used to record all these users’ activities.

![Figure 1. Information about a new monument in Civilization IV](image)

3.1 Civilization IV

Civilization is a game created by Sid Meier for MicroProse on 1991. Current version of the game is Civilization IV, which can be played as single player game or online multiplayer game. Civilization IV has been awarded (http://www.2kgames.com/civ4/awards.htm) by various organizations such as the Academy of Interactive Arts & Sciences, Gamespot, IGN, etc. This game was ideal for this study, since it is not an educational game. In an educational game the player’s goal is self evident: the acquisition of knowledge. In this case the goal was to measure the information retained by a player during a game, that its sole purpose is not educational. Furthermore, the facts presented in this game seem real, but they are not real. Some of the city names are real, but most of the dates and locations are determined by each particular game plot and they are not real. Therefore, the measurements of knowledge acquisition after the game could not be biased by
previous player’s knowledge. This was extremely important for the integrity of our data: measurements after the game correspond to information received by each player during this game and this game only.

Each game in Civilization IV starts at age 4000 BC and players attempt to create and maintain an empire up to contemporary dates. Each player starts the game by selecting a civilization (i.e. Greek civilization) and controlling two units and competes against other civilizations (controlled by other players or the computer). Each player acts in turns and has to make decisions related to politics, religion, diplomacy, trading and war. During the game the players receive historical and geographical information (like the birth of historical people in particular cities and dates, the development of new cities, the creation of historical monuments, etc.). An example of such information is shown in figure 1 and it must be noted that this information is not historically and geographically correct, since it is depended on each game’s progress and civilizations that participate in the game. For example Parthenon could actually be built in Athens in the 5th century BC, but could also be built in Rome in the 10th century BC!

3.2 Software Quality Laboratory

Experimental usability evaluation methods take place in laboratories or in users’ locations using mobile equipment and tools. Such a laboratory (Software Quality Assessment laboratory) has been developed and used by the SQRG team. Although this laboratory has equipment and tools for various aspects and not only usability evaluation hereinafter we will focus exclusively in the usability features. A schematic of the laboratory is shown in figure 2.

The Software Quality Assessment Laboratory of SQRG consists of one room designated as the testing room and a second room designated as an observation and control room. The observation room is separated by the testing room by a one-way mirror so that the members of the quality assessment team can watch the game players whilst the players cannot see them. The maximum number of individuals inside the testing room is two users and –in some cases– a member of the assessment team, depending on the case of study. All other members of the quality assessment 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 assessment 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 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 in actual players’ locations or remote computer laboratories, special recording tools, such as Usability Logger (Kostaras et al, 2008) developed by SQRG, 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).

3.3 Research Methods and Tools

For measuring perceived usability for each player, different methods were used depending on the player’s location, as outlined in table 1. Action logging took place in all cases using the laboratory equipment in the first case and the Usability Logger combined with input from the user’s microphone and camera in the other two cases. Following each experiment the individual knowledge acquisition for each user, was measured using a questionnaire (in all cases) and users’ interviews in the first and third case.

<table>
<thead>
<tr>
<th>Location</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software quality laboratory</td>
<td>Co-Discovery &amp; Actions Logging</td>
</tr>
<tr>
<td>Player’s location</td>
<td>Thinking Aloud Protocol &amp; Actions Logging</td>
</tr>
<tr>
<td>Computers laboratory</td>
<td>Questions-Answers Protocol &amp; Actions logging</td>
</tr>
</tbody>
</table>

Generally speaking experimental methods (such as co-discovery, thinking aloud protocol and questions-answers protocol) involve the observation of individual users performing specific tasks with the game under evaluation. In these methods the users perform representative tasks, under the discrete attendance of usability experts (the personnel supporting the experiment). Papaloukas and Xenos (2009) have already used combined experimental methods for the measurement of usability in the game Second Life in usability experiments conducted in the HOU’s quality laboratory. The results have shown that the combination of methods amplifies the progress of the experimental procedure, providing that the conducting conditions simulate reality adequately.

**Actions Logging** (or User Logging) includes recording of all user’s activities by the use of special equipment like cameras, microphones and specialized logging software. **Thinking Aloud Protocol** is the method in which the users express verbally their thoughts, feelings and opinions while interacting with the system. All these thoughts and feelings are recorded and compared with user’s actions. **Co-Discovery** is the method in which the users attempt to perform tasks together while being observed. The difference of this method over the thinking aloud protocol is two-fold: most people have someone else available for help and the interaction between the two participants can bring out more insights than a single participant voicing his or her thoughts. Moreover this method has a better effect on the learning process and especially in cooperative learning. Finally, **Questions-Answers Protocol** is a method in which the evaluator provokes the player to express thoughts and feelings while playing by asking the player specific questions. It is similar to thinking aloud protocol, but the player is not expected to be focused on the game and commenting at the same time.

The Inquiry methods we have used are questionnaires before and after the conduct of the experiment, in addition to a personal interview of the users who were asked open-ended and closed-ended questions immediately after playing the game. Inquiry methods provided us initially (before the game) with subsidiary characteristics of the users such as their computer familiarity, the time they usually spend on videogames and the Internet. After the game the focus of the inquiry was to determine the knowledge each player gained while playing the game. All the aforementioned data, collected from various devices in all three locations was inserted into Observer XT (http://www.noldus.com). The Observer XT allows to synchronize all data for each player, to identify actions (point events and state events), to comment player actions and to code player behaviors. Using the Observer XT we have coded and monitored all players’ activities throughout playing time. It must be pointed out that, although the synchronization of the data and the identification of events are automated, the analysis of data is a time-consuming process that requires at least 3 times the amount of time actually spent by each player during the game.
4. EXPERIMENTS RESEARCH RESULTS

Although the initial population of people registered to participate in our experiments was higher than 50, for this survey we selected only people that had never played the Civilization IV game before. This is important in order to be able to identify reactions that indicate perceived usability during the first hours they are playing the game. Thus the final number of participants was restricted to 50 people. Before starting to play the game, they were asked to fill in a questionnaire that offered demographic data to us such as gender, age, hours spent on the computer, hours spent on the Internet, hours spent playing games, experience in similar games, game preferences (single vs. multiplayer), etc. An outline of the data is shown in Table 2.

Afterwards the actual experiments took place. The experiment involving 10 users in the software quality laboratory took two days to complete and during that time data was collected from 12 users that played the game in their own computers. After the completion of these two experiments five more experiments took place in five separate school computer laboratories collecting data from 28 players (20 students and 8 adults) that had played Civilization IV for the first time. The data from all these experiments was introduced into Observer XT and analyzed.

<table>
<thead>
<tr>
<th>Table 2. Players’ demographic data</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

The Observer XT allows the evaluator to categorize all user data, comment them and assign a value to similar cases in order to affect the overall perceived usability for each user. The overall perceived usability for each user is calculated as a number from 0 to 1. The data that contributes to this degree of user perceived usability are introduced from concurrent analysis of video (screen capture and player face), audio, and logging and includes gestures, facial expressions, twitches, indication of discomfort, indication of pleasure or excitement, direct comments of appreciation or disapproval, groaning, discussion among players, expression of fun or enjoyment, time consumed for particular tasks, hesitation logged as idle time, wrong movements (not in terms of gaming but in terms of usability), nervous mouse movements, the game score, etc. The analysis of the video recordings based to the DEtailed Video ANalysis method (DEVAN); a method that was developed to detect usability problems in task-based products for adults (Vermeeren et al., 2002) and has been adjusted for the detection of usability problems in a strategy game.

<table>
<thead>
<tr>
<th>Table 3. Players’ usability range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adults’ Average</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
</tbody>
</table>

Based on data from all 50 players the degree of perceived usability range from 0.37 to 0.83 (average 0.61, standard deviation 0.11), while minor differences were identified among various categories. The results from the inquiry methods were compared with the actual information from the game video capture and graded ranging from 0 to 1. The highest score represents correct responses in all questions (therefore a player that retained all information will receive a high score, whilst a player that retained less information will receive a lower score). Based on data from all 50 players the degree of knowledge acquisition range from 0.30 to 0.90 (average 0.68, standard deviation 0.14) while minor differences were identified among various categories.

<table>
<thead>
<tr>
<th>Table 4. Players’ acquisition range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adults’ Average</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
</tbody>
</table>
Table 5. Players’ correlation results

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Adults</th>
<th>Teenagers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>0.72</td>
<td>0.67</td>
<td>0.78</td>
</tr>
<tr>
<td>Female</td>
<td>0.74</td>
<td>0.79</td>
<td>0.68</td>
</tr>
<tr>
<td>Total</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Comparison of the data from perceived usability degree and knowledge acquisition shows that these values are correlated. Table 5 presents the results from the CORREL function between usability range knowledge acquisition range for the 50 players, while figure 3 illustrates the collective results of usability degree and knowledge acquisition degree. Even from observing the figure 3, it is obvious that these values are correlated significantly.

![Figure 3. Collective results for all players: Usability degree (black) and knowledge acquisition (white)](image)

5. CONCLUSIONS

Several studies, some of them referenced in the literature review, have proven that there is an unquestionable relationship between gaming and education. This relationship has been the object of research and studies since seventies and the scientific community seems convinced that during game playing learning takes place. In this paper non-biased players (in terms that they have never played the particular game before) had been selected. These players were not experienced in Civilization (any version of it), but were relatively experienced in gaming. Focusing on these players and utilizing the specialized measurements equipment, we have used diverse methods in order to determine perceived usability degree for each player and the correlation between usability and acquisition of knowledge. What is evident from the results presented in this paper is that perceived usability of a game affects in a significant degree the learning that took place during play. Game engineers’ aim to develop educational games should focus not only on the educational value of the game, but in the usability per se.

ACKNOWLEDGEMENTS

The authors would like to thank the 50 users participating to the experiments presented in this paper. Part of the research presented in this paper is funded by the EC Project: SimSafety: Flight Simulator for Internet Safety, Life Long Learning Program, Project: 143689-LLP-1-2008-GR-KA-KA3MP. This project is about the development of an educational game teaching internet safety to pupils (age 9-12).

REFERENCES


Papaloukas, S., Xenos, M., Enhanced socializing through the usability of a videogame: a case study on Second Life, Submitted In Journal of Interacting with Computers.


