On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE

Maria Virvou [*], George Katsionis
mvirvou@unipi.gr; gkatsion@kman.gr

Department Of Informatics, University of Piraeus, 80, Karaoli and Dimitriou St., Piraeus 185 34, Greece

Abstract

Educational software games aim at increasing the students' motivation and engagement while they learn. However, if software games are targeted to school classrooms they have to be usable and likeable by all students. Usability of virtual reality games may be a problem because these games tend to have complex user interfaces so that they are more attractive. Moreover, if the games acquire an educational content they may lose the attractiveness and appeal that they have on users who are familiar with commercial games. Consequently, likeability may also be questioned. In this paper, we address the issue of usability and likeability of a virtual reality game that is meant to teach students geography. We describe the evaluation experiments conducted, which involved three categories of students in terms of their level of game-playing expertise: novice, intermediate and expert game players. The evaluation results showed that the game was indeed usable and likeable but there was scope for usability and likeability improvement so that the educational benefits may be maximised for all categories of students. The evaluation studies reported in this paper, revealed important issues about further research on virtual reality educational games.

Keywords: evaluation of CAL systems, learning environments, teaching/learning strategies, virtual reality, educational games, edutainment

1. Introduction

In the recent years there is an increasing popularity of software games among children and adolescents. Many empirical studies confirm that software games constitute the principal way that children and adolescents become acquainted with the use of computers and devote a lot of their time on them (Mumtaz 2001). This has resulted into a rapidly expanding industry of commercial games. However, these games are mainly created for entertainment and do not aim at educating the targeted users. On the other hand, educational technologists seek ways to render software more attractive in their effort to create software that will motivate and engage student-users in learning. In this respect, the popularity of games may be exploited for the purposes of education through the creation of educational software games. Moreover, the gaming environments can provide the means for a constructivist approach to learning. Constructivist approaches (Papert, 1980) suggest that children should acquire knowledge through experience.

Indeed, there have already been quite a lot of research projects towards the development of software games for education that aim at increasing the students' motivation and engagement while they learn (e.g. Amory et al. 1998; Conati & Zhou 2002; Kearney 2004). Being actively engaged in a learning activity has repeatedly been shown to be beneficial for learning; engagement with a learning environment facilitates exploration and conversely exploration promotes engagement (Price et al. 2003). The resulting hybrid applications that combine education with entertainment are often referred to as edutainment. However, none of these research projects has been reported to have been introduced in real school classroom settings. In any case, before educational software games are introduced in classrooms there are several issues that have to be examined very carefully and the games have to be designed or redesigned accordingly.

More specifically, if games are to be introduced in classrooms they have to be usable and likeable by the majority of students. Otherwise, there is no point employing gaming aspects in education. Despite the great popularity of software games there are still many children and adolescents who are not familiar with them. For example Griffiths and Hunt (1995) in a study they conducted found that approximately one third of the adolescents play computer games every day whereas the same proportion play once a month or less. This shows among other things, that there is a difference at the level of expertise and competence on software game playing among children and adolescents of the

---

[*] Corresponding Author: (Maria Virvou, Associate Professor), Tel: +301-4142269, Fax: +301-4112463, E-mail Address: mvirvou@unipi.gr

same age. Children and adolescents who play games every day must have acquired considerable skills on game playing whereas the rest of the children have not. This may also imply that there is a difference in the likeability of these games by their potential users. The fact that there are children and adolescents who do not play games may mean that they have not had the chance to play more often but it may also mean that they do not like games very much. Likeability can be affected by usability: the less usable a game, the less likeable it may be. For example, if some children and adolescents find it difficult to interact with games then perhaps they do not like these games so much as they could. Hence, if computer games are introduced in classrooms and are included in the compulsory assignments and tests that are given to students then these games might cause problems to some students instead of help them in their learning process. As pointed out by Markopoulos & Bekker (2003a), usability may be a crucial stumbling block for learning or enjoyment.

However, there are also other issues that may affect likeability of educational games. One such issue is the inevitable competition of educational games with commercial games. This competition may not be literal in terms of marketing but will certainly take place in the targeted users' minds and will affect their tastes, especially if these users have already had quite a lot of experience with commercial games. In this respect, users may not like the educational content of educational games or they may not like the educational games if they seem less attractive than commercial games. Indeed, there are criticisms from researchers about the quality of the existing educational games. For example, Brody (1993) points out that the marriage of education and game-like entertainment has produced some not-very-educational games and some not very-entertaining learning activities. Such issues concerning the likeability and the resulting motivation of educational games have to be examined before educational games can be proposed to be included in classrooms.

In this paper, we describe the evaluation experiments that we have conducted so that the usability and likeability of an educational virtual reality game could be evaluated. The game is called VR-ENGAGE and is meant to teach children geography through a virtual reality adventure game (Virvou et al. 2002). Virtual reality games are among the most popular games for children and adolescents. Nevertheless, their complex 3-D interfaces constitute an issue to be examined in terms of usability and likeability.

Therefore, we have conducted the present evaluation that focuses on the usability and likeability aspects of the educational game. The primary aim of our evaluation experiments is to reveal the special conditions with respect to usability and likeability that may hold in an environment that is neither solely educational nor solely entertaining, but a combination of the two. The evaluation experiments are described and analysed in the subsequent sections. The results of this analysis provide useful findings about the scope of edutainment environments and lessons that can be used for the specification of design guidelines of such software applications.

2. Related work

Usually, evaluation of educational software mainly concerns the pure educational effects of the software. However, in most cases, usability is considered to be an important factor that affects the educational effectiveness (Jones et al., 1999; Mayes & Fowler, 1999; Squires & Preece, 1999). For example, the JIGSAW model (Squires & Preece 1996) or the 'set of learning with software heuristics' (Squires & Preece 1999), address the problem of integrating both usability and learning issues in the evaluation of educational software. In the JIGSAW model, the evaluation is performed in three levels. In Level 1, the subtasks of the learning and operational tasks are considered independently of each other. As we move from Level 1 to Level 2, integration within the learning and the operational tasks is considered. At Level 3, integration between the learning and operational tasks is considered. On the other hand, the set of 'learning with software' heuristics, are an adaptation of the "usability heuristics" presented in Nielsen (1994), so as to relate them to socio-constructivist criteria for learning.

Moreover, in addition to the importance of usability in educational software, some researchers have also acknowledged the fact that depending on the underlying rationale of the educational software, the evaluation methods should vary accordingly. The above views are explicitly supported in the CIAO! framework (Jones et al. 1999) which is mainly created for educational software. In particular, the CIAO! framework outlines three dimensions to evaluate: (i) context; (ii) interactions; and (iii) attitudes and outcomes. One important aspect of context is the reason why CAL is adopted in the first place, i.e. the underlying rationale for its development and use; different rationales require different evaluation approaches. According to the framework, the reason for looking at students' interactions with the software is in order to understand more about their learning processes; such interactions can provide protocol data for later analysis. Finally, at the "outcomes" stage, information from a variety of sources needs to be used including, pre and post-achievement tests, interviews and questionnaires with students and tutors.

In the case of a virtual reality educational game, it is very important to evaluate the usability of the typically complex virtual reality environment of the game since it is not our aim to make the students' learning process more difficult than it already is. Moreover, it is very important to evaluate the likeability of the game because if the game is not likeable then there is no point using a game environment for educational purposes.

2.1 Evaluation of educational software

In VR-ENGAGE, first we conducted an evaluation that focuses mainly on educational effects. This evaluation is fully described in (Virvou et al. 2005). That educational evaluation aimed at finding out whether the educational game environment achieved better educational effects in comparison with educational software of no gaming environment.

The evaluation of VR-ENGAGE with respect to educational effects was conducted using three groups of students, those who used to have good academic performance, those who used to have mediocre academic performance and those who used to have poor academic performance in the domain taught by VR-ENGAGE. The results from the evaluation showed that all of the students had benefited more from the educational game than the conventional educational software. Moreover, one important finding that came out was that when the subgroups of students who previously had good, average and poor performance respectively were compared separately, it was revealed that the subgroup of students who used to be poor performers had benefited the most from the game environment whereas the subgroup of good students had benefited the least from the game environment. This coincides with findings about the benefits of multimedia in general (Mayer, 2001).

However, it must be noted that during that experiment all of the students had as much help as they needed from lab instructors concerning their interaction with the VR environment of the game. If the students had used the software on their own at school or at home, then perhaps they might have had more usability problems, especially those who were not sufficiently experienced in virtual reality game playing. Such problems might have resulted in less good educational results. Moreover, such problems may have affected the likeability of the game and thus the motivational asset of the game may have been refuted. The likeability of the game may also have been affected by the fact that the game is educational and as such it is based on content that is pertinent to school syllabus.

Consequently, the evaluation with respect to the learning outcomes was not sufficient since it could not possibly reveal all the usability problems that students may have had while they interacted with VR-ENGAGE. Moreover, it could not possibly show the extent to which students have actually had a more enjoyable experience with the program than with other educational programs. Moreover, even if a first evaluation of the game has shown that the educational game is better in terms of the learning outcomes than a non-game educational software application, this does not necessarily mean that it could not have been even better if possible usability and likeability problems were resolved.

Therefore, issues concerning the usability and likeability of the educational VR-game remained to be investigated. These issues are addressed in the present evaluation which involves real users from schools. Unlike the educational effectiveness evaluation, this evaluation takes into account three groups of users with respect to their level of experience in playing virtual reality games rather than the quality of academic performance. The reason for this classification is that usability and likeability degrees may vary depending on the level of users' expertise.

2.2 Usability of educational games for young learners

As already mentioned, usability is a very important aspect of educational software games that has to be examined. Usability of software that is targeted to children has recently attracted a lot of research energy (e.g. Markopoulos & Bekker 2003b; Hoysmiemi et al., 2003), since children constitute a special group of users (Bruckman & Brandlow 2003). In the case of educational games, usability plays an even more important role than in the case of other educational software since the user interface of the games is typically more complex so that it may support amusement, narrative, adventure and so on. As pointed out in (Yacci et al. 2004), edutainment environments that include educational games, demand a certain amount of effort and learning that is not related to the instructional goals of the school lesson that is taught; such irrelevant learning concerns the plot or mission of the game and the "legal" movements and actions that a player can make while "inside" the game. Thus, they point out that there is a very important question associated with educational games: how much student effort is an expense when engaging in edutainment. More specifically, they identify three types of learning which are not necessarily usable and valuable outside the edutainment environment: 1) Operations, 2) Strategy, and 3) Instructional Goals and Outcomes. Operations refer to the "legal" movements and actions that a player can make inside the game. Strategy learning refers to the overall plot or mission of the game. Finally, instructional goals and outcomes refer to educational goals and outcomes that have value beyond the game itself. In the case of VR-ENGAGE the classification of usability characteristics has to take place in relation to the 3D virtual reality environment of the game, which adds complexity to the user interface on top of the operations, and strategies of game playing.

VR-ENGAGE usability evaluation involved user-based methods of evaluation. The testers in both parts were children. Sweeney, Maguire and Shackel (1999) suggest that the typical methods for user-based evaluation involve observation, computer-logging, video recording of user interactions and the user's self reporting which includes post-hoc comments and questionnaires. In the case of our evaluation, we did not use video cameras for recording user interactions so that students would not feel uncomfortably while working with the software programs. Indeed, Hoysniemi, Hamalainen, and Turkki (2003) argue that usability testing can be a very unnatural experience for children if they are taken to try out an unknown computer product while video cameras are recording. Hence, we used extensively computer-logging and self-reporting methods.

2.3 Likeability

There are studies that are concerned with the likeability of games, which are not necessarily educational (Fabricatore et al. 2002). Moreover, there are studies that relate heavy use of computer games with certain categories of children. For example, in a study conducted by Tsai (2004) it was reported that there were significant differences between boys and girls in motivation of online game playing. This may be the case for other categories of students other than the boy/girl classification. What is important is that not all children are reported to like computer games to the same extent and not all computer games are likeable by their targeted users.

Likeability and the resulting motivation for students is the principal asset of games as compared to other educational software. Motivation has been an important issue in educational software that has been examined in several ways, such as motivational diagnosis (De Vincente & Pain 2002) or instructional planning to increase motivation (Matsubara & Nagamashi, 1996). However, so far the likeability of games as a factor of motivation has not been examined but rather has been taken for granted in most cases. Despite this view, if the likeability of a gaming interface is not high for the majority of the targeted users, then there is no point having a game introduced in a classroom in the first place.

Our approach for evaluating likeability in VR-ENGAGE is quite similar to the one adopted for the Snark game (Price et al. 2003). In the Snark adventure game the children had to find and capture as much as they could about the imaginary creature called the Snark. Although, the Snark game is quite different from the virtual reality adventure game of VR-ENGAGE, the aim of the evaluation of likeability of the game environment in both cases is quite important. In the case of the Snark game, one important clue for the assessment of excitement and engagement of players was their desire to continue interacting and playing. Similarly, in VR-ENGAGE, the length of time of interaction of children with the game was considered a very important clue for the evaluation of likeability.

3. The educational VR game

VR-ENGAGE is a virtual reality game, which is highly interactive. The environment of VR-ENGAGE is similar to that of the popular game called "DOOM" (ID-Software 1993), which has many virtual theme worlds with castles and dragons that the player has to navigate through and achieve the goal of reaching the exit. Similarly with DOOM, VR-ENGAGE has also many virtual worlds where the student has to navigate through. There are mediaeval castles in foreign lands, castles under the water, corridors and passages through the fire, temples hiding secrets, dungeons and dragons. The main similarity of VR-ENGAGE with computer games like DOOM lies in their use of a 3D-engine.

The story of VR-ENGAGE incorporates a lot of elements from adventure games. The ultimate goal of a player is to navigate through a virtual world and find the book of wisdom, which is hidden. To achieve the ultimate goal, the player has to be able to go through all the passages of the virtual world that are guarded by dragons and to obtain a score of points, which is higher than a predefined threshold. The total score is the sum of the points that the player has obtained by answering questions. In particular, while the player is navigating through the virtual world, s/he finds closed doors, which are guarded by dragons as illustrated in the example of Figure 1. A guard dragon poses a question to the player from the domain of geography. If players give a correct answer then they receive full points for this question and the dragon allows them to continue their way through the door, which leads them closer to the "book of wisdom".
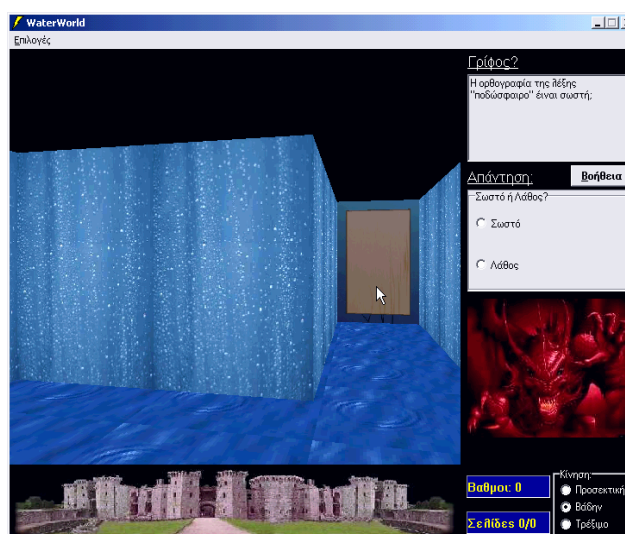

Figure 1: Virtual water world

As part of the adventure of the game the player may come across certain objects where s/he may click on. These objects appear at random and give hints to students or guide them to read another part of the domain being taught. For example, Figure 2 illustrates a hint in the form of a blue ball in a virtual volcano world. However, these hints or the parts of the theory that are visited, are not immediately usable by the students, since they refer to questions that the students will have to answer at a location of the virtual world other than the one that they are currently at. Hence, the students will have to remember these hints or parts of the theory so that they may use them when the time comes. Educationally, these hints motivate students to read and memorise important parts of the theory.
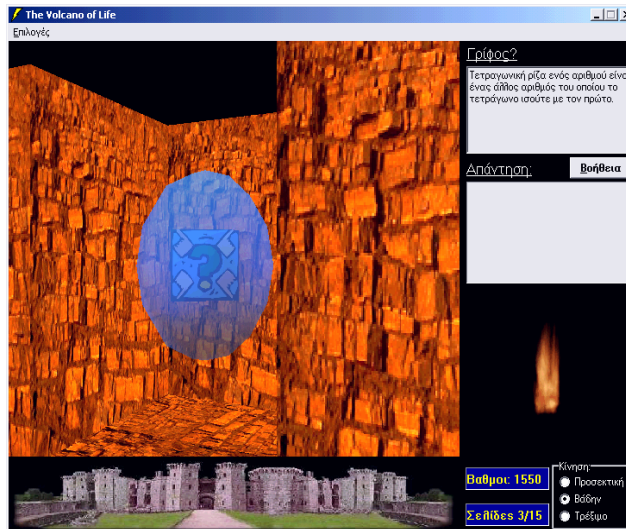

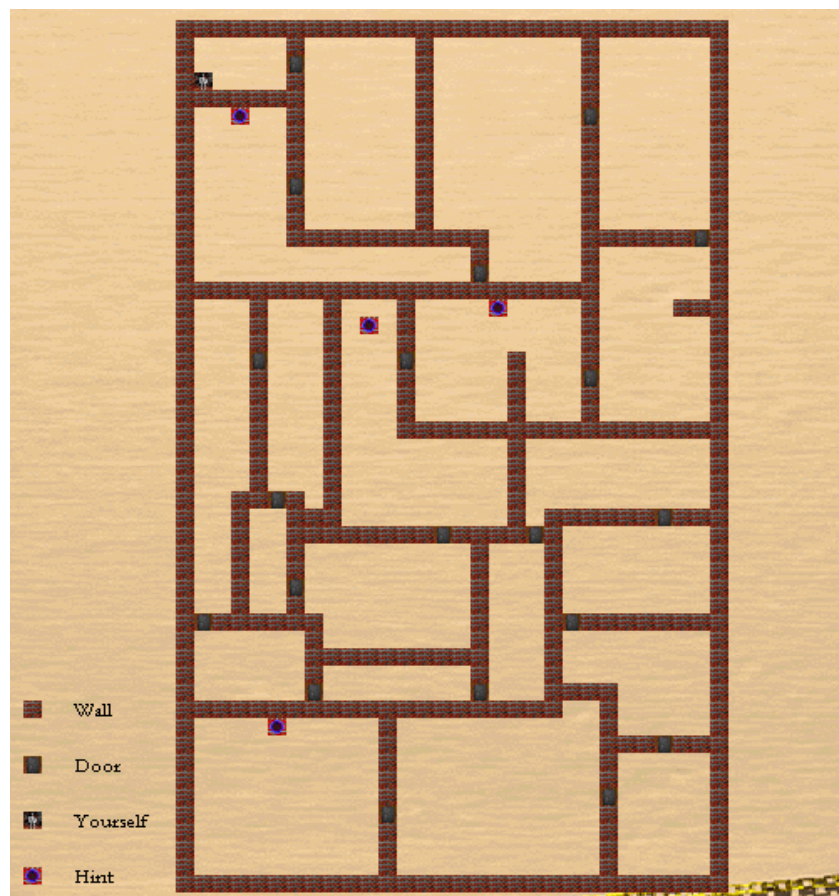Figure 2: A hint in the form of a blue ball in a virtual volcano world

Figure 3: An instantiation of a user's map

Another part of the adventure of the game involves giving players some bonuses in the form of "keys" which may be used to open doors from the virtual world. More specifically, if a student-player has answered 10 questions correctly the player is given a key as a bonus. If the user has a key, s/he is allowed to get through one guarded door of his/her choice without having to answer a question correctly. Thus, in a case where a student does not know an answer posed by a dragon, s/he can use a key that s/he may keep in his/her inventory list in order to open the door. S/he will not have to answer a question to get through the guarded door.

The user interface of VR-ENGAGE involves the navigation of the player through the virtual worlds of the game using the mouse and the keyboard. In case players are lost in the virtual world that they are currently in, they may use a map, which is a 2D view of the virtual world. All players have access to the map, which is available on line to show them where they are. The map is an essential part of the game and it illustrates an overview of the structure of the world, the position of the player in it, the doors that have not been opened yet and also the tutor-hints that are available in the world. An example of a map instantiation is illustrated in Figure 3.

The game was evaluated with respect to its educational effectiveness. For a full account of that evaluation the reader is referred to (Virvou et al. 2005). In that evaluation, we conducted an experiment where the educational game was compared to a similar educational application that had a conventional user interface without any virtual reality game. Both educational software applications had the same underlying reasoning mechanisms as well as the same help and theory functionalities. The main difference between the two educational software applications (game and non-game) was that one had a gaming approach whereas the other one did not have any gaming approach at all. In fact, the software with the simple user interface had a hypertext display of domain theory and exercises that were communicated to students through forms, dialogue boxes, buttons, drop-down menus etc. However, these exercises were not part of any story as in the gaming approach. Moreover, there was no virtual reality environment and no animated-speaking agents.

4. Aims, methods and settings of the evaluation experiments concerning usability and likeability

Many studies and common experience show that the prevailing preoccupation of students with computers at home is playing games. In contrast, games are not played at school. Mumtaz (2001) points out that this has created an enormous gap between home and school perceptions and use of computers by students. In this respect, the social context of the classroom or home of the use of computer games may affect their likeability. For example, in the context of a classroom, the game may seem more appealing to students in comparison with other media of education that students are used to, such as reading texts; on the other hand some students may dislike the game under the pressure of a school exam. The social context of the interaction is considered very important by many researchers. For example, Markopoulos & Bekker (2003a) point out that a successful study of children's interaction with technology requires that we understand the purpose and context for which children will interact with technology as well as their own needs.

In view of the above, the usability and likeability of the game had to be examined separately in the school environment and home environment of each subject. As a result, our evaluation consisted of two parts. The first part was conducted at schools and the second one at the students' homes. However, both parts of the evaluation involved the same students. The experiment aimed at estimating the usability and likeability of the game interface of VR-ENGAGE in the following 3 ways:

a. On VR-ENGAGE own stand
b. In comparison with the usability and likeability of educational software that had a simple user interface with no virtual reality and no game at all
c. In comparison with a commercial game of no educational content.

More specifically, both parts of the evaluation experiment involved 50 school children of 11-12 years old from 5 geography classes. The students that participated were selected from each geography class on the basis of their game playing expertise. For this purpose, all the students of the five classes were interviewed concerning their experience in game playing. For example, they were asked to estimate how long they used to spend weekly on playing virtual reality computer games. Moreover, they were asked about how long they had been familiar with such games, and they were requested to navigate for 5 minutes in the virtual worlds of a commercial game that was given to them. As a result, the students were classified into three groups: experienced, intermediate and novice players. Then, some students from each group were selected at random to participate so that there was sufficient representation of all three categories. Indeed, among the participants there were 15 novice, 20 intermediate and 15 expert game players. The novice players were given a short training before they used the game on their own. Both parts of the evaluation involved user-based methods of evaluation. The two parts of the evaluation are described in more detail in the following subsections.

4.1 First part of the evaluation

In the first part of the evaluation, all students were asked to play the educational game for about 2 hours in the environment of their school classroom as part of their assignments. During their interaction with the game, they were not allowed to receive any help from their co-students or lab assistants. While they played, all their actions were collected in protocols that were analysed later. The analysis of their actions aimed at evaluating the usability of the game for all three categories of student-player, namely expert, intermediate and novice student players. This kind of evaluation of the usability of the game was based on computer logging. As Chou (1999) points out, the major advantages of computer-logging are that it automatically and continuously collects objective data for further analysis and interpretation and does not interfere with users during their interactions with the system, whereas the major advantage of self-reporting is that it directly collects first-hand feedback from users.

Then, on a different day, after the 2 hours that all students had used VR-ENGAGE, they were given another educational software application that had a simple user interface with no virtual reality and no game at all. This session was given to children on a different day so that they would not be too tired by the previous session before they started a new one. Druin (2002) notes that short testing sessions, e.g. one afternoon are found to give useful input. Similarly with the case of VR-ENGAGE, the students were asked to use this application for 2 hours so as to get acquainted with its user interface, explore its usability and its educational content.

Both educational software applications (game and non-game) had the same underlying reasoning mechanisms as well as the same help and teaching theory functionalities. The main difference between the two educational software applications was that one had a gaming approach whereas the other one did not have any gaming approach at all. More specifically, the software with the simple user interface had a hypertext display of domain theory and exercises that were communicated to students through forms, dialogue boxes, buttons, drop-down menus etc. However, these exercises were not part of any story as in the gaming approach. Moreover, there was no virtual reality environment and no animated-speaking agents. For example, the way that the exam question, "Ethiopia is in Africa. Right or Wrong?" is presented to the user of VR-ENGAGE is illustrated in Figure 4, and the way the same question is presented to the user of the software with a simple user interface is illustrated in Figure 5.

On a different day, after the two hours spent on each application, all the students were given the choice to use for a maximum of 1 hour either VR-ENGAGE or the non-game application to make revisions on the lessons that they had been taught. Students who had spent most of their 1-hour time with VR-ENGAGE were considered to have preferred it over the other kind of educational application. This part of the experiment aimed at evaluating the likeability of the game against the non-game application for all three categories of student-player, namely expert, intermediate and novice student players. The students were also interviewed, after the one-hour of free use of the two applications, about the usability and the likeability of the two applications.

At a first glance, one might consider it obvious that school-children would prefer VR-ENGAGE since it would be more amusing. However, in reality it could not be foreseen whether the students who were novice game-players would find playing the game amusing and whether the expert game-players would find VR-ENGAGE interesting enough to bother to play. Finally, students might find the game distractive and thus might prefer the other kind of software for their learning purposes.
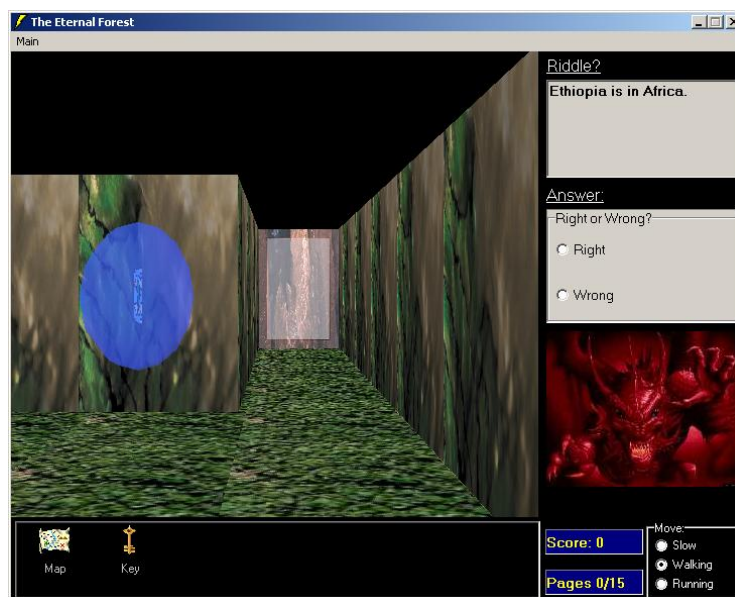

Figure 4: A question posed in the VR-ENGAGE

Figure 5: A question posed in the non-game educational application

4.2 Second part of the evaluation

The second part of the evaluation aimed at finding out the extent to which VR-ENGAGE could be liked and used by children and adolescents at their leisure time. The underlying rationale of this part of the evaluation was to find out whether VR-ENGAGE could replace other computer games, which did not have any educational value, in the children's preferences for their entertainment. In such a case, the children's game culture could be enriched with educationally beneficial games. Moreover, the educational game could be used both at work time and leisure time and thus would have a greater educational impact on children.

In view of the above aims, the usability and likeability of the educational software game interface was compared to the usability and likeability of a popular commercial game that had no educational content. The commercial game that the students were given to use during the weekend was a common adventure game and had similar game play functionality with VR-ENGAGE. However, as expected, the game play features of the commercial game, which focused solely on enjoyment, were numerous and in many cases more advanced than VR-ENGAGE. The commercial game had larger and more detailed virtual worlds, and its sound effects, graphics and environment were extremely professional. Also in order to navigate through the worlds of the commercial game, the players had to fight their way through by killing opponents as well as by solving riddles. In contrast, VR-ENGAGE was not violent at all since violence was not considered to be in accordance with the pedagogic principles that VR-ENGAGE was meant to serve.

Thus, after the 3 hours in maximum that the students had spent with VR-ENGAGE in the classroom (2 hours of exclusive use and 1 hour where they had a free choice between two programs) the same students were given the commercial game to play for a maximum of 3 hours time at school so that they could get acquainted with the environment and the plot of the game. Then the students were given VR-ENGAGE and the commercial game to take away to their homes during the weekend. Students were encouraged to play both of these games for as much as they liked but it was also made clear that they did not have to play if they did not want to. After the weekend the students were asked to fill in questionnaires that contained questions about the usability of each of the two games, the time spent on each one of them and the likeability of VR-ENGAGE as a game for their leisure time. These questionnaires were used to examine the usability and likeability of the educational software game interface, in comparison with the commercial game.

The second part, which was conducted at the students' homes, was mainly based on users' self-reporting methods since in that case the experiment aimed at evaluating the usability and likeability of the game during students' leisure time and thus, the time length of the experiment could not be predefined. In such conditions computer-logging was not feasible. However, the validity of the results of users' self-reporting methods is also quite high even when there is no other method for cross-examination of the results. This is so because children tend to be honest about their experiences. Indeed, Druin (2002) reports that children are "incredibly" honest when they are asked about computer programs that they have used and they can even be quite critical about them.

In the subsequent sections, first all the issues that relate to usability are analysed and discussed and then all the issues that relate to the likeability of the game are analysed and discussed. Finally the interplay of usability and likeability and important findings from the evaluation studies are analysed and discussed.

5. Usability evaluation issues

The protocols that were collected during the students' interaction with the educational game in classroom, at the first two hours of the first part of the evaluation, were analysed with respect to usability features. The analysis of the protocols led to the specification of three important characteristics of the students' interactions with the game. These characteristics were mainly observed in the novice and intermediate players' protocols and were the following:

1. *Game User Interface Acquaintance*, which showed the player's level of understanding of the Virtual Reality user interface of the game.
2. *VR Navigational Effort*, that showed how well the user could navigate through the Virtual World.
3. *VR Environment Distractions*: There were many cases when the Virtual Environment drew the player's attention to such an extent that s/he may have missed the main point of the educational game (which was learning a specific topic).

5.1 Game User Interface Acquaintance

While playing a game, a player has to become familiar with its user interface. This is very important because the level of understanding of the Virtual Reality user interface of a game is an essential part of the game itself. Playing becomes much easier if the user is aware of the components, functions and tools that are available for use.

More specifically, the way a student used or not used the functionalities that the game provided, revealed how acquainted he/she was with similar games. User interface concepts such as the "Inventory", the "Map", the "Tutor" etc., can make the gaming experience easier and more enjoyable. However, the student-players may have not been able to use these gaming user interface concepts due to their low level of user interface acquaintance. As we have mentioned earlier, we had a group of 50 students that used the virtual reality game. Among them there were 15 novices, 20 intermediate and 15 experienced game players.

By observing the students while they played during the 2 hours that they used the game in class we were able to measure their missed opportunities to use parts of the functionality of the gaming user interface in difficult situations. Our observations led to measuring how much time the students wasted, if they quitted and how much they improved during time or not. These conclusions were connected to the usability of the gaming environment. The more time the users wasted, the less usable the game was. The frequency of drop out (especially in the novice and intermediate protocols) revealed the degree of frustration of users due to usability problems. Finally, the level of improvement of students in the use of the user interface revealed the learnability level of the user interface.

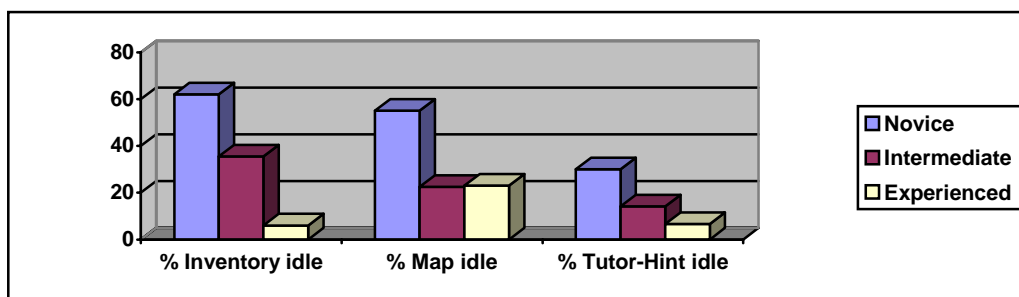Parts of the data that were collected are illustrated in Figure 6 and Figure 7.

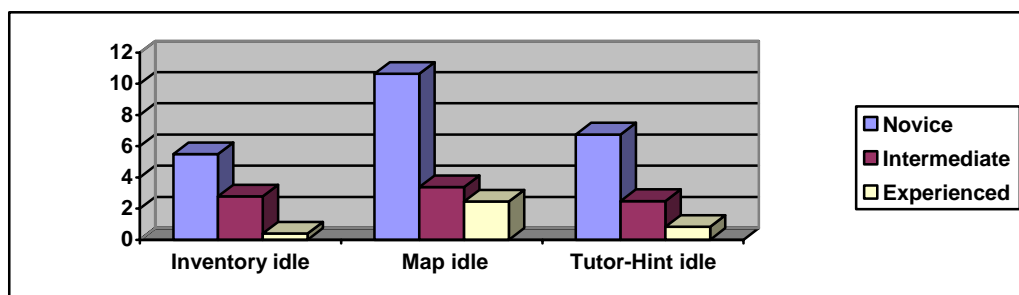Figure 6. The percentage of each functionality that was not used in cases it was needed.

Figure 7. The mean value of time (minutes) that each student of each category wasted by not using the functionality.

In cases the student players were in need of their inventory, novice users did not use it in 62% of cases, intermediate users in 35.5% of cases and experienced users in 6% of cases. Each time the inventory was not used by a student the mean value of time wasted was 53 seconds for a novice student, 47 seconds for an intermediate student and

40 seconds for an experienced student. The total time that each student had wasted for not using the inventory had a mean value of 5.48 minutes, 2.78 minutes and 0.4 minutes for each novice, intermediate and experienced student player respectively. Only twice the result was the student to quit the game. This occurred once for a novice student and once for an intermediate one. It was evident that the more gaming experience a student had the more s/he had used the inventory.

The use of the map was not activated by novice users in 55% of the cases needed, by intermediate uses in 22.5% and by experienced users in 23% of the cases needed. Each time the map was not used by a student the mean value of time that s/he wasted was 58 seconds for a novice student, 45 seconds for an intermediate student and 32 seconds for an experienced student. The total time that each student wasted for not activating the map, had a mean value of 10.63 minutes, 3.38 minutes and 2.45 minutes for each novice, intermediate and experienced student player respectively. It was evident that the novice players were not familiar with the use of a map. Also, the experienced users used the map less than the intermediate ones. It is likely that they thought it was easy to find their way without using the map and so they did not use it so much.

The use of hints was not activated by novice users in 30% of appearances, by intermediate users in 14.16% and by experienced users in 6.67%. Each time the hint was not used the mean value of time wasted was 225 seconds for a novice student, 175 seconds for an intermediate student and 125 seconds for an experienced student. The total time that each student wasted for not activating hints had a mean value of 6.75 minutes, 2.48 minutes and 0.83 minutes for each novice, intermediate and experienced student player respectively. Only once the result was an experienced student to quit. What can be concluded about novice users is that they missed the Tutor –Hint most of the times and not that they did not want to use it. Missing the tutor-hint was quite time consuming for students. That is because when a student missed the tutor-hint, s/he had to go back and find the tutor-hint and then use it.

Comparing the first hour of play with the second there was an improvement in using the inventory by 44%, 19% and 4% in novice, intermediate and experienced students respectively. Also there was an improvement in activating the map by 42%, 5% and 0% in novice, intermediate and experienced students respectively. Finally, there was an improvement in using the tutor-hint by 15.5%, 5% and 0% in novice, intermediate and experienced students respectively. It is easy to see that during the game the novice students became much more acquainted with the user interface than the intermediate, while the experienced users had minimum changes in their game play. This showed that the learnability of the game was quite high.

The above information, for the three characteristics of the user interface acquaintance that we measured, led to the following overall result about the total time wasted for each category of player. Each novice student player wasted a mean time of approximately 22.86 (5.48+10.63+6.75) minutes, each intermediate student player wasted a mean time of approximately 8.63 (2.78+3.38+2.48) minutes and each experienced student player wasted a mean time of approximately 3.69 (0.4+2.45+0.83) minutes, due to problems of the user interface of the educational game.

5.2 VR Navigational Effort

It is unrealistic to expect all students to know how to play a Virtual Reality Game. In cases when students are unfamiliar with such games, it may be quite difficult for them to get used to movements in a 3D environment. The main problem that inexperienced users have is the navigation in the parts of the virtual scenes.

The main navigational malfunctions that were observed in users' protocols of playing were twofold: aimless movement and disability to move. Aimless movement was observed in cases where the user moved around the same position without being able to steer him/herself around the virtual room. On the other hand, inability to move was observed in cases where the user had been stuck between virtual objects or next to walls and became unable to move again in a reasonable time.

By watching the students we were able to measure the times when they had navigational problems. The observation time spent per student was about 2. Parts of the data collected by the observation are illustrated in Figure 8 and Figure 9. We measured how much time students had wasted due to VR navigational problems, if they quitted and how much they improved during time or not.
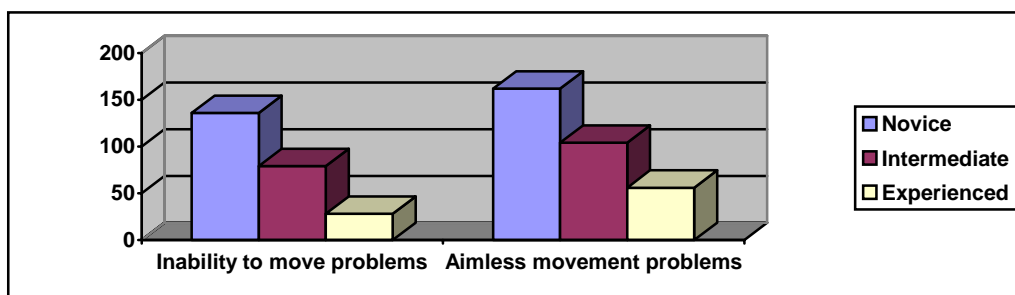

Figure 8. The number of occurrences of each navigation problem for all the students.
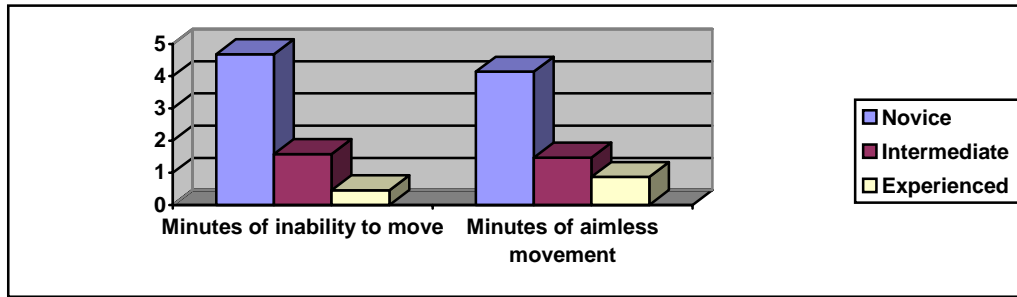
Figure 9. The mean value of time (minutes) that each student of each category wasted due to each navigation problem.

In particular, in cases the students had navigational problems, novice students were stuck between or beside objects and walls 136 times, intermediate students 79 times and experienced users 28 times. Each time a student was stuck, the mean value of time wasted was 31 seconds for a novice student, 24 seconds for an intermediate student and 15 seconds for an experienced student. The total time that each student wasted being stuck had a mean value of 4.68 minutes, 1.58 minutes and 0.47 minutes for each novice, intermediate and experienced student player respectively. Only once the result was a novice student to quit the game. It is obvious that the experienced users had no difficulty with problems of inability to move.

The aimless movement in the virtual world occurred for novice students 162 times, for intermediate students 104 times and for experienced users 56 times. Each time a student was confused the mean value of time wasted was 23 seconds for a novice student, 17 seconds for an intermediate student and 14 seconds for an experienced student. The total time that each student had wasted in aimless movements had a mean value of 4.14 minutes, 1.47 minutes and 0.87 minutes for each novice, intermediate and experienced student player respectively. No student quitted due to aimless movement.

Comparing the first hour of play with the second there was an improvement in problems of inability to move by 18, 11 and 6 less problems of this kind in novice, intermediate and experienced students respectively. Also there was an improvement in aimless movement by 10 and 6 less problems of this kind in novice and intermediate students respectively.

All the above information for the two characteristics of navigational problems result to the following. Every novice user wasted approximately 8.82 (4.68+4.14) minutes, each intermediate user wasted approximately 3.05 (1.58+1.47) minutes and each experienced user wasted approximately 1.34 (0.47+0.87) minutes, due to navigational difficulties in using the educational game.

5.3 VR Environment Distractions

The main aim of the virtual reality game was to make learning more motivating for students. However, the VR-environment might result in the distraction of students, which would be an undesirable effect with respect to the educational aims of VR-ENGAGE. Indeed, there were some cases that players were observed not paying so much attention to the real purpose of the game, which was to answer all the questions and get out of the labyrinth of the virtual world. Instead, they wandered around the virtual world.

Although it was desirable to have very specific measurements about the distraction problem, it was difficult to know when a student was actually in a distraction condition and to measure how much time he\she wasted by being distracted. For example in a case that a player did not know how to answer a question in order to have a door opened in the labyrinth and went back to find the tutor-hint to get help, the time that passed until the player went back to give the answer was affected by many parameters. The student may have had VR navigation problems; he/she may not have used his/her inventory or may have been distracted. Of course every problem can occur more than one time in such a case.

Despite these problems there were cases where it was clear that the students had delayed their playing due to inability to focus on the purpose of the game. An example of a clear case of virtual environment distraction is the continuous use of agents and tutor-hints to get help. That shows that the student likes to play with the agents and hear their voice. Some other examples are: 1. Navigating around in the virtual worlds of the game and visiting rooms that the player has already passed through before, 2. Always giving wrong answers and not trying to think about a question just to continue his/her way in the virtual world although s/he will get a bad grade and will not be able to explore the whole virtual world.

Some statistical information about such cases, which was collected during the 2 hours of observation time for each student, is illustrated in Figures 10 and 11. This information was gained in cases it was obvious that the student was delaying due to not focusing on the purpose of the game. These data show how many times a student was distracted and what percentage in the total delay of the student was due distraction.

For novice users, distraction occurrence averaged about 3.5 times, for intermediate users 6 times and for experienced users 4 times. The total time that each student wasted being distracted, had a mean value of 2.9 minutes, 5.2 minutes and 4.2 minutes for each novice, intermediate and experienced student player respectively.
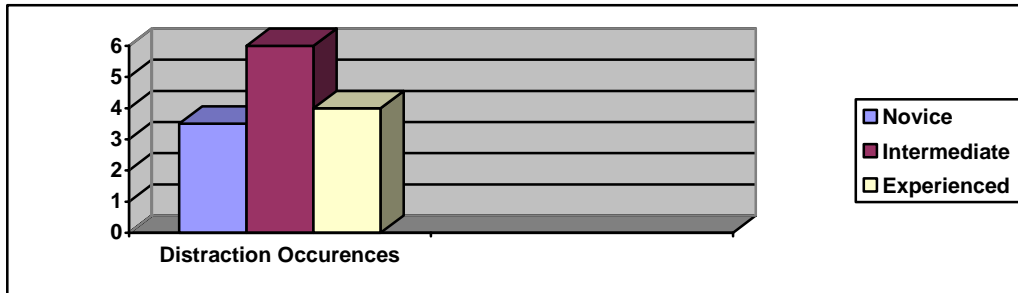

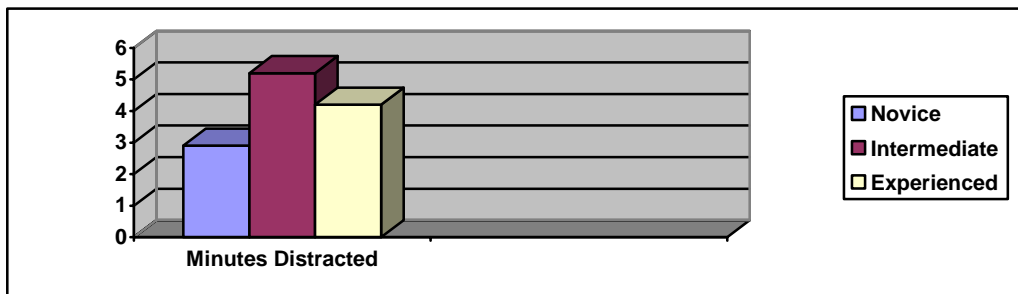Figure 10. The mean value of distraction occurrences of each category.

Figure 11. The mean value of time (minutes) that each student of each category wasted due to distraction.

5.4 Total time wasted due to VR usability problems

By collecting the data from the above three important characteristics of usability problems of the students' interactions with the game we measured how much time each student had wasted in total. Table 1 presents the mean values of time (minutes) that every student of each category wasted due to each VR environment usability problem. Novice users had wasted the longest amount of time and experienced users had wasted the shortest amount of time due to these problems. In fact, the main delaying problem for experienced users was due to virtual environment distractions unlike the other two categories of users who did not have so many virtual environment distractions but had VR-navigational problems and gaming user interface problems.

	Mean value of minutes wasted due to			
	UI acquaintance	Navigation problems	VR environment Distraction	Sum
Novice student	22.86	8.82	2.9	34.58
Intermediate student	8.63	3.05	5.2	16.88
Experienced student	3.69	1.34	4.2	9.63

Table 1: Results of the collected protocols on students' time wasted while using the VR-ENGAGE for the first 2 hours.

5.5 Usability comparison based on users' opinions

In the four previous sections we examined three important characteristics of the students' interactions with the game during the first 2 hours spent using VR-ENGAGE. In this Section we describe what the impact of these operational problems has been on the usability of the VR-ENGAGE in comparison with the non-game educational application, according to the opinions of users. Additionally, at the second part of the evaluation, which took place at

students' homes during the weekend, we made a similar comparison on the usability of our virtual reality game and the commercial one.

These comparisons are based on the usability part of the students' interviews. The first interview took place after the 1 hour of free time class experiment, and the second after the weekend. The interviews concerned both the usability and the likeability of the systems that the students had used and their comparison: VR-ENGAGE vs non-game educational application, and VR-ENGAGE vs commercial game of no educational content. In this Section, we analyse and discuss the parts of the interviews that concerned usability. In Section 6, we further analyse and discuss the interviews in terms of likeability issues.

Figure 12 illustrates important parts of information concerning the opinions of every category of students, depending on their gaming experience, about which application between VR-ENGAGE and the non-game educational application was easier for them to use. As expected, the virtual reality software was considered more complicated and more difficult than the non-game software. Only 26% of the students found easier to play with VR-ENGAGE. This percentage consisted of 6.7%, 25% and 46.7% of the 15 novice, 20 intermediate and 15 experienced students respectively.
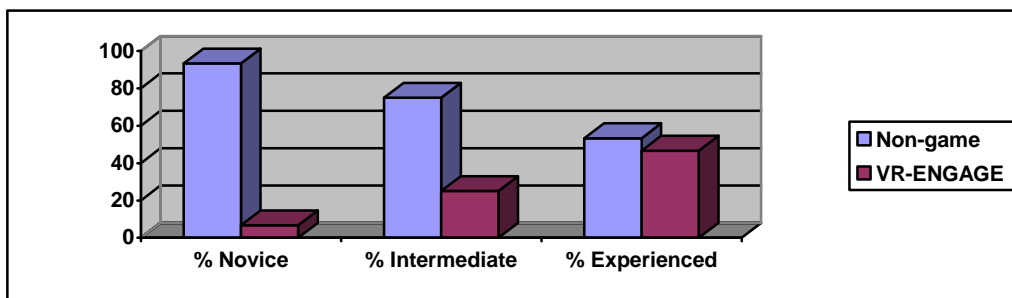

Figure 12. Which of the VR-ENGAGE and the non-game educational application was easier to use.
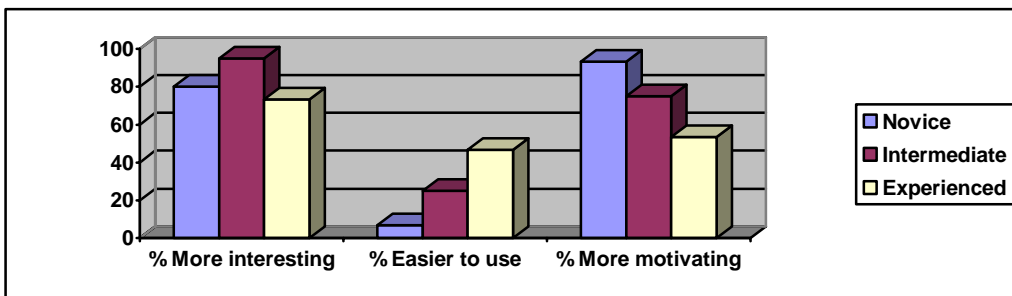

Figure 13. Which of the VR-ENGAGE and the Commercial game was easier to use.

Figure 13 illustrates similar data for the comparison between VR-ENGAGE and the commercial game that the students were given to use at their leisure time at home. Again as expected the commercial virtual reality software proved to be more complicated and thus more difficult to use than VR-ENGAGE. 76% of the students found VR-ENGAGE easier to play with. More specifically, 86.7% of the novice users, 80% of the intermediate users and 60% of the experienced users found VR-ENGAGE easier to use than the commercial game.

6. Likeability evaluation issues

The likeability of VR-ENGAGE was evaluated based on two comparative studies: 1) One study in the context of a classroom that compared the likeability of VR-ENGAGE and the likeability of the non-game educational software application. 2) One study in the context of the children's leisure time at their homes, which compared the likeability of VR-ENGAGE and the likeability of a commercial game of no educational content. In the first study, likeability was assessed in terms of the amount of time that children had spent on each application as this was recorded in the students' protocols. The longer times in favour of one application were considered to signify the children's preference for this application. Moreover, these findings were confirmed by the students' answers to direct questions concerning the likeability of each application and their comparison.

6.1 Comparative study in classrooms

As mentioned earlier, both parts of the likeability evaluation involved 50 students of 11-12 years of an elementary school. The students were separated into three groups of 15, 20 and 15 children of novice, intermediate and expert users depending on their gaming expertise. After all the children had used VR-ENGAGE and the non-game educational application for 2 hours respectively, they were given the choice to use for a maximum of 1 hour, either of the applications to make revisions on the lessons that they had been taught. Moreover, after they had completed the use of both applications the students were interviewed concerning the likeability and usability of the programs.

6.1.1 Analysis of observed students' behaviour

The results during the one-hour of free time in class showed a significant preference of the student-users for VR-ENGAGE in comparison with the non-game educational software. In particular, for novice, intermediate, and expert users the results showed a 59%, 62% and 74% preference for the virtual reality game respectively, as compared to 37%, 32% and 19% respective preference for the other application. In total, all the groups of students spent 64% of their time using VR-ENGAGE and 30% of their time using the non-game application. The reason that the sums of the respective percentages are not 100% is because there was a small amount of time for some students spent in neither of the two applications. The total results showed that VR-ENGAGE had achieved its aim of being more attractive and motivating, for the students than the non-game educational software.

In more detail, the total time in minutes that was available for the students of each group was 15x60 minutes = 900 minutes for the groups of the novice and experienced student-players, 20x60 minutes = 1200 minutes for the group of the intermediate student-players and 50x60 minutes = 3000 minutes for all the students. In total, the students of novice, intermediate and expert game users used the VR-ENGAGE application for 532 of 900, 741 of 1200, and 664 of 900 minutes respectively. On the other hand the same categories used the conventional educational software for 333 of 900, 387 of 1200, and 168 of 900 minutes respectively. In total, all the groups of students (when summed up) had used VR-ENGAGE for 1937 of the 3000 minutes available, while they had used the non-game application for 888 minutes.

The statistical analysis, which took place after the 1-hour use of either of the two applications, concerned the significance of the difference on the time spent in each of the two applications for each of the groups of students and for all the students. There were 4 t-tests performed to compare the time spent on VR-ENGAGE and the time spent on the application with the simple user interface. Each of the t-tests performed referred to one of the below categories respectively:

- Novice game-players (students),
- Intermediate game-players (students),
- Expert game-players (students),
- All the students.

The null hypothesis, H_0, was that there was no difference between the lengths of time spent for each of the two applications. The research hypothesis, H_1, was that there was significant difference between the lengths of time spent for each of the two applications. The t-value results of 2.71 for the intermediate students-players, 4.26 for the expert students-players and 5.1 for all the students were significantly greater than their critical values 1.73, 1.76 and 1.68 respectively. This showed that the difference in the time spent between the two applications was statistically significant for the groups of intermediate users, expert users and all the students together. The t-value result of 1.99 for the novice students-players was greater than its critical value 1.76. Although this difference was also statistically significant the small difference from its critical value reveals potential difficulties in using the VR-ENGAGE due to lack of experience of novice users. So for each of the four t-tests we can reject H_0 and accept H_1. The time spent on the VR-ENGAGE was significantly greater than the time spent on the other educational application.

The results of the above four t-tests are summarised in Table 2. In particular, Table 2 illustrates the mean values of the differences in the time spent for each of the groups and for all the students in using VR-ENGAGE and the non-game application. Additionally it includes the respective results of the four t-tests. These are the results of the t-tests, to find out any difference in the time spent between the two applications. These results involve the standard errors of the differences, the T values and the Critical values of each of the t-tests.

Variable	VR-ENGAGE	Non-game educational application			T_Value, C_Value
	Mean Value of minutes played (0-60)	Mean Value of minutes played (0-60)	Mean of Differences	Standard error of Differences	
Novice student respective results	35.47	22.2	13.27	6.64	Tv = 1.99 Cv = 1.76
Intermediate student respective results	37.05	19.35	17.07	6.53	Tv = 2.71 Cv = 1.73
Experienced student respective results	44.27	11.20	33.07	7.76	Tv = 4.26 Cv = 1.76
All the student respective results	38.74	17.76	20.98	4.11	Tv = 5.10 Cv = 1.68

Table 2: Results of the analysis of the students' time spent by each group either on VR-ENGAGE or the non-game educational application.

In the above t-tests, the t-value of each t-test is calculated by performing a t-test for correlated samples for the time spent on each of the applications (Voelker 2001). The critical value for each t-test is the value taken from Table T for a one-tailed research hypothesis depending on the sample number. The t-test results show that there is a statistically significant difference for the two samples in favour of the time spent using VR-ENGAGE for every category of students.

The results of the t-test for the *novice game-players* have shown that t value (1.99) is greater than the critical value (1.76). Therefore novice students prefer to play VR-ENGAGE than to use the non-game educational application during free time at school, or to repeat lessons. However, we can see that the difference is not great. This is expected to some extent because novice users have more operating difficulties with a virtual reality system than experienced users and thus they may be put off from the use of the game for this reason. In the case of *intermediate game-players* the t value (2.71) is much greater than the critical value (1.73). Therefore intermediate students strongly prefer to play VR-ENGAGE than use the non-game application during free time at school, or to repeat lessons. For the *experienced student-players* the t value (4.76) is extremely greater than the critical value (1.76). Therefore experienced students almost all the time preferred to use VR-ENGAGE during free time at school. This may be explained because experienced users have no operating difficulties with a virtual reality system. Moreover, the fact that they are experts means that they usually spend a long time playing games, which means that they are used to the culture of games and are fascinated by it and in some cases, they are even addicted to it. These users were very pleased to see the games introduced in the classroom. So they play to have fun. Finally the t value (5.10) for *the whole group of students* is extremely greater than the corresponding critical value (1.68). Therefore the total of students prefer to play VR-ENGAGE than use the non-game educational application during their free time in class, or to repeat lessons at the school classroom.

6.1.2 Analysis of students' answers to interviews concerning the likeability of the two applications

After the one hour use of either of the applications to make revisions on the lessons that the students had been taught, they were interviewed about the two applications used, the virtual reality educational game and the non-game educational application. These interviews included questions about the usability and the likeability of the two applications.

Among the questions about likeability asked during the interviews were the following:
1. Which application was more motivating?
2. Which one was more interesting?
3. Did you get tired of any of the applications?
4. What do you think of VR-ENGAGE as a game?
5. What would you add to VR-ENGAGE?
6. What do you think of learning while playing in class?
7. Do you have any other comments?

Figure 14 illustrates important parts of the interviews information concerning the opinions of students about which application between VR-ENGAGE and the non-game educational application was more interesting The virtual reality

game was reported as more interesting by 84% of the students. This consisted of 80%, 95% and 73.3% of the 15 novice, 20 intermediate and 15 experienced students respectively. There is no doubt that the students were more interested in the game than the other application.
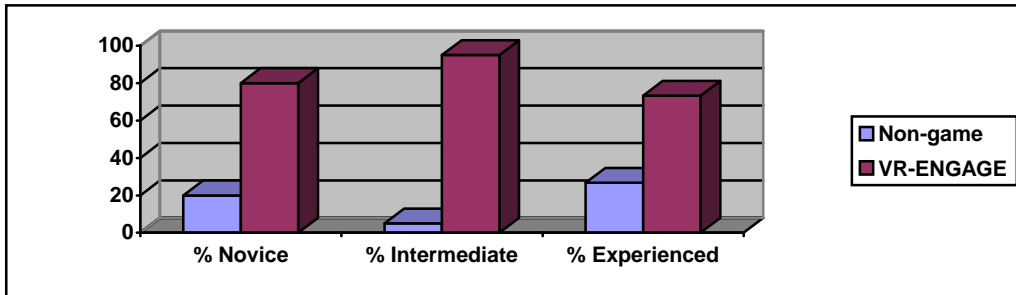

Figure 14. Which one between VR-ENGAGE and the non-game educational application was more interesting.

Concerning the interface of each application, the tasks that needed to be accomplished for each of them, and the opportunities offered, the players of the two applications commented on their motivation. The results are gathered in figure 15 below. 74% of the students thought of VR-ENGAGE to be more motivating than the non-game application. More specifically, 93.3% of the novice, 75% of the intermediate and 53.3% of the experienced in game playing students reported that VR-ENGAGE was more motivating.
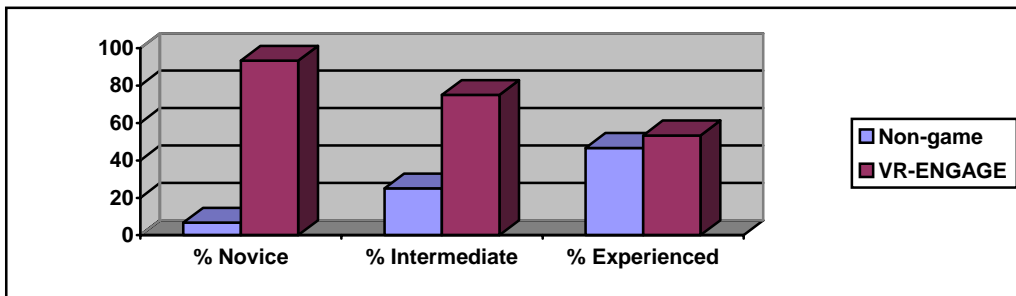

Figure 15. Which one between VR-ENGAGE and the non-game educational application was more motivating.

6.2 Comparative study during users' leisure time

The second phase of the empirical study was based on the 50 students' interviews after they had used VR-ENGAGE and a commercial game at their homes during the weekend. The students were again separated into the same three groups of 15, 20 and 15 children of novice, intermediate and expert users depending on their gaming expertise. The students' answers to the questionnaires of the interviews were used for showing what the comparison results were for the usability and the likeability of the two applications. However, the students' answers were also analysed to reveal how students had liked or disliked VR-ENGAGE irrespective of its comparison with a commercial game.

The questionnaires included among others the following questions:
1. How long did you spent using VR-ENGAGE?
2. How long did you spent playing with the commercial game?
3. Which application was more motivating?
4. Which was more interesting?
5. Did you get bored of any of the applications?
6. What do you think of VR-ENGAGE as a game?
7. What would you add to VR-ENGAGE?
8. What do you think of learning while playing at home?
9. Do you have any other comments?

6.2.1 Analysis of students' answers to interviews concerning the time spent on the two applications during the weekend

The results showed a significant preference of the student-users for the commercial game in comparison with the VR-ENGAGE. In particular, for novice, intermediate, and expert users the results showed a 53%, 62% and 73% preference for the commercial game respectively, comparing to a 47%, 38% and 27% preference for the virtual reality educational game. In total, all the groups of students spent 65% of their time using the commercial game and 35% of their time using VR-ENGAGE. The reason that the sums of the respective percentages are 100% is because the percentages are calculated by finding the part of the time spent in each of the applications to the total time using both of them, for each of the student categories. It is certain that the mean value of the total time spent using any of the two applications defers between the three groups. For example the experienced game players would definitely spend more time using the applications, because they would be more attracted to them.

In more detail, the total time in hours that was spent by the students of each group using any of the applications was 60 hours for the 15 novice student-players, 109 hours for the 20 intermediate student-players, 102 hours for the 15 experienced student-players, and 60+109+102 = 271 hours for all the students. In total, the students of novice, intermediate and expert game users used the commercial game for 32 of the 60 hours, 68 of the 109 hours, and 75 of the 102 hours respectively. On the other hand the same categories used VR-ENGAGE for 28 of the 60 hours, 41 of the 109 hours, and 27 of the 102 hours respectively. All the groups of students combined in using the commercial game for 175 of the 271 total hours, while using our educational application for 96 hours.

The statistical analysis, which took place after the questionnaires of the students had been examined, concerned the significance of the difference on the time spent in each of the two applications for each of the groups of students and for all the students. There were 4 t-tests performed to compare the time spent on the commercial game and the time spent on VR-ENGAGE. Each of the t-tests performed referred to one of the below categories respectively:

- Novice game-players (students),
- Intermediate game-players (students),
- Expert game-players (students),
- All the students.

The null hypothesis, H_0, was that there was no difference between the time spent for each of the two applications. The research hypothesis, H_1, was that there was significant difference between the time spent for each of the two applications. The t-value results of 2.08 for the intermediate students-players, 4.05 for the expert students-players and 3.95 for all the students were significantly greater than their critical values 1.73, 1.76 and 1.68 respectively. This showed that the difference in the time spent between the two applications was statistically significant for the groups of intermediate users, expert users and all the students together. So for each of the three t-tests for intermediate, expert and all the users we can reject H_0 and accept H_1. The time spent on the commercial game was significantly greater than the time spent on VR-ENGAGE for leisure time at home.

The t-value result of 0.56 for the novice students-players was much smaller than its critical value 1.76. This result revealed that there was no statistically significant difference in using the two applications for the novice users. So for the category of novice student-players we can accept H_0. Therefore novice students do not prefer to play a commercial game more than VR-ENGAGE for leisure time at home.

Variable	Commercial game	VR-ENGAGE			T_Value, C_Value
	Mean Value of hours played	Mean Value of hours played	Mean of Differences	Standard error of Differences	
Novice student respective results	2.13	1.87	0.27	0.48	Tv = 0.56 Cv = 1.76
Intermediate student respective results	3.40	2.05	1.35	0.65	Tv = 2.08 Cv = 1.73
Experienced student respective results	5.00	1.80	3.20	0.79	Tv = 4.05 Cv = 1.76
All the student respective results	3.50	1.92	1.58	0.40	Tv = 3.95 Cv = 1.68

Table 3: Results of the analysis of the students' leisure time spent by each group either on VR-ENGAGE or the commercial game during the weekend.

The results of the above four t-tests are summarised in Table 3. In particular, Table 3 illustrates the mean values of the differences in the time spent for each of the groups and for all the students in using the commercial game and VR-

ENGAGE. Additionally it includes the respective results of the four t-tests. These are the results of the t-tests, to find out any difference in the time spent between the two applications. These results involve the standard errors of the differences, the T values and the Critical values of each of the t-tests.

In the above t-tests, the t-value of each t-test is calculated by performing a t-test for correlated samples for the time spent on each of the applications. The critical value for each t-test is the value taken from Table T for a one-tailed research hypothesis depending on the sample number.

The results of the t-test for the novice game-players have shown that t value (0.56) is much smaller than the critical value (1.76). Therefore novice students do not prefer to play a commercial game more than VR-ENGAGE for leisure time at home. This can be explained because novice users have operating difficulties with virtual, which is even more difficult for them to handle. As a result the time spent by novice users in either of the applications was the smallest as compared to the other two groups of users as a consequence of these difficulties. In the case of intermediate game-players the t value (2.08) is much greater than the critical value (1.73). Therefore intermediate students prefer to play a commercial game more than VR-Engage for leisure time at home. For the experienced student-players the t value (4.05) is extremely greater than the critical value (1.76). Therefore experienced students almost all the time played commercial games for leisure time at home. We can understand this because experienced users have no operating difficulties with a virtual reality system. So they prefer commercial games, which are more sophisticated than VR-ENGAGE and thus seem more amusing to them. Concluding the t value (3.95) for the whole number of students is extremely greater than the corresponding critical value (1.68). Therefore students in total prefer to play a commercial game more than VR-Engage for leisure time at home.

The above t-tests' results were expected to some extent. Indeed, it was almost certain that the commercial game would be more appealing than VR-ENGAGE. This is so because commercial games have very sophisticated virtual environments and can be more challenging in terms of adventure since they do not have to care about educational content.

6.2.2 Analysis of students' answers to interviews concerning the likeability of the two applications

The answers to questionnaires provided useful data about the likeability of the virtual reality game in comparison with the commercial game. Figure 16 illustrates important parts of the questionnaires information about which application between VR-ENGAGE and the commercial game educational application was more interesting. The virtual reality game for teaching was characterised more interesting by the 38% of the students, consisting of 60%, 40% and 13.3% of the 15 novice, 20 intermediate and 15 experienced students respectively. There is no doubt that the students in total were more interested in the commercial game.
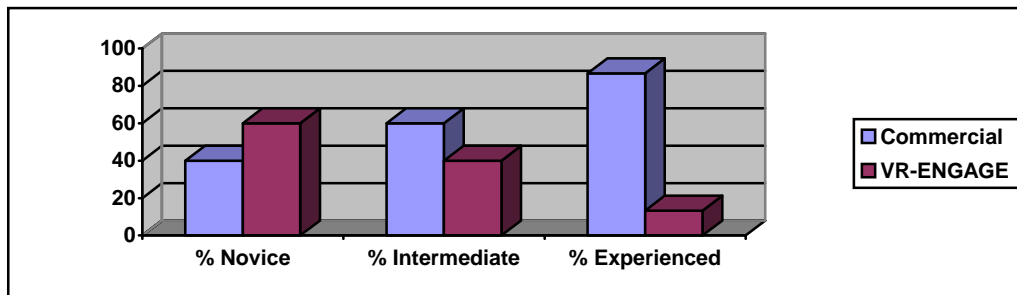

Figure 16. Which one between VR-ENGAGE and the commercial game was more interesting.

The virtual reality user interface of each application, the tasks that needed to be accomplished for each of them, and the components and functions offered, gave the players the opportunity to comment on their motivation. The results are presented in figure 17 below. Only 30% of the students thought of VR-ENGAGE to be more motivating than the commercial application. In detail it was 33.3% of the novice, 35% of the intermediate and 20% of the experienced in game playing students who claimed that.
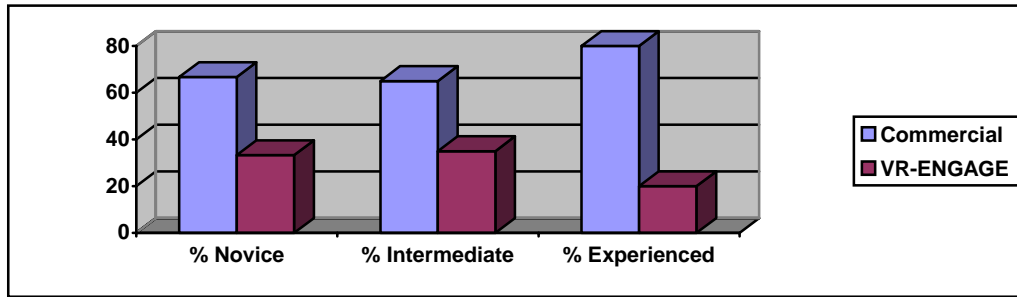
Figure 17. Which one between VR-ENGAGE and the commercial game was more motivating.

The results from the second part of the evaluation at student's homes were quite different from the first part. Since children were not given the game to work with it as an assignment, they considered it merely as a game similar to the commercial games they were familiar with. Therefore their judgment on it focused on the game environment. The students' interviews revealed many interesting comments about what they expected and what they would like. Most of the students (62%) pointed out that the game would be better as a game if it had more virtual objects, more background sounds and more adventure. These comments came to a large extent from experienced game players rather than novice ones. This was due to the fact that most of them were familiar with commercial virtual reality games therefore they compared VR-ENGAGE with them and had higher expectations in this aspect. Some of the students (8%) criticised it for being non-violent. Again, this was probably due to the fact that the culture of commercial VR-games has penetrated the world of adolescents and children in a way that they expect all games to be similar even if this is not good for them.

However, although most children had given comments for the enhancement of the entertaining aspect of the game, a very large percentage of them (84%) said that they would like to have this game at their homes and play it at their leisure time together with other computer games they had. This was a very encouraging result.

7. Discussion of the results

The virtual reality user interface of an educational software game has been analysed based on the interactions of students with it. The students were classified into three categories depending on their expertise in game-playing. These categories were experienced, intermediate and novice game players. Then usability and likeability issues were examined with respect to the three categories of user. The results of this evaluation led to valuable insight about how virtual reality educational games can be improved in the near future concerning usability and likeability issues.

7.1 Usability

The results of the analysis of the students' protocols in terms of usability of the VR educational game, revealed that the students' actions could be interpreted in terms of three interaction features: user interface acquaintance, navigational effort and VR environment distractions. As a result, in VR-ENGAGE, evaluation of usability was based on a similar classification of learning and playing behaviour as the one made by (Yacci et al. 2004) for edutainment environments. However, in the case of VR-ENGAGE the classification that has been made, additionally takes into account the fact that the user-player has to play in a 3-D virtual reality game, which constitutes a special broad category of gaming environments and poses additional special problems with respect to usability and instructional goals of the educational games. In particular, in VR-ENGAGE, Game user interface acquaintance refers to what is identified as "operations" and "strategy" in (Yacci et al. 2004). VR navigational effort is a behaviour characteristic that concerns the player's skill on navigation in the 3-D virtual reality worlds of a VR-game. Finally, VR environment distractions refer to a player's behaviour characteristic that shows that the user is not pursuing the educational goals of the educational game due to virtual reality distractions.

The user interface acquaintance showed the extent to which the students knew how to use the user interface of the game in terms of what was considered as a "legal movement" in it as well as game strategies that would help the player win the game. This feature was affected by the narrative of the game but not from the virtual reality environment. In contrast, the navigational effort and the VR environment distractions, were affected by the virtual reality environment. The navigational effort was greater for students who used to lose their way in the virtual worlds than for students who did not. On average, for the total of students who participated, this did not happen very frequently. However, when the protocols of each category of student-players were examined separately, it was revealed that novice student-players made quite a lot of navigational effort. This means that on average they had wasted quite a lot of their time trying to find their way in the virtual reality worlds and thus they had been left less time for reading the theory and answering questions that would help them extend and consolidate their knowledge. In such cases, students did not benefit as much as they could from the educational content of the game due to the virtual reality interface. As a consequence, this

feature has to be improved for the novice game players as well. Improvements may include the addition of more maps, which should be more explicit and thus more comprehensible and also adaptive help that would take into account each individual user's circumstances to guide them out of problematic situations.

Finally, the virtual reality distractions were observed in many students' protocols but not to a great extent. The distractions occurred when users behaved as if they had forgotten what their ultimate goal was which was related to reading the lessons' theory and answering questions. Instead, they repeated actions without any particular meaning. The explanation of this interaction feature can be attributed to several causes. One cause is that the user is possibly distracted by VR features of the game, such as movements of animated agents, virtual objects, music etc. However, it may also be because the users have been absent–minded. A possible improvement for this problem could be the addition of more user options concerning the number of virtual objects, the speed of the movement of animated agents, the existence of music or the kind of music played in the background. If students are given more options to customise the virtual environment components then they may be less distracted.

The observations showed that the majority of students did not have major usability problems while interacting with the game. However, one important finding is that the first two kinds of usability problem (user interface acquaintance and navigational effort) affect mostly the less experienced whereas the third kind of usability problem (virtual reality distractions) affects mostly the more experienced users. These findings show that all categories of user may benefit less than they could from the educational content of a VR-educational game due to usability problems. Thus the design of VR-game interfaces has to attract a lot of attention for the elimination or improvement of these three kinds of problem.

7.2 Likeability

In the environment of school, VR-ENGAGE was considered more likeable than the non-game educational application by the students. In fact, the interviews revealed that the students were fascinated by the idea of a game in the classroom and they were certainly more enthusiastic about this kind of software than the non-game educational application. In this respect, the results of the evaluation definitely show that it is worth trying to include educational games in classrooms since these games can be very appealing to students, resulting in their extra motivation to learn the material taught. This finding is in accordance with the findings that came out from the evaluation that focused on the pure educational effects of VR-ENGAGE (Virvou et al. 2005). The results from both evaluations show that students are indeed motivated by an educational game in class and they learn better from it than from other kinds of educational software. This is especially true for students who used to have poor academic performance, which include students that are not easily disciplined in class.

However, in the home environment, the commercial game was considered more likeable than VR-ENGAGE. This can be explained by the fact that commercial games are made just for fun and they are not related to the educational syllabus of schools. The educational syllabus in a game may be stressful for some student-players because these students are reminded of school at their leisure time. This finding confirms to some extent Brody's criticism (1993) about educational games that the marriage of education and game-like entertainment has produced some not very-entertaining learning activities. However, this problem appears only when the educational game is compared to a commercial game. Otherwise, on its own stand the educational game is quite likeable by the students at their leisure time if they have not any other game to compare it with.

One design lesson from these results is that educational games have to provide very sophisticated virtual reality environments to compete with commercial non-educational games. At the same time they have to provide as much help as possible concerning usability issues to avoid discouraging non-expert game –players from playing the game.. This is so because commercial games have very high quality VR- user interfaces and features. However, due to the numerous operating difficulties that some of the students faced (mainly novice users) with the commercial game they found VR-ENGAGE application more interesting and they devoted to it similar amount of time. On the other hand the small interest that the experienced users showed to VR-ENGAGE in comparison to the commercial game, is due to the fact that these users are used to playing commercial games very often and thus they are difficult to satisfy. Another important issue is that many players are used to features, which are not considered to be in accordance to pedagogic principles. Such features include violence of VR-agents. In this respect, many users like the educational game less than the commercial game. However, in our opinion it is worth the effort to produce educational games, which would be very competitive to commercial games without compromising with non-pedagogic gaming aspects, so that both the pedagogic and educational principles are served to the best.

7.3 Likeability in association with usability

A vital question is how likeable it is for a student to interact with a particular application (educational game, non-game educational application or commercial game) depending on its usability. What is very important to note is that paradoxically, students seem to have preferred the applications that were more difficult to use but had a challenging game interface. Indeed, from the answers of students to interviews, the majority of all kinds of students had stated that VR-ENGAGE was more motivating and interesting than the non-game educational application and that the commercial game was more motivating and interesting than VR-ENGAGE. However, in usability comparison the commercial game

was considered more difficult to use than VR-ENGAGE, which in turn was considered more difficult to use than the non-game educational application. It seems that usability problems do not prevent students from enjoying a virtual reality user interface, which they rank higher in their preferences.

However, the above results do not undermine usability problems, which certainly have a negative effect on both the educational benefits of students and on the likeability of the interfaces. This is also confirmed by the length of time that users had used the applications when they had a free choice. For example, the novice players who had the highest number of usability problems with VR game interfaces had used VR-ENGAGE for very little longer than the non-game educational application despite the fact that the vast majority of them had claimed that VR-ENGAGE was more interesting and more motivating than the non-game educational application. Obviously, they were put off from using VR-ENGAGE for longer times by the usability problems they had. Similarly, novice players had used the commercial game for less time than VR-ENGAGE despite the fact that the majority of them had claimed that the commercial game was more motivating than VR-ENGAGE (although not more interesting). The time length that students had spent on the educational game is quite important for the educational aspect of the application. The more time students spend on the application the more chance they have to revise topics and learn. Similarly, the less they are distracted by usability problems the more they can benefit from the educational content.

8. Conclusions

The evaluation of the educational VR-game in terms of usability and likeability has shown that usability problems do exist to some extent in virtual reality environments and they affect mostly the novice VR-game players. However, they do not discourage users from playing the educational game. In fact, a very important finding of this evaluation is that likeability is proportional to the sophistication of the VR-environment of the game.

Moreover, if VR educational games are designed to reach a high degree of sophistication they will be quite competitive to commercial games of no educational content. This will provide the asset of having educational applications that can be equally attractive for the students in the school classroom and at home. In this way the potential users will be more motivated to make better use of both their school time and leisure time to the benefit of educational goals. Presently, from the evaluation of the virtual reality educational game, it was shown that users were more difficult to satisfy for home use of the educational game than in classrooms. In fact, we had very good results of likeability of the game in classrooms. If these results are combined with the results of the evaluation of the same VR-game concerning the educational effects, it means that VR educational games can provide highly motivating software that can achieve better educational effects than other kind of software. Thus, it is worth inviting the culture of such games in classrooms.

As a lesson from the evaluation analysis, the design of educational VR-games has to be such that the VR-environment is as sophisticated and attractive as possible but special care has to be taken about usability problems. Usability problems may cause distractions to student-players from the main educational goals of the games. If such problems are addressed, then the educational application can be more effective educationally for more groups of student-players (including novice players) than it would be otherwise.

References

Amory, A., Naicker, K., Vincent, J. & Claudia, A. (1998). "Computer Games as a Learning Resource". Proceedings of ED-MEDIA, ED-TELECOM 98, World Conference on Education Multimedia and Educational Telecommunications, Vol. 1, pp. 50-55.

Bruckman, A. & Brandlow, A. (2003) "HCI for kids" In: Jacko, J. & Sears, A. (Eds.), Human Computer Interaction Handbook, Lawrence Erlbaum, Hillsdale, NJ, pp. 428-440.

Chou, C. (1999) "Developing CLUE: A formative evaluation system for computer network learning courseware" In *Journal of Interactive Learning Research,* 10(2), pp. 179-193.

Conati, C. & Zhou, X. (2002) " Modeling students' emotions from cognitive appraisal in educational games". In S. A. Cerri, G. Gouarderes and F. Paraguacu (Eds.): *Intelligent Tutoring Systems 2002, LNCS, 2363*, pp. 944-954, Springer-verlag Berlin Heidelberg 2002.

De Vincente, A. & Pain, H. (2002) "Informing the Detection of the Students' Motivational State: An Empirical Study". In S. A. Cerri, G. Gouarderes and F. Paraguacu (Eds.): *Intelligent Tutoring Systems 2002, LNCS, 2363*, pp. 932-943, Springer-verlag Berlin Heidelberg 2002.

Druin, , A. (2002) "The role of children in the design of new technology" In *Behaviour and Information Technology,* 21(1), pp. 1-25.

Fabricatore, C., Nussbaum M. & Rosas R. (2002) "Playability in action videogames: A qualitative design model" *Human-Computer Interaction* 17(4), pp. 311-368, 2002.

Griffiths, M.D. & Hunt, N. (1995) "Computer game playing in adolescence: Prevalence and demographic indicators". In *Journal of Community and Applied Social Psychology*, **5**, pp. 189-193.

Hoysniemi, J., Hamalainen, P. & Turkki, L. (2003) "Using peer tutoring in evaluating the usability of a physically interactive game with children" In *Interacting with Computers* 15 (2003), pp. 203-225.

ID Software (1993), Doom - Virtual Reality Computer Game, Id Software Company Texas.

Jones, A., Scanlon, E., Tosunoglu, C., Morris, E., Ross, S., Butcher, P. & Greenberg, J. (1999). Contexts for Evaluating Educational Software. Interacting with Computers, 11(5), 499-516.

Kearney, P. (2004) "Engaging young minds – using computer game programming to enhance learning" In *Proceedings of ED-MEDIA World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004,* pp.3915-3920.

Markopoulos, P. & Bekker, M. (2003a) "Editorial: Interaction design and children" In *Interacting with Computers* 15 (2003), pp. 141-149.

Markopoulos, P. & Bekker, M. (2003b) "On the assessment of usability testing methods for children" In *Interacting with Computers* 15 (2003), pp. 227-243.

Matsubara, Y. & Nagamachi, Y. (1996) "Motivation System and Human Model for Intelligent Tutoring". In C. Frasson, G. Gauthier and A. Lesgold (Eds.) *Proceedings of the 3rd International Conference on ITSs,* pp. 139-147, Berlin 1996, Spriger-Verlag.

Mayer, R.E. (2001) "Multimedia Learning". New York: Cambridge University Press.

Mayes, J.T. & Fowler, C.J. (1999) "Learning technology and usability: A framework for understanding courseware" In *Interacting with Computers*, 11(5), pp.485-497.

Mumtaz, S. (2001) " Children's enjoyment and perception of computer use in the home and the school" In *Computers & Education*, 36 (2001), pp. 347-362.

Papert, S. (1980) "Mindstorms: Children, Computers and Powerful Ideas" Basic Books, New York.

Price, S., Rogers, Y., Scaife, M., Stanton, D. & Neale, H. (2003) "Using tangibles to promote novel forms of playful learning" In *Interacting with Computers* 15(2003), pp.169-185.

Squires, D. (1999) "Editorial. Usability and Educational Software Design: Special Issue of Interacting with Computers" In *Interacting with Computers*, 11(5), pp.463-466.

Squires, D. & Preece, J. (1996): "Usability and learning: Evaluating the potential of educational software" In *Computers & Education*, 27(1), pp.15-22.

Squires, D. & Preece, J. (1999): "Predicting quality in educational software: Evaluating for learning, usability and the synergy between them" In *Interacting with Computers*, 11(5), pp.467-483.

Sweeney, M., Maguire, M. & Shackel, B. (1993): "Evaluating user-computer interaction: A framework." In *International Journal of Man-Machine Studies,* 38, pp. 689-711.

Tsai, M. L. (2004) "Gender differences in online behaviors, motivation and attitudes" In *Proceedings of ED-MEDIA World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004,* pp. 4409- 4415.

Virvou, M., Manos, C., Katsionis, G. & Tourtoglou, K. (2002): "VR-ENGAGE: A Virtual Reality Educational Game that Incorporates Intelligence". In *Proceedings of IEEE International Conference on Advanced Learning Technologies* (2002), Kazan, Russia, September 16-19.

Virvou, M., Katsionis, G. &, Manos, C. (2005): "Combining Software Games with Education: Evaluation of its Educational Effectiveness", Educational Technology & Society, Journal of International Forum of Educational Technology & Society and IEEE Learning Technology Task Force , April 2005, Vol. 8, No 2.

Voelker, D.(2001): Statistics, Wiley Publishing, Inc. New York 2001.

Yacci, M., Haake, A. & Rozanski, E. (2004): "Operations and Strategy Learning in Edutainment Interfaces". In *Proceedings of ED-MEDIA World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004,* pp.4416- 4421.