Human-Computer Interaction - INTERACT'03
M. Rauterberg et al. (Eds.)
Published by IOS Press, (c) IFIP, 2003, pp. 705-708

Empirical evaluation of usability and fun in computer games for children

Wolmet Barendregt, Mathilde M. Bekker & Mathilde Speerstra

Faculty of Industrial Design, TU Eindhoven, P.O.Box 513, 5600 MB, Eindhoven, The Netherlands

W.Barendregt@tue.nl, M.M.Bekker@tue.nl

Abstract: This paper aims at identifying difficulties encountered by evaluators when they conduct a formative evaluation of both usability and fun in computer games for children. Towards this end, two experimental studies were conducted. The first study suggests negative effects of the use of tasks for fun assessment in computer games. The second study shows difficulties in determining with which set of heuristics (usability vs. fun heuristics) identified problems should be explained. Based on the findings of the second study a procedure is suggested to support evaluators in determining the right set of heuristics for each problem.

Keywords: Formative evaluation, observation, usability, fun, computer games, children

1 Introduction

In the last few years, researchers and developers of technology have started paying special attention to children as users of technology (Druin, 1999; Bekker et al, 2002). Children have different abilities and interests than adults and therefore it is also important to think about how to design and evaluate products for this user group. In this paper we will discuss some methods for the formative evaluation of educational computer games for children through observation.

2 Quality of computer games

Usability is seen as one of the major criteria to ensure the quality of work related products. The usual attributes of usability are effectiveness, efficiency and satisfaction (ISO, 1998). Satisfaction in this context refers to the satisfaction of the user regarding the efficiency and effectiveness of the product. Additionally, fun should also play a role in the evaluation of computer games for children because enjoyment is one of the major motivations for children to interact with technology (Inkpen, 1997). However, it is not sufficient to look for fun alone. Pagulayan et al (2003) wrote 'The ease of use of a game's controls and interface is closely related to fun ratings for that game. Think of this factor as the gatekeeper on the fun of a game'. Thus, the quality of a computer game depends on both usability and fun. Furthermore, when testing either fun or usability it is very likely that problems of the other type will be encountered as well. As a consequence the aim of a formative evaluation procedure is to create a list of both fun and usability problems and causes to be fixed by the developers. The global procedure for a formative evaluation through observation consists of two parts: data gathering and data analysis. This paper describes two studies exploring some issues concerning this procedure when conducting an evaluation of computer games for children. The first study focused on the common practice of giving tasks during data gathering. The next study examined how both usability and fun heuristics may support evaluators in creating a list of problems and problem causes during data analysis.

3 Tasks during game evaluation

3.1 Introduction

The use of tasks during the evaluation of work related products is common practice and certainly has some advantages. By giving tasks the evaluator tries to ensure that all relevant parts of the software are covered and reactions of participants can be

compared because they have all done similar activities. However, in the context of work related products research has shown that the order, wording and type of tasks influences users' behaviour (Vermeeren, 1999). For games, there are reasons to assume that giving tasks might influence the number and kinds of fun problems found. The major reason comes from the work of Malone and Lepper (Malone & Lepper, 1987). They defined the important heuristic 'Challenge' for good game design. According to this heuristic, providing good goals is a way to create challenge. Suppose we define tasks for the evaluation of a game. The tasks themselves become goals and depending on the user they will compete with goals that are provided by the game. We may see frustration when tasks are too difficult or boredom when too easy and we may see deviations of the tasks when the goals of the game are more attractive than the tasks. The next section describes a study that examines the influence of tasks on the behaviour of children during the evaluation of a game.

3.2 Influence of tasks

This study examines the influence of tasks on the behaviour of children playing a game. We defined three measures to determine this influence: number of screens visited, number of positive and negative verbal and non-verbal indications and the number and kinds of problems found in each condition.

Because in contrast to work related products it is difficult to do a task analysis of games, we first prepared a pilot study to determine realistic tasks for the educational computer game 'Oscar de ballonvaarder en de geheimen van het bos' (Oscar the balloonist and the secrets of the forest). The tasks were determined by observing four children play the game without giving tasks. After this pilot study we set up a study with seventeen children of 8 and 9 years old. Half of the group performed the tasks (8 children) and the other half (9 children) could play the game as they liked for fifteen minutes. All sessions were videotaped.

Although we did not find any difference in verbal and non-verbal indications or number and kind of problems, we did find a significant difference in the number of visited screens in each condition. Children in the task condition visited more screens than children in the free play condition (T (13) = -2.43; p=0.03) (see Table 1).

Tasks		Free play		Totals	
Avg.	SD	Avg.	SD	Avg.	SD
23.4	7.3	12.6	4.1	17.7	7.9

Table 1: *Average number of visited screens and standard deviation in the task condition, the free play condition and overall.*

An explanation for this difference is that the children in the task condition just seemed to browse a screen to see if they could fulfil the given task with this screen. If the screen did not seem to be the right one they quickly tried another one. Children in the free play condition did not have this task so they could take more time to look for interesting things on each screen. The major problem with giving tasks for the evaluation of computer games is that the goals set by the tasks interfere with the intended goals of the game. So even though we explicitly tried to create realistic tasks that matched the goals of the game by doing a pilot study, the tasks still influenced the children's behaviour. Of course, the aim is to evaluate the fun of the game not the fun of our tasks, so giving tasks to evaluate fun is something one should be very careful with. The fact that we did not find a difference in the number or kinds of problems strengthens our belief that it is not absolutely necessary to give tasks to find all relevant problems in an educational computer game.

4 Usability and fun evaluation

4.1 Analysis of video data

A common way to analyse the data gathered in an empirical evaluation is to review videotapes of the test sessions. The evaluators create a list of problems and problem causes based on verbal and non-verbal behaviour of the user. Research has shown that there can be a lot of disagreement between different evaluators because of vague evaluation procedures and problem criteria (Hertzum & Jacobsen, 2001). One way to help the evaluators to decide on a cause for an observed problem and make the analysis more reliable is to provide heuristics or guidelines like Nielsen's (Nielsen, 1994). For example, when a user is confused about which button to choose, a heuristic like 'Use recognition rather than recall' can help the evaluator to decide that the cause may be that the buttons lack labels and that by using only colour o

convey their meaning people are forced to recall what colour means.

When analysing videotapes of children playing a game the evaluator will probably not only see usability problems that can be mapped to usability heuristics but also fun problems. One obvious way to try to help the evaluators do the analysis is to provide them with an additional set of fun heuristics, like that of Malone and Lepper (Malone & Lepper, 1987). Their set of fun heuristics includes Challenge, Curiosity, Control, and Fantasy. Research has shown that this set of heuristics provides a promising start to analyse the causes of fun and frustration in computer games (Kersten-Tsikalkina & Bekker, 2001).

The next section describes a study into the use of a combination of these usability and fun heuristics.

4.2 Combining usability and fun

We set up a study to determine whether evaluators can find both fun and usability problems with the help of two sets of heuristics, we set up a study. In this study four evaluators had to evaluate four videotaped evaluation sessions. Each session contained two children in the age group 7-9 years old playing the same game that was used in the first study. The children had been asked to think aloud while performing some tasks. The evaluators first received a training session on how to apply the heuristics to evaluate fun and usability of computer games. Subsequently the evaluators received a copy of the game and the tasks so they could get familiar with them before doing the analysis. During the actual analysis the evaluators had to fill in a template for each problem they found based on the children's verbal and non-verbal behaviour. For the description of the cause(s) the evaluators got two sets of heuristics: (Nielsen, 1994; Malone & Lepper, 1987).

The results showed that the evaluators had problems deciding which set of heuristics to apply. The evaluators found 31 problems altogether. For 26 of the 31 problems (84%) both usability and fun heuristics were mentioned to explain the occurrence of the problem; either one evaluator mentioned heuristics from two sets, or different evaluators mentioned heuristics from different sets.

4.3 Discussion

We tried to find an explanation for the problems the evaluators had in deciding on the set of heuristics. One of the reasons could be that the evaluators, although experienced in usability testing, did not have enough experience in applying both fun and usability heuristics. We tried to solve this by giving all evaluators the same training session, but apparently this did not help. Afterwards, most evaluators complained about the difficulty in choosing the right heuristic. For example, when a child could not find the game that was supposed to be played the evaluators had difficulty in choosing between a violation of the fun heuristic 'challenge' or a violation of the usability heuristic 'recognition rather than recall'. Because of these comments we took a closer look at both sets of heuristics. It became clear that the sets of heuristics sometimes seem to contradict each other. Although in a work related product challenge is usually an unwanted feature because it may decrease efficiency, it can be a desirable feature of a game. Still the quality of a game depends on both usability and fun, and games can have both usability and fun problems. We think a procedure to distinguish usability and fun problems before choosing the right heuristic might help. The next section describes a proposal for such a procedure.

5 Distinguishing usability from fun problems

Based on our own experiences in testing with parents and children we came up with a procedure for evaluators to distinguish usability from fun problems. This procedure is depicted in Figure 1.

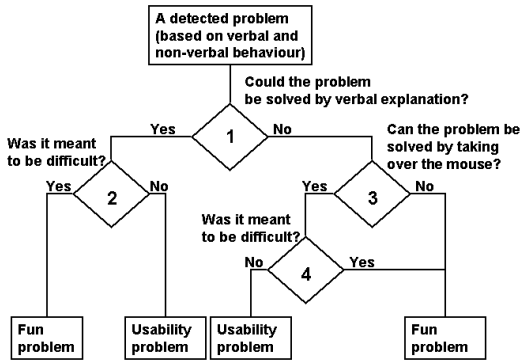

Figure 1: Procedure to distinguish usability problems from fun problems

Suppose the evaluator has decided that there is a problem based on verbal and non-verbal behaviour of the user. To determine the appropriate set of heuristics the evaluator has to answer the following questions (numbers refer to the decisions in Figure 1):

1. Could the problem be solved by verbal explanation? If 'yes' it is probably a usability problem, if 'no' it is probably a fun problem. *Example*: the user wants to go to another part of the game but does not know how to do it. The evaluator can explain by saying: "It's the purple button in the lower right of the screen".

2. Was it meant to be difficult? If 'yes' it is a fun problem related to high challenge, if 'no' it is a usability problem. *Example*: the user cannot find a key to open a door and wants to give up. The evaluator could tell where it is, but it was meant to be difficult, therefore it is a fun problem.
3. Can taking over the mouse solve the problem? If 'yes' it concerns a motor skill problem that can either be a fun or usability problem, if 'no' it is a fun problem. *Example*: a child wanting to stop the game because of the scary characters has a fun problem because neither explaining nor taking over the mouse can solve it.
4. Was it meant to be difficult? If 'yes' this is a fun problem because the challenge is too high, if 'no' it is a usability problem. *Example*: shooting a spaceship while avoiding the rockets aimed at you is meant to be difficult, but is can be too difficult.

We assume that this procedure can help evaluators determine which set of heuristics to use to find a cause for the problem. We plan to set up a study to explore the practical use of this procedure.

6 Suggested evaluation procedure

We suggest a practical method for the formative evaluation of fun and usability in computer games for children. To make sure that usability problems do not completely spoil the fun the facilitator helps to overcome problems when the child asks for it. Every time the facilitator had to help is noted as a problem area for later analysis. Because we have evidence that parents often help their children to play a game for the first time it seems a very natural approach to have a helping facilitator in the room with them.

Furthermore, the children are allowed free play, but to ensure that each important part of the game is evaluated the children can be directed to certain parts of the game where they are allowed free play in combination with some free exploration of the game. This is especially useful when the game consists of several independent parts, which is common in many educational games for young children.

Finally, during analysis the procedure to distinguish fun and usability problems enables the evaluator to classify all encountered problems and use the appropriate heuristics to find causes.

Although we conducted our studies with young children playing an educational game, we think many of our suggestions can also be applied for the evaluation of games for adults.

Acknowledgements

We thank Ramona, Jan, and Lineke for their research on giving tasks to evaluate fun in computer games.

References

Bekker, M.M., Markopoulos, P., Kersten-Tsikalkina, M. (2002). *Proc. of the Conf. on Interaction Design and Children IDC2002*, Eindhoven, August 2002.

Druin A. (1999). *The Design of Children's Technology*. Morgan Kaufmann Publishers, San Francisco CA, 1999.

Hertzum, M., Jacobsen, N.E. (2001) The Evaluator: A Chilling Fact About Usability Evaluation Methods. *Intern. Journal of Human-Computer Interaction*, 13(4), 421-443.

Inkpen, K. (1997) Three Important Research Agendas for Educational Multimedia: Learning, Children, and Gender, *Proc. of Educational MultiMedia '97* Calgary, AB, June 1997, pp. 521-526.

ISO (1998) *Ergonomic requirements for office work with visual display terminals (VDTs) Part 11. Guidance on usability* (ISO No. 9241-11). Geneva: International Organization for Standardization

Kersten-Tsikalkina, M., Bekker, M.M. (2001) Evaluating Usability and Fun of Children's Products, *Proc. of The International Conf. on Affective Human Factors Design*, Asean Academic Press, London, 2001.

Malone, T.W., Lepper, M.R. (1987) Making learning fun: a taxonomy of intrinsic motivations for learning. In: Snow, R.E. and Farr, M.J. (Eds*.) Aptitude, learning and interaction III. Conative and affective process analysis*. Hillsdale, N.J.: Erlbaum.

Nielsen, J. (1994) Heuristic evaluation. In Nielsen, J., and Mack, R.L. (Eds.), *Usability Inspection Methods*. John Wiley & Sons, New York, NY.

Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R., & Fuller, T. (2003). User-centered design in games. In J. Jacko and A. Sears (Eds.), *Handbook for Human-Computer Interaction in Interactive Systems* (pp. 883-906). Mahwah, NJ: Lawrence Erlbaum Associates.

Vermeeren, A.P.O.S. (1999) Designing Scenarios and Tasks for User Trials of Home Electronic Devices. In *Human Factors in Product Design, current practice and future trends*. Eds. Green W.S. and Jordan, P.W. pp. 47-55, 1999. Francis & Taylor, London.