Main assignment for INF2260, Fall 2012

Project report for Child's Play Unit (CPU)

by

Arild Birkeland, Martin W. M. Evensen, Emil Lie Hatlelid and Persijn Kwekkeboom

University of Oslo

November 1, 2012

Innhold

1. Delivery

1.1 Summary of report

Chapter 1 - Delivery, focuses on the background and basis for this project. It shows what the project is about, the goals we have set for our project and the methods used to collect data.
Chapter 2 - Empirical data, gives an overall impression and the basis for the data that we have collected.
Chapter 3 – Analysis, gives an in depth analysis of the data presented in chapter 2. We try to answer the main questions that we present in chapter one.
Chapter 4 – Evaluation, we discuss in debts the methods that we have used to gather data and evaluate whether or not the methods used were suitable for the data-gathering that we required.

1.2 Description of the group

Our group consists of four members; Arild Birkeland, Martin W. M. Evensen, Emil Lie Hatlelid and Persijn Kwekkeboom. Arild has a background in computers and law at the University of Oslo. Martin have previously worked with children at kindergardens and after school programs. Emil and Persijn are sentral actors at the room for experimental informatics at IFI (Åpen sone for eksperimentell informatikk) and have previous experience with game based programming. We are all bachelor students of Design, use and interaction at the Department of Informatics at the University of Oslo, and we are all interested in computer games.

As the basis for this report, we have created a game that's interactive, using Microsoft Kinect. The game is developed for Oslo Children's Museum. We have presented our prototype as an oral presentation. In this report we will focus mainly on the data collection and analysis of our project.

1.3 Purpose of project

In this project we wanted to create a game that would both be fun and stimulate learning. Our goal was to create an interactive experience where children can play, be creative and engage children to learn more about the world around them.

1.3.1 Inspiration:

While we were initially exited about the prospect of designing an interactive experience for children, a contributing event was the trip to video game designers SnowCastle^{tm} and their offices where Martin together with Alma was introduced to their vision about how to design

video games for children. The most impressive part of the presentation was not their actual games themselves, but their business model/design model about Game-Based Learning, which focused on making entertaining games first, and educational games second.

1.3.2 SnowCastle Business Model[1]:

SnowCastle representative Erik Hoftun talked about the "tran med sjokoladetrekk" (En. "chocolate covered fish oil") that educational games tend to be. The basic approach seems to have been to decide on what exactly the children are supposed to learn, cram as much of it as possible into a product, and then try to make a game of it. Our experience has been that these kind of games can be boring since the only intention of the game is to teach, despite being called a game. The word "game" will make most people think of fun. Therefore is fun a crucial part of a game.

1.3.3 Entertainment vs education

While entertainment value is not directly listed in the design brief from Oslo Childrens' Museum, the intent from the customer is clear since it focuses on the importance of play with quotes like this:

> "*Play is the beginning of knowledge*"
>
> -- George Dorsey, American Anthropologist 1868 - 1931

This, combined with our playful nature, experience with video games, as well as inspiration from the Snowcastle presentation with Alma, made us want to focus on entertainment first and educational value second.

1.4 Design brief

We wanted to create an interactive experience where children can play, be creative and engage children to learn more about the world around them. We could freely choose our topic, but we had to use a Kinect-device and it had to fit the requirements of the Oslo Childrens' Museum. We had regular meetings with Katie, the representative from the museum. We presented our vision to her early in the semester. We got positive feedback, and we have gotten continuous positive feedback from Katie, especially during our interactions with the children at Uranienborg school.

[1] SnowCastle GBL workshop (see appendix).

1.5 Project plan

Week	Plan:	Deadlines:
Week 38	(17-23.09.2012): Low-fidelity prototype, user testing and evaluation.	
Week 39	(24-30.09.2012):	
Week 40	(01-07.10.2012): Work on high-fidelity prototype and user testing.	
Week 41	(08-14.10.2012):	
Week 42	(22.10.2012): Presentation of a working prototype,	Deadline, working hi-fidelity prototype.
Week 43	(23-28.10.2012): Finish our high fidelity-prototype.	
Week 44	(29-04.11.2012): Last user testing.	
Week 45	(05-11.11.2012): Focus on report writing.	
Week 46	(12-18.11.2012):	
Week 47	(19-25.11.2012):	
Week 48	(26.11.2012): Finished report!	Deadline, finished report.

The group has maintained this timeline throughout the project. We have had regular meetings at least two times per week, debating the problems of the week, and worked individually throughout.

1.6 Define the problem space

At the start of this project we already had several requirements relating to our prototype. We knew that the Oslo Children's Museum wanted something that used Kinect and that the user group were children in the age of 6 to 10 years old. The requirements were: No language needed, no queues, engage multiple senses, bring strangers together, fit child-size proportions, easy transportation and low cost.

We wanted to define tree central stages, what is the initial state that we are in, what is the goal of the end of this project, and what are the "...possible intermediate states that must be searched in order to link up the beginning and the end of the task."[2] In this part of the paper we will discuss the problems we have encountered and the solutions to these problems.

1.6.1 The initial state

The main question is: "what makes a good game for children?". Finding out what makes a good game means that we needed to design a concept of what we wanted to make, identify the solutions to the constraints mentioned above and find out how children can learn from interacting with a game.

1.6.2 Intermediate states

The first problem was deciding who will develop the concept of the game, the group or children? Given more time, the group would want to have a closer design process with the children, but

[2]"Problem space" Accessed November 15, 2012. http://www.answers.com/topic/problem-space

that would take too long time compared to the deadlines that we were given. This meant that we had to develop the concept of the game, and then get feedback from the children.

The next question is: “who can identify and solve the problems?”. We have encountered problems during the design of our project, for example finding out how hard it should be to win the game. We had already introduced garbage-objects in the game, so the question then becomes finding out how many garbage-objects should appear in the game, to make the game harder or easier. Another problem was how the fish should look like in the game. Here we thought that the children would be more creative than us. We conducted a workshop with children at Uranienborg SFO (an after school facility), were the children drew several of the fishes that are in the game. When solving design issues, one should consider who is the domain expert.[3] When deciding who could identify how hard the game ought to be, we were sceptical to whether the children were able to differentiate between the subtle changes in the game that could make the game easier or harder. We decided that our group members were the closest experts. Although, when it came to identifying the best fishes in the game, we decided that the children were the experts.

The next question is; who can evaluate the solutions? Only the children can evaluate the concept and the solutions of our game. The group can make assumptions about the evaluations, and set the conditions for were and how the children can evaluate the game. Our main concerns here are that we focus on evaluating the central parts of the prototype and that we minimize the chance for bias. We will discuss our evaluations in chapter 2 of this paper.

We also wanted to incorporate learning into our game. This led us to three central questions; “What does children in the age group know”, “who can identify the learning”, and “who can evaluate the learning?”. The group had some trouble identifying what we could teach children with our game and within our timeline. We wanted to inspire learning about the ocean, but still keep the game fun for other children outside of our age group. The game can only stimulate to learning. The children are the only ones who can tell us if they learned anything, which points to that they are the domain expert, but since the age group are so young it is questionable whether they realise there own learning. It is through our interviews with the children that we have identified whether the children have learned anything or not. We therefore see that it takes a collaboration between the group and the children to identify the learning. When it comes to evaluating the learning, it is clearly the group, since that is the main part of this paper.

1.6.3 Goal state

[3]Subject-matter expert. Last modified 5 October, 2012. Accessed November 15, 2012. http://en.wikipedia.org/wiki/Subject-matter_expert

At the start of this project we created this goal state: "*When we are finished with this project we have created a high-fidelity prototype of a game that fulfills all of the requirements from the Childrens' Museum. The game is fun and is educational for the children in our age-group, but can also be played by all ages.*"

1.7 Methods for data collection

1.7.1 Interview, observation and triangulation

The group has used multiple research methods in this project in order to triangulate and "*...map out, or explain more fully, the richness and complexity of human behavior by studying it from more than one standpoint*."[4] The group had two interactions with the children at Uranienborg. The first time we conducted a workshop (formative usability test) and the second time we conducted a summative usability testing session. We presented the results from the workshop during our oral presentation. We will therefore only focus on the data from the summative usability-session in the next chapters.

We structured the usability-session to include a formal part, the interview, and an informal part, playing the game. We wanted to minimize the formal parts of the workshop, in order to establish trust between us and the children.[5] By introducing a more playful environment we believed that we would get more honest answers from the children.

Interviews and observations were conducted in both the sessions. Interviews with children can be more challenging than adults. A child's ability to comprehend height, weight, age, time or there ability to recall experiences may be more challenging than for an adult. This means that the group had to create questions that did not collide with these constraints. By compensating for the constraints in the interviews we made it a priority to triangulate with both interviews and observations. This can for example be seen when we both asked the children "What do you think about the game?" and count how many times the children asked to play the game one more time.

1.7.2 Qualitative data

The group has conducted in depth observations and interviews during this project. We have used open ended questions during the interviews and asked for their opinions. We have had observations of behavior conducted in depth recordings of behavior. The combination of these creates a rich picture of the children who partook in the usability testing. The data collected is therefore qualitative data.[6]

[4]How to combine multiple research methods: Practical Triangulation. Last modified August 20, 2009. Accessed November 15, 2012. http://johnnyholland.org/2009/08/practical-triangulation/

[5] Sosiologisk blikk på brukerorientert design, page 10.

[6] Preece. *Interaction Design: Beyond Human-Computer Interaction*, page 272.

1.7.3 Usability Testing

The goal of usability testing is "(...)to improve the quality of an interface by finding flaws in it."[7] It can be used on low-fidelity paper prototypes to already implemented software, but the focus is on the interface itself, and the pros and cons of it. For example, how a menu works, how visible it is, easy to navigate, understandable for the user, are all factors, while the color scheme or music is not.

With this in mind, traditional usability testing may not be appropriate for a computer game at all, or at the very least, is insufficient. On the other hand you can make a computer game that runs well, is easy to understand and operate, fast and efficient. And it may still be the most boring game ever conceived. Games are generally intended to entertain and please the player(s). Exceptions may be in professional sports and intellectual games. Attempts have been made of making videogames where education value has been the priority, while entertainment factor is added as an afterthought. The result is often a game that is neither very entertaining nor very educational, as a game that is not played because it's boring, by extension fails to educate. Thus using usability testing with it's traditional focus on interface, efficiency, error seeking will not by itself tell us if this is a good game or not.

Usability testing has its main focus on humans. Traditional usability testing focuses on the smoothness and speed of action. These ultimately are for the comfort of the user and improving the product. When applied to a web-page the purpose of it is that users have a pleasant experience using that web-page. The speed of actions and the efficiency of the web page are what gives a pleasant experience of the web-page. When the testing is used for games the speed is important but the purpose of the game is not to finish it as fast as possible. Since usability testing is used to find spots where a product can be improved it is easily used for a web-page or a game.

We have employed many elements of Usability testing, because it will give some results about our game. In our usability test we have followed this plan:

1. Develop Test Plan
2. Set up test environment
3. Find and select participant
4. Prepare test materials
5. Conduct test sessions
6. Debrief participants
7. Analyze data and observations
8. Report findings and recommendations

[7] 2010, Lazar et al *Research Methods in Human-Computer Interaction*, p. 252

(Rubin and Chisnell, 2008)[8]

The systematic approach in Usability Testing lends itself well to our goals; Testing the requirements of the game, especially how many can play it, if it is intuitive enough, the level of enjoyment from the target users, and testing how our users interact with the prototype. Thus can we evaluate the prototype and the project that led to it, uncover flaws that could be fixed in later prototypes, or how to expand the game. Although not directly a part of usability testing, we were also investigating the user experience in the test session, mainly if the children thought the game fun. More on this in chapter 3.

2. Empirical data

The central part of this chapter is to use the data that we have collected "...to gain an overall impression of the data and to start looking for patterns."[9] In order to establish the overall impression we needed to find out:

1. How the children interact with the game with no information beforehand
2. Whether the children cooperate or simply compete in multiplayer
3. Any observable differences in how the different age groups and genders react to the game
4. Further validation of our concept
5. If the children can learn from our game , or be inspired to learn about the sea.
6. If the game (and Kinect interface) can handle more than one player

Who were the participants?

We wanted participants from 1st, 2nd and 3rd grade, and we wanted both males and females. Thus we made a table that would give us the correct amount of children, with specified gender and age. For example, one group would consist of a pair of girls, both first graders. This was repeated for both genders and all 3 grades (total of 12 children). After that we had 3 further pairs, with mixed gender (for all 3 grades). This would have given us a total of 18 children. However, we did not have time to conduct the last test with the 1st graders (one girl+one boy). Despite this we now had data from 16 children, which is more than needed for usability testing[10].

The participants were divided into three age groups according to their grade; first, second and third grade. There were two first graders and three second and third graders, a total of 16 children.

[8] 2010, Lazar et al *Research Methods in Human-Computer Interaction*, p. 262
[9] Preece. *Interaction Design: Beyond Human-Computer Interaction*, page 285.
[10] 2010, Lazar et al *Research Methods in Human-Computer Interaction*, p. 263

Table of participants:

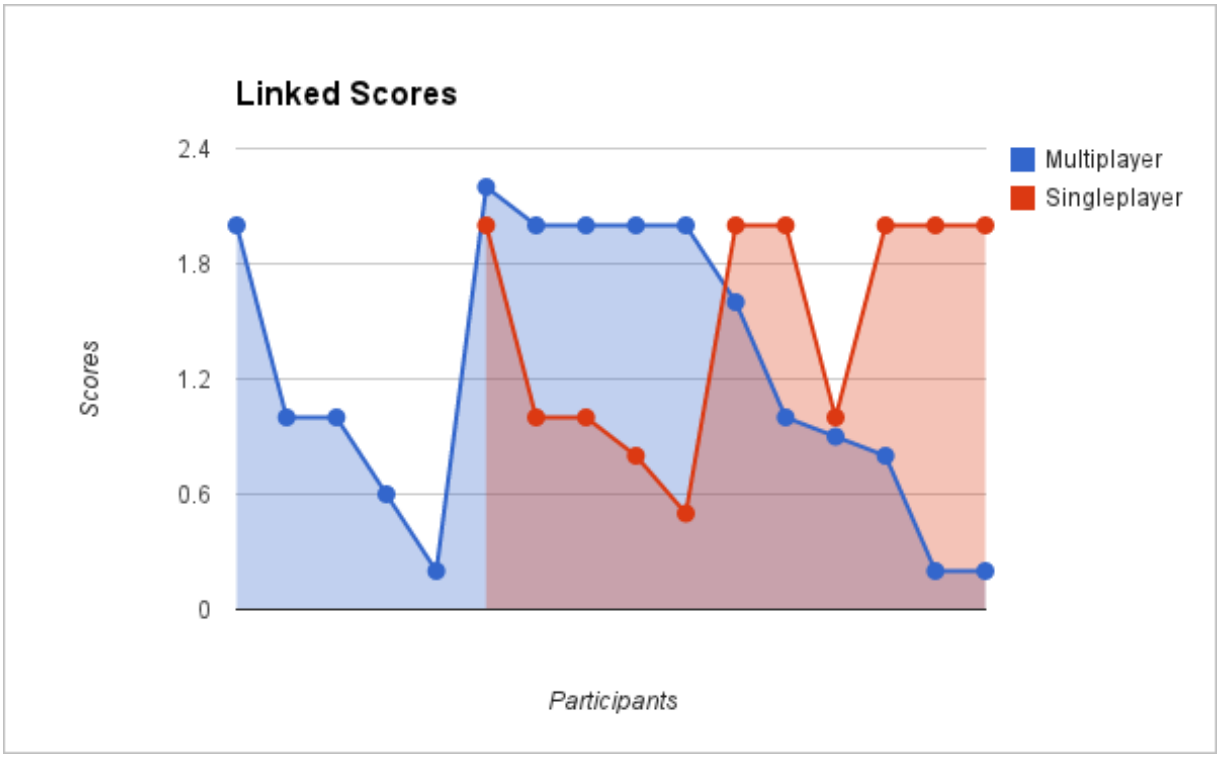
Groupnumber:	**Total gametime:**	**Players:**
1	10 min	B1 + B1[11]
2	10 min	G1 + G1
3	10 min	B2 + B2
4	10 min	G2 + G2
5	10 min	B3 + B3
6	10 min	G3 + G3
7	5 min	G1 + B1
8	5 min	G2 + B2
9	5 min	G3 + B3

2.1 Observation

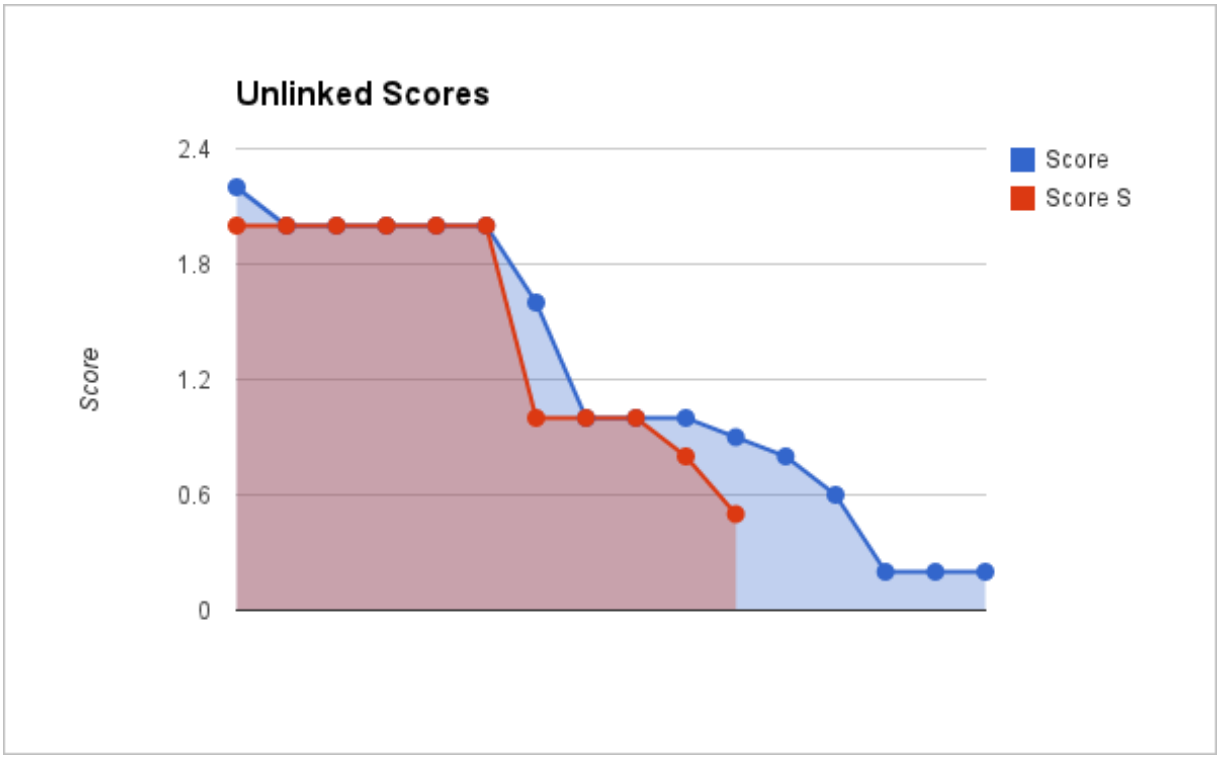
2.1.1 Player Scores

Player score is determined by how fishes that the player has “eaten” in the game. Each fish earns the player 0.1 score, if the player hits a hostile fish the player is reduced to the minimum score of 0.2. The game is completed when a player reaches a score of 2.0 and manages to eat the boss (shark).

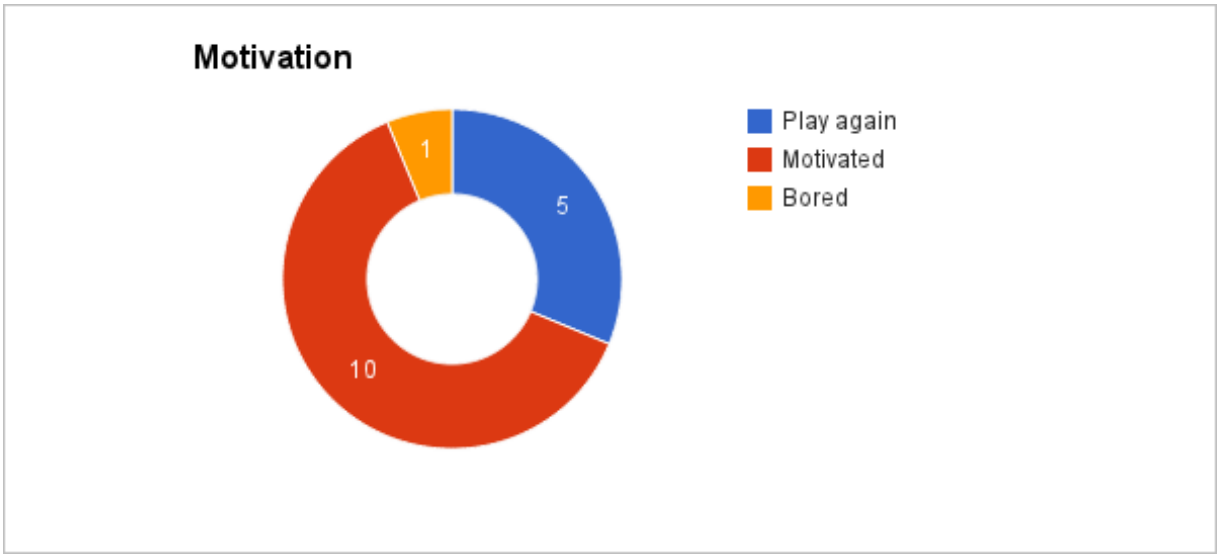
[11] B1 + B1 means boy from the first grade and another boy from the first grade.
[11]G1 + G1 means girl from the first grade and another girl from the first grade.



The chart "Linked Scores" describes how the participants scored, sorted by their multiplayer score, decending from high to low. The first 5 participants on the chart did not play the game alone and lacks singleplayer scores. These scores show that most participants (11/16) won at least one game, whether playing alone or with another participant.

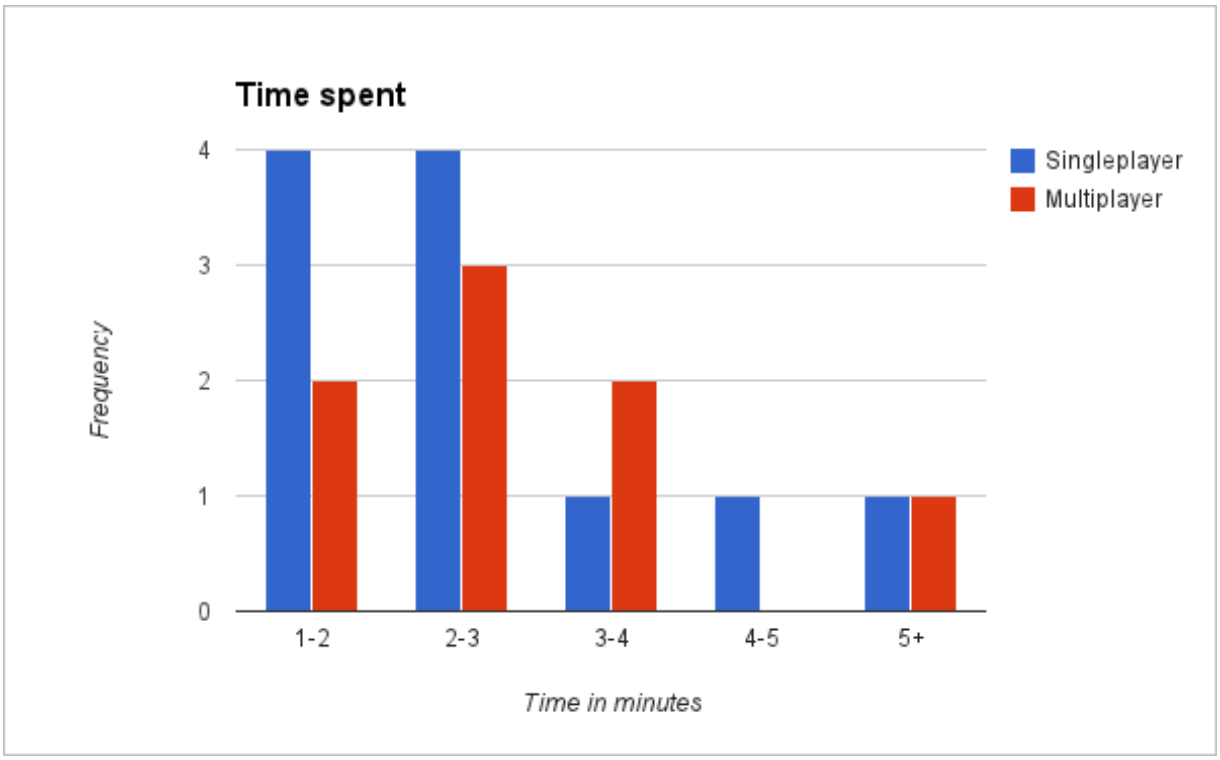


The "Unlinked Scores" chart describes all scores, descending from high to low.



Out of the 16 participants there was only one who did not seem to enjoy the experience. First she played the game with another participant, but when we asked her to play alone she did not want to play. She seemed afraid and we did not want to scare her so she was let out of the test chamber before asked what she was afraid of. Five of the remaining 15 participants asked to play the game again.

2.1.2 Game Lengths



The average time for a solo playthrough were about 2:44 min, the average multiplayer was 2:39 min.

2.1.3 Cooperation

Most of the participants competed in order to be the one to win the game, but often discussed how they should play the game when they played together. The participants who talked less would instead try to mimic what the other was doing. In some cases this turned out to beneficial as the other would have figured out the proper interaction with the game. In most cases this would lead to one trying to understand the controls by doing a gesture and the other mimicking the gesture and then they would accept this as the correct form of interaction. This lead to the errors regarding the game-controls and interaction.

2.1.4 Errors

The errors that the participants made while playing the game was mostly while exploring how the game was played. There were in all 8 different errors while trying to figure out the controls in the game. The game is controlled by moving your right hand, the participants tried; pointing with their right hand (2), using the left hand (2) or using their entire bodies; leaning and jumping (2) or sidestepping and jumping (2). There were also 5 participants that had some initial trouble understanding the objective in the game. Of the 5 participants only two seemed to continue having trouble understanding the concept and we decided to hint and explain some parts of the game.

2.2 Interviews

After conducting the usability tests participants were interviewed in pairs. The first few questions were identical to the control questions asked before the testing took place. This would give the children a chance to change their previous answer in response to what they had seen in the prototype.

While most groups changed their answers somewhat after playing, some added information unrelated to the game, as they recalled more of their previous knowledge. A few answers were even forgotten the second time around. To the question about what a fish should try to avoid, only a single group answered “trash” after playing the game.

As for the questions regarding the game itself, the children were more unified in their responses. All the groups said they thought the game was fun, out of 16 children only one of them (a 1st grader girl) didn’t want to play a second time, and was visibly intimidated by the whole process. We can’t be sure if it was the game itself that was the cause of anxiety, or the presence of four men with no familiar adults around for comfort. We did not ask the children to rate the game, as we assumed that children of these ages don’t have the capacity to rate the game on a numerical

scale we can compare. Adjectives used by the players were “good”, “very good”, “cool”, “nice.” The children seemed to enjoy themselves, and several were upbeat and enthusiastic after playing. Despite this, many if not most of the children said that the game was difficult to play.

Most of the children achieved a certain understanding of how to play the game after trying it once. The oldest of the children specifically mentioned that it was important to avoid sinking garbage, and almost every pair understood that they had to “eat” the other fishes in order to win.

We wished for suggestions from the children on how to improve the game. Here are their suggestions:

-more (different) “fishies”
-easier to see that the fish are eaten
-less garbage falling down
-”when is the commercial coming?”
-implementing levels
-the shark should be the toughest fish
-fishes in different sizes, and increasing ability to eat larger fishes the bigger you get
-fishes need to be more scary
-game should be easy to play for left-handed
-controller

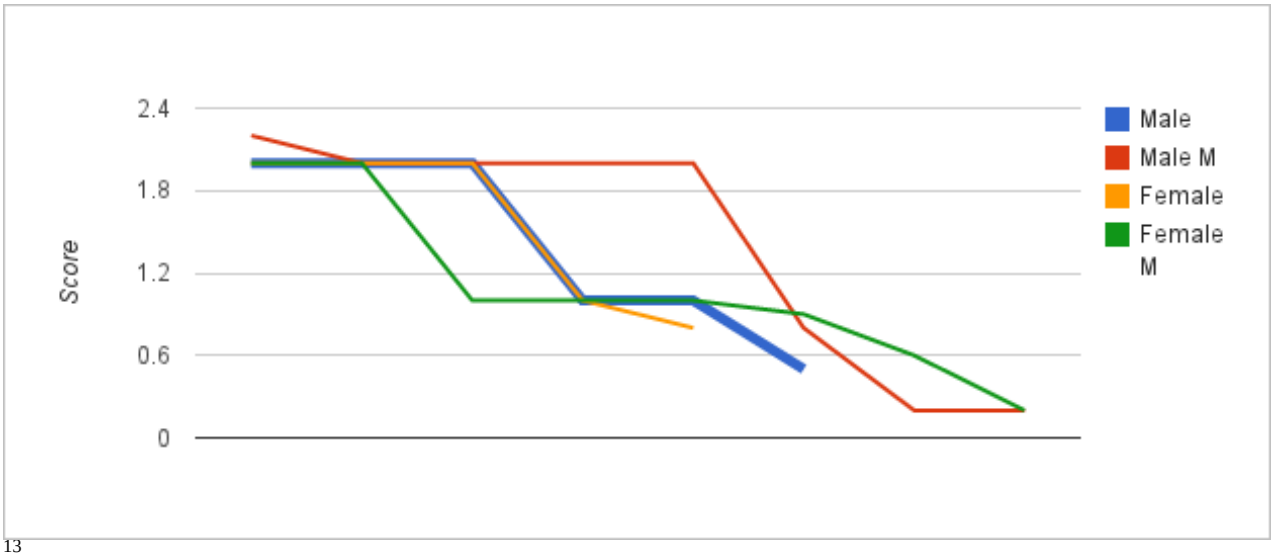
3. Analysis of data

Analysis of data means to do a “...more detailed work using structured fremeworks or theories to support the investigation.”[12] We have in the previous chapter presented the significant parts of the dataset. The next thing we need to do is to see if the data that we have collected can be used to answer the central questions to our project.

3.1 Gender and age differences

Gender

[12] Preece. *Interaction Design: Beyond Human-Computer Interaction*, page 269.



[13]

The score distribution is very much the same, especially in singleplayer mode. There is a difference between male and female scores in multiplayer, but this is too small scale to determine if this is accurate. Also in the two groups where there were one of each gender; the female participants won.

3.2 Childrens' Museum's Requirements

The game can not present any text. This gave us a challenge; will the children be able to walk up to the game and understand how the game is played, without any indication of how it will be played? We told the children, before they entered the play-room, that they will be playing a fish-game, but that we would not tell them was how the game was played. Telling the children what they were going to be playing was necessary in order to get them to participate in the testing. By not telling the children how to play the game we got the opportunity to see if they could learn the game for themselves. We told the child to stand in front of the camera and behind a line, before the child was going to play the game. This shows us that we still lack a tested scenario were the child enters a empty room and can start playing the game. Given this minimal information, the children did not have much problem understanding the game.

A central part of our user testing was to make sure that both children who knew each other and children who did not know each other could play equally well together. This is one of the reasons why we chose to have pairs consisting of only boys and only girls, but also boy and girl. We assumed that if we brought inn two boys or two girls from the same age, that they were friends, but that the link would not be that strong when bringing in a boy and a girl. In the latter case, the boy and girl may know each other, but not be as close as for instance two boys. When comparing these two groups we observed that both groups preferred to play multiplayer. The group were initially worried that the game did not have a strong enough cooperation element, but we still

[13]The scores on the chart are split into groups depending on gender and gamemode (multiplayer or singleplayer)

wanted to create a game that would be enjoyable for single player. Since we see that almost all of the groups preferred multiplayer we believe that we have created a platform that can bring strangers together in meaningful interaction.

Kinect can only track two people at once. This meant that we needed to see how long it took for a child to complete the game. The average game-time for multiplayer was 2 and a half minute. This time period is in itself short enough to prevent much queuing. If someone wants to play the game twice or need a longer time, which we also saw, there might be some queue. The game also has the ability to change players during a game. This means that if a queue has formed, one child can leave the game so that another can enter. The children do not care much about actually winning the game. The interviews show us that the children describe the game equally fun, whether they won the game or not. We wanted from the start to create a game that would be enjoyable to play, not to just win. We therefore believe that the game will not create a queuing problem as long as the children are able to share.

The game combines the use of perception and movement with the entire body. The game uses large images that are easy to see, even if a child has poor eyesight. When it comes to movement of the body, the game offers usage for many types of children, even if they have handicaps. We can see from the interviews that even though some of the children felt that the game was hard, they still enjoyed playing. We therefore believe that even if a child has the same fine motor skills in the arms, the child will still enjoy the game on an equal ground as children in our usability test.

3.2.1 Fun

By fun we mean "... the enjoyment of pleasure and...an absolutely primary category of life...".[14] We have therefore tried to measure indication that the children received pleasure from interacting with our game.

The question of whether or not a game is fun can not just be measured by just looking at children playing the game, but also the context of where the game is placed. So the real question therefore becomes; was the game fun for the children in the context of being set up in an after school program, and can we extrapolate from this how the game will work when placed in the Oslo Children's Museum?

We see from the observations that all except for one of the children finished the game time they were allowed to play, playing both multiplayer and single player. The interpretation from the group members that observed the children were that the children were having fun. 7 out of 8 groups said that the game was fun when interviewed, and the last group was not interviewed at

[14]Fun. Last modified November 22, 2012. Accessed November 25, 2012. http://en.wikipedia.org/wiki/Fun

all. This can give an indication that the game was enjoyable, but it can also mean that the children were just following our instructions, a concept well established in psychology.[15] Five of the children asked if they could play the game one more time. We did not ask the children any question regarding playing again, they did this out of their own free will. The group also saw that there were a long line of children outside of the classroom were we tested. We interpret this that the feedback from the children who had played the game to the children who wanted to play the game was good.

On the other hand, we see that at least one of the children from each group said that the game was hard. If the group were creating an application, then the fact that so many are saying that it is hard, would be a huge failure from our part. A game is supposed to be challenging, and as described in the chapter on methods for data collection, we did not want to ask the children to rate how challenging the game was. We also see that as the number of errors increase, the children rate the game as harder. This correlation may indicate that the game is to hard. Despite this, we do not see that as the errors increase, the positive feedback decreases. If we for instance look at group number 1. They had a total of 20 errors during play, but still rate the game as 'fun'.

Based on this analysis of whether or not the game is fun, the group believes that the children enjoyed playing the game. The group can not extrapolate from the test that the game will be equally enjoyable placed in Oslo Childrens' Museum. We see that the game may lack the ability to hold a child interested, given that they rate the game as to hard. We have neither tested how long the game will hold a child focused on the game or how it will compare to other games, given the child has the chance to choose freely between two or more games.

3.2.2 Playable by children of all ages?

How cartoon like the characters are in the game and the difficulty level is adjusted so that its not too easy and not to hard for children. This may indicate that the game is to easy for an adult, how ever it is fully possible for an adult to play the game since the Kinect-sensor only needs a body to track.

3.2.3 Universal Design

Principles of Universal Design are:

1. Equitable use
2. Flexibility in use

[15]Milgram experiment. Last modified November 15, 2012. Accessed November 25, 2012. http://en.wikipedia.org/wiki/Milgram_experiment

3. Simple and intuitive
4. Perceptible information
5. Tolerance for error
6. Low physical eff ort
7. Size and space for approach and use [16]

"Make it accessible for the maximum amount of people, but not everyone" (Jo Herland, spoken in lecture)
Not everyone can use everything, and many technological products are intended for specific users. What should be kept in mind though is how you can design your technology so that as many as possible can benefit from it?

Our game, being a video game is of course totally reliant on eyesight. Our current prototype was based on using your hand to control, but the Kinect can easily be programmed to accept other limbs (legs, head, shoulders etc...) - or even the body itself. In fact some of the children in the test held their arms to their side and jumped up and down, and still managed to play the game. Our prototype was independent of sound, so deaf people could play it, but such a game would be enriched by adding sound to it.

The fact that the game is language independent and simple should make it accessible to mentally impaired people, but unfortunately we did not have any such in our test group of children. To determine how suitable it is for children with special needs is outside the scope of our current project, but could be a possible continuation of it.

3.2.4 Educational

It's hard to say that the children learned much about fish during the game. As mentioned in part 2.2, the children either knew beforehand the control questions we asked about fish, or repeated their previous incorrect (or even random) answer. It may be that our questions were too easy for the children, or that the children simply thought the game was a work of fiction, and thus did not consider what was shown as a simulation of real life. This may have been exasperated by the fact that we used several of the children's' drawings in the game, instead of using more life-like depictions.

After looking at the game ourselves, we have to admit we did not attempt to make it realistic or lifelike. Most of the fish we employed were generic representations rather than real species (with the exception of the biggest "boss" fish in the game - the whale shark). Also the player fishes were able to eat every other fish out there (just not the trash). At best, they could learn that fish eat other fish, but most of the children knew this already.

[16] Lecture by Jo Herland on Universal Design, October 2012

However, the goal was not just to educate the children, but to inspire them to seek out information themselves and bring a greater awareness of the ocean, to the issue of pollution in it. In that effect we could say they were inspired, as the children were happy to talk about fish after the game, often enthusiastically so. And while the game might not have been educational as such, due to our constraints in time, and the requirement that the product should not be language dependent or have any writing, we are not sure if it would have been possible in the first place.

3.3 Anomalies in Player Scores

The individual player scores (2.1.1 Player Scores, page 10) showed that the participants who scored 2 while playing multiplayer more often scored lower while playing singleplayer, and vice-versa. Thi

4. Evaluation of the methods

4.1 Choice of evaluation method

The choice of methods to apply to our project was not an easy one. In the early phase we were of course influenced by our coursework in Experimental Design, but it did not seem useful for us to invent hypothesises and research questions in order to make the experience we wanted. Instead we opted to use Usability testing as the basis for evaluating our game and comparing our results to the requirements we had. In the end phase we learned about Grounded theory, which also might have been a valid method to reach a scientific theory based on our game and the testing of it, but we decided not to due to time constraints and there being no need to create such a theory. Consequently we decided to stick with Usability testing as the main method.

4.2 Reliabilty

The central question of reliability is; if another researcher conducted the same usability test, would the researcher get the same results? The answer to this question is dependent on the methods used to gather data.[17]

Observing users in there natural setting will have a variable reliability. We observed children only when they were playing a game. By focusing in on one task we can say that the reliability is higher than for instance if we were to observe at the entire day at the after school program. The

[17] Preece. *Interaction Design: Beyond Human-Computer Interaction*, page 471.

number of children observed at once are also relevant for evaluating the reliability. We observed a maximum of two children at once. Members of the group have previously experienced that placing many boys in a room and letting them play may produce unexpected interaction than two boys in front of a video game.

The types of observations we focused on are also relevant for the reliability. How long the children played or their score in the game are data that can be reproduced with a higher degree of reliability than for instance when we observed that some children did not use their right hand to guide the play-fish, but using their entire body by jumping around the room to guide the play-fish.

We conducted unstructured interviews[18], which have low reliability. Although, we would say that there is a difference between asking an open question to an adult, getting a longer, in-depth response, and asking a child what she thought of the game and getting the answer "it was fun". Despite getting short answers, we still see that it would be close to impossible to repeat exactly the same interview.

4.3 Validity

The purpose of the usability testing was to measure if the children were enjoying and learning from our game. The central question here therefore becomes; did the methods that we use to collect data, and the way they were conducted, actually capture whether or not the children had fun and learned from our game?

The first part is reflecting upon the methods that we used in the test. As discussed in Methods for data collection, we knew that we could not just use interviews to ask the children about what they thought of the game. We therefore triangulated the methods in order to get more accurate data. In chapter 3, Analysis- Is the game fun? we combine certain observations and certain questions that overlap. We believe that this strengthen the validity of our questions because they were central in order to answer one of the main question of this project.

Another point that strengthen the validity of our research was the use of the same interview-questions before and after the game was played.[19] This was done in order to see if the children had learned anything from our game.

The second part is to reflect upon the way we conducted our test. It was conducted in a room at Uranienborg skole, not at the Oslo Childrens' Museum. We noted in chapter 3 that where the game is presented may affect how children will interact with the device. We also had to restrict

[18] Preece. *Interaction Design: Beyond Human-Computer Interaction*, page 228.
[19] The questions were: *"Hva spiser fisker?", "Hva må en fisk passe seg for?" and "Hvilken fisk er størst?".*

the flow of children in and out of the room. We needed to make sure that we were able to collect the data that we needed. This means that we can not conclude fully how the the game will cope with a constant flow of children using the game without any supervision.

4.4 Ecologically Validity

Not being able to conduct the test in the natural environment of the game weakens the ecological validity. Although, the school also is a place where the children felt safe, so the validity is not as weak as for instance if the test had been done in a classroom at the institute for informatics.

The children certainly knew that they were being monitored during the test, which may have created an Hawthorne effect.[20] We were four men in the room. We were not all observing the children at all time, but this fact was probably not relevant in the child's mind.

4.5 Bias

This project has a selection bias when it comes to participants. Both our formative test and summative usability tests were conducted at the same after school program. We can therefore not say that our participants were chosen random. Even within the school we can not say fully that the children chosen were random. Many of the children stood in line outside the test-room, so it is fair to say that we did not choose the children, the children chose us.

There were a few occasions where we had to guide the children, by telling them how to play the game or what they could answer in a question. We also categorize this as a cognitive bias on our part, since we knew that we had the choice of not helping the child, but we did not want to be mean and not help. One of the groups were not able to answer one of the control-questions "who is the biggest fish in the sea?" A control-question means that they had already answered the question before they played the game. The interviewer therefore guided the children by saying what they had answered before they played the game.

There may also be confounding biases in our study. One of the independent variables were dividing the groups by age. This was to see if age affected ability to play and learn. Age may be an indicator to cognitive ability, but age can not say how fast a child will be able to interpret how to play the game. How smart a child is therefore an extraneous variable.[21]

[20]Hawthorne Effect. Accessed November 25, 2012. http://psychology.about.com/od/hindex/g/def_hawthorn.htm
[21] Extraneous variable. Last modified November 13, 2012. Accessed November 25, 2012. http://en.wikipedia.org/wiki/Extraneous_variable

4.6 Are the results generalizable, i.e., what is their scope?

If we divide the methods in to two categories the enjoyment- and the learning part we see that we can generalize when it comes to the learning, but not enjoyment.

The way we collected the questions of enjoyment was by asking how the children felt about the game and by a subjective judgement of how it looked like the children were enjoying themselves. Both the answers and the observation were highly based on individual conceptions. This means that if the test was conducted in a different setting, for instance if the children had different games they could play, then mabye the enjoyment would fade in comparison.

The learning part may be more generalizable. Asking the same questions before and after an experiment is a generally approved scientific method.

5. Conclusion

We have created a working high-fidelity prototype based on the requirements given to us from our employer/customer, Oslo Children's Museum. The prototype was based on the group's ideas, yet inspiration and visual design was taken from the children of Uranienborg SFO. After validating our vision in a formative usability test session, we employed a summative usability test with children from the same SFO on the high-fidelity prototype. We tested if the requirements we had been given, as well as the ones we had set ourselves, had been fulfilled. We conclude that these requirements have been met:

-The game requires no reading and is not language dependent
-can be played by two people at once
-new players can enter in the middle of the game without it being affected
-we have seen that the game is fun for both children and adults
-the feedback from Katy, our customer from the Oslo Children's Museum, have been positive
-we have continued to develop the prototype based on the feedback we got from the children, making our project part of iterative design

When we started this project, all we knew was that we were to design an experience for children, using the Kinect interface, on behalf of Oslo Children's Museum. In our design brief we set this goal:

"Our goal is to create an interactive experience where children can play, be creative and engage children to learn more about the world around them"

Our game certainly can be played, it is also fun and engaging. The game itself is not about being creative as such (very few games are), however the children that helped us make it were. The educational value of the game is low, however if the enthusiasm the children played with, and talked about fish and things that live in the sea after playing is any indication, we assume that the last part of that goal has been fulfilled. Are children not easily affected by what they see, how they play, and what they hear?

In our data collection and analysis we have looked at different ages and genders in children, and looked for interesting patterns that might be of scientific value. This is a nod towards experimental design, but without having defined any hypothesis. We found no real differences between the genders when it came to enthusiasm in playing - they both seemed to have fun. When it comes to how well each gender played the game, we did not find significant differences in their average scores. The age groups had differences only in the interviews, generally older kids knew more about fish and the sea beforehand, but with a single exception the first graders seemed as entertained by the game as the third graders. In fact it was the 2nd graders that scored the highest in the game, with the 1st graders on second place. Our concerns that the game might be too hard for the youngest or too easy or childish for the eldest thus didn’t seem to have any base in reality - which is maybe not surprising since also adults are entertained by playing it.

6. Sources

6.1 Internet sources

About. Cherry. Kendra. “Hawthorne Effect”. Accessed November 25, 2012. http://psychology.about.com/od/hindex/g/def_hawthorn.htm

Answers. “Problem space.” Accessed November 15, 2012. http://www.answers.com/topic/problem-space

Johnny Holland. “How to combine multiple research methods: Practical Triangulation.” Last modified August 20, 2009. Accessed November 15, 2012. http://johnnyholland.org/2009/08/practical-triangulation/

Wikipedia. “Extraneous variable”. Last modified November 13, 2012. Accessed November 25, 2012. http://en.wikipedia.org/wiki/Extraneous_variable

Wikipedia. “Subject-matter expert.” Last modified 5 October, 2012. Accessed November 15, 2012. http://en.wikipedia.org/wiki/Subject-matter_expert

Wikipedia.“Fun”. Last modified November 22, 2012. Accessed November 25, 2012. http://en.wikipedia.org/wiki/Fun

Wikipedia. “Milgram experiment.” Last modified November 15, 2012. Accessed November 25, 2012. http://en.wikipedia.org/wiki/Milgram_experiment

6.2 Books

Interaction Design: Beyond Human-Computer Interaction, 3 Edition. By: Helen Sharp, Yvonne Rogers, Jenny Preece, 2011.

Research Methods in Human-Computer Interaction. By Jonathan Lazar, Jinjuan Heidi Feng and Harry Hochheiser. 2010.

Sosiologisk blikk på brukerorientert design. By Svein Hovde.

http://www.uio.no/studier/emner/matnat/ifi/INF1510/v11/undervisningsmateriale/Sosiologi%20og%20brukerorientering.pdf

Universal Design. By Jo Herland.
http://www.uio.no/studier/emner/matnat/ifi/INF2260/h12/undervisningsmateriale/inf2260october2012.pdf

7. Appendix

7.1 Observation table:

#	Sender	Group	Finished M?	Finished S?	Replay?	Motivated	Time M	Bored	Time S	Score	Score S	Errors
1	f	8	1	DNP	1	1	2:00	0	DNP	0.60		1.00
2	m	8		DNP	0	1		0	DNP	2.00		1.00
3	f	9	0	DNP	1	1	1:30	0	DNP	1.00		4.00
4	m	9		DNP	1	1		0	DNP	0.20		0.00
5	m	1	0	0	0	1	3:50	0	1:30	2.00	1.00	9.00
6	m	1		0	0	1		0	3:24	2.00	0.50	11.00
7	m	4	1	1	0	1	2:10	0	1:14	2.20	2.00	6.00
8	m	4		1	0	1		0	2:20	0.80	2.00	5.00
9	f	6	1	0	0	1	3:00	0	7:07	0.90	1.00	5.00
10	f	6		0	0	1		0	2:30	2.00	0.80	3.00
11	m	5	1	1	0	1	5:00	0	4:00	0.20	2.00	2.00
12	m	5		0	1	1		0	2:54	2.00	1.00	1.00
13	f	3	0	1	1	1	2:00	0	1:11	0.20	2.00	0.00
14	f	3		1	0	1		0	2:24	1.00	2.00	0.00
15	f	2	0	1	0	1	1:41	0	1:30	1.60	2.00	1.00
16	f	2		0	0	0.00		1	DNP	1.00	DNP	1.00

Columns:
"#" : Participant's number
"Gender" : Participant's Gender
"Group" : Group number
"Finished M?" : Did the participant finish multiplayer? (0 means no and 1 means yes)
"Finished S?" : Did the participant finish singleplayer? (0 or 1)
"Replay?" : Did the participant ask to play again?
"Motivated" : Did the participant seem to enjoy playing the game?
"Time M" : Multiplayer timer
"Bored" : Did the participant find the game boring
"Time S" : Singleplayer timer
"Score" : Player score, multiplayer
"Score S" : Player score, singleplayer
"Errors" : Number of errors

7.2 Intervjuguide

Før spill:

Hva spiser fisker?
Hva må en fisk passe seg for?
Hvilken fisk er størst?

Etter spill

Hva synses du om spillet?
Hva kunne vært bedre?
Hva spiser fisker?
Hva må en fisk passe seg for?
Hvilken fisk er størst?

8 Other

Legg ved PDF fra SnowCastle her