UNIVERSITY OF TARTU VILJANDI CULTURE ACADEMY

Design and Development of Virtual Environments

Rene Rebane, Kaspar Roost

# VIDEO GAME ENGAGEMENT TESTING WITH PROTOTYPES: COMPARING VIDEO GAME PROTOTYPES IN DIFFERENT DEVELOPMENT STAGES BY TESTING THESE ON SIX TO ELEVEN YEARS OLD KIDS.

Master's Thesis

Supervisor: Ander Tenno, PhD

Consultant: Hegle Sarapuu

Approval for defence: /signed digitally/

Viljandi 2014

# EXECUTIVE SUMMARY

The goal of this thesis is to research how fun and engaging the game is during different development and prototyping stages. Need for this research comes from real life. Game developers usually make several prototypes or proof of concepts of a game to find out if the core mechanics are fun. This thesis helps to find out how comprehensive these prototypes have to be to test the engagement of the game so in the future game developers can maybe spend far less time developing the prototypes. This is important because from the authors' experience every next logical step in the prototype design increases the scope of the development multiple times. For example creating a paper prototype may take only several hours to make but developing a working playable game prototype may take several days up to several weeks. There is very little academic research done in the field of game development and game design as the field itself is still quite young — only about 30 years. There is very little academic work regarding playtesting especially with children. Testing object of this research is an educational kids' iPad game that authors have been developing with a small team themselves. This game is also going to be released in 2015 to the public. So the research focuses on playtesting with kids who are 6 to 11 years old. Testing was conducted mostly at the public school with the help of teachers who already use video games in their curriculum. Playtesting has been divided into three parts as there were three game prototypes to represent the different stages of game prototyping: paper prototype, low fidelity prototype and high fidelity prototype. Pretest and posttest surveys were created based on Mihaly Csikszentmihalyi's theory of flow and psychology of optimal experience but during the tests authors also used the observation of kids and how they played the game. Authors consider the low fidelity prototype to be optimal in the context of game testing and measuring the flow with elementary school kids. There has to be bare representation of testable game mechanics and few visual or semantically correct cues for the users testing it. Game mechanics have a bigger impact on the playing experience than the game art, animatsions or sounds.

# TABLE OF CONTENTS

## USED TERMS

In this paper authors have used these terms and phrases in the meaning and context that is explained here:

**AAA title** – "triple A" is a term used for games with the highest budgets of development and promotion by the game industry giants.

**Game elements** – Character and everything they communicate and interact in the game

**Game mechanics** – Constructs of rules intended to produce gameplay

**MVP** – Minimum Viable Product as Eric Ries describes it (2011)

**NPC** – Player Controlled Actors

**NPCA** – Non Player Controlled Actors

**Sandbox type game** – a game where player can move around freely in the game world and approach the game objectives in the their preferred order

**Lean** – "to maximize customer value while minimizing waste. Simply, lean means creating more value for customers with fewer resources." (lean.org 2014)

**Grind** – in the context of video games grinding is the activity that player has to do for extended periods of time to achieve something. Grinding can also be considred to be hard, annoing or even frustrating.

# INTRODUCTION

The goal of this thesis is to research how fun and engaging the game is during different development and prototyping stages. The subject of this research is important because developers may spend too much time to make an elaborate game prototype to test if the game is engaging and fun when similar results can be accomplished with far less. The research project aims to find out what the MVP can be for game development.

## Problem Statement

The problem that authors are solving comes from game development process. Game development takes usually a lot of time and money. But it is all wasted, if target audience finds the result not attractive, engaging or fun. As attractiveness, engagement and fun are all so personal, subjective and hard to measure, then authors have used theory of flow by Mihály Csikszentmihalyi (1990, 1997) to research the topic in question.

So it is essential for game development studio to find if game mechanics, graphics, sound effects etc are engaging enough before actually developing the product. Only quick way to get to know this is to test the planned game on users. This is where MVP comes handy for every kind of research on the product.

## Need and importance of the thesis

Small game studios don't have big budgets to spend on R&D. For them it is essential to get the customer feedback as soon as possible in the very beginning of the process. Important is to know, how polished and refined a prototype and MVP as a whole has to be to get correct feedback. Bad alternatives would be:

A. prototype is too low-fidelity and do not show all the functionality (necessary for deciding) clear enough to users and/or customers;
B. working hours and effort spent on prototype is too big and most of it might be thrown

away according to test results.

The uttermost important is to find the secure balance between those two options - how to waste less money/time and gain test results that can be trusted.

## Existing research

There is very little research to be found regarding playtesting and fun of video games as the game design and development field itself is fairly new to the academic world. Game design books cover somewhat cover these topics but can not be considered as academic work. Books are written by the game industry veterans who have based the books on their experiences and personal observations. Very little can be found on the topic of measuring fun and engagement in games. Most of the information that can be found is in the form of presentations that game designers and developers have created based on their experiences for different game development conferences and industry related events. There are also some blogs and blog posts on the topics.

In 2006 "All work and no play: Measuring fun, usability, and learning in software for children" by Gavin Sim, Stuart MacFarlane, Janet Read was published. Abstract of the study:

> "This paper describes an empirical study of fun, usability, and learning in educational software. Twenty five children aged 7 and 8 from an English primary school participated. The study involved three software products that were designed to prepare children for government initiated science tests. Pre and post tests were used to measure the learning effect, and observations and survey methods were used to assess usability and fun. The findings from the study demonstrate that in this instance learning was not correlated with fun or usability, that observed fun and observed usability were correlated, and that children of this age appeared to be able to differentiate between the constructs used to describe software quality. The Fun Sorter appears to be an effective tool for evaluating products with children. The authors discuss the implications of the results, oVer some thoughts on designing experiments with children, and propose some ideas for future work." (All work and no play, 2006)

Some parallels can be drawn from this study. For example the kids' age range is somewhat similar. Also one of the study subjects is fun in games but it is more related to how it affects learning. Similarly game prototypes are used for testing but in their case these are three different

games not different versions of one game. Process of the test itself is somewhat similar although authors of "All work and no play" do not mention playtesting at any point and rather use usability testing methods, the process is still very similar and comparable results can be achieved through the observation.

## Goal of the thesis

Goal of this thesis is to optimize one aspect of the game development process — development of a playable prototype with the purpose of testing if core game mechanics are fun and engaging. With the results of this thesis future game developers can assess how much effort should be made to find out if the game idea is fun and engaging for the target group.

## Context

## Code2Kids - Project's goal and background

Code2Kids is an iPad game that that authors are developing for the kids 6 and up. They will learn the basics of any programming language and problem solving through play. Authors started Code2Kids game project to prove that most of the current programming games/tools meant for children are seemingly missing the fun game element thus these games can't be commercially successful and self-spreading amongst the target groups.

Hypothesis for creating the Code2Kids was that children would play the educational game more and even voluntarily when it had similar fun game mechanics that commercial entertainment games use. The idea was to "hide" the educational content inside the game mechanics and force them to learn through the facts that typical educational games provide in between the game sessions but lean through the gameplay itself. From interviews with parents, coaches and advisors it has appeared that programming *per se* is not a very important argument, instead developing logical thinking and problem solving skills is rather preferred and will cover wider audience's interest.

#### Code2Kids - Need and importance

There is the will and actions taken (code.org 2014) to teach kids programming for few years now but much of the effort is wasted because of missing the correct methods, tools and ways of doing

it according to target the age group it is meant to.
Children grow up playing and learning at the same time, this is how they learn fastest - not using it is waste of time and effort from both sides (Brown, 2008).
As a market share of educational games is growing rapidly almost doubling every year (see table 1), there is a great need for better quality educational games. Best way to achieve this is by doing rapid development and working together with kids to test the game during every development phase that is clickable and playable. This kind of research is needed to make this game development process more effective and result more user centered.

## Code2Kids - Problem

Programs for teaching programming to kids are more tools than games - educational content overrides fun.
Those tools are developed programming-centered way, nor user-centered or kid-centered way. Kids are being taught programming with tools that don't fit with their age (methodically, not visually).
How to find the right game mechanics that work best to hide educational content so that kids are willing to use the game for fun and learn new problem-solving techniques during it.

## Code2Kids - Success criteria

1. To create an **engaging game** with
2. **the educational content "hidden"** in it
3. to guarantee break even or **profit**

## Code2Kids and the need for playtesting

To guarantee that game development team has been on the right track, every new functionality has to be tested and validated by target customers. Meanwhile team tested the prototype with smaller groups of students but bigger and documented testing sessions were conducted in March and in May 2014. Prototype was improved in between tests according to first test results.

## Background

### The current state of educational games

Educational gaming market is characterized by boring and mediocre games. Educational gaming has historically grown out more of the academic side rather that the entertainment industry side of the field. Today the educational games industry and entertainment games industry have a big

gap between them. Both of these industries have lost something on the way. Educational games have lost the fun which is the core of the entertainment side of the industry. And AAA games have lost the educational side for the most part (Floyd 2014, 00:43 – 01:11). This also means that most of the gaming industry's talents have left from the educational games niche to work on blockbuster and AAA games. Still times are changing and there are signs that in the future this gap between two industries might get narrower.

*Table 1. Global edu-gaming and overall market size comparison*

| | Global edu-gaming market size:<br>(GSV Advisors, Edu Factbook 2014) | Global games market overview:<br>(Schutte 2014) |
|---|---|---|
| 2012 | 2.0B$ | 66.3B$ |
| 2015 | 4.4B$ | 80.5B$ |

As seen in the Table 1. educational games market size is only a fragment of the global overall games market size. From GSV report it can be seen that edu-gaming market size is estimated to grow by 55% by the year 2015, while overall market size is estimated to grow only by 25%. This means that there is a great overall need for educational games in the world. Also the means necessary to play these games in educational establishments are becoming cheaper and more accessible. Classrooms can easily be equipped with computers and in some cases tablets.

Not only classrooms bring the digital content closer to kids but also parents themselves. 42% of adults in the US owned a tablet in January 2014. Only four years before that the penetration rate of tablet devices in US was mere 3% (Pew Research Centre, E-Reading Rises as Device Ownership Jumps). The adoption rate of tablet devices has been very fast. Only in couple of years nearly half of the adults in the US own a tablet. This has also a great impact on how devices are used between family members and how digital content is used. Example from the Nielsen Group's research "The rise of gadgets is ushering in a new generation of kids who are growing up digital. According to a Nielsen survey of adults with children under 12 in tablet-owning households, in Q4 2011 seven out of every 10 children in tablet-owning households used

a tablet computer (…)" (*Nielsen, 2014*).

## Educational games vs entertainment games

Although the educational gaming market is growing rapidly the games that lead this growth are mostly entertainment games that educational establishments have adopted and repurposed for their own needs. One example of this phenomenon is Minecraft which originally wasn't built as an educational tool but rather a sandbox type open world exploration and building game. The whole game world is built out of blocks that represent different materials. These materials can be reused to build, or crafted into new materials or tools. Minecraft has become highly popular among the kids around the world. Educators saw the popularity and the potential of the simple and flexible system that Mojang had built and started working on curriculums to support the game as an educational tool. Today Minecraft is considered to be one of the most popular educational games with its wide acceptance as an educational tool by more than 2500 schools worldwide (The Edublogger 2014).

To put educational game studio and triple-A game studio sizes into perspective — Mojang, the company behind beforementioned Minecraft, has 43 employees (Mojang 2014), but from the entertainment games industry side there is an award winning and record breaking triple-A title "Grand Theft Auto V" that sold 11.2M units in the first 24 hours (IGN 2014). "GTA V" is developed by Rockstar Games, where more than 1000 people have worked on this very game only (Develop-Online 2014).

## Prototyping in Small vs Big game Companies

Some of the biggest problems that smaller studios face compared to bigger companies is the lack of time and funds to create games. The development cycles have to be much shorter and more cost effective than they might be in bigger companies. There is also the fact that game design itself is very much a creative field and it is hard to predict the scope of the game's prototyping phase. What makes this problem even more important is what Schell states as the rule of the loop: "The more times you test and improve your design, the better your game will be" (Schell, 80). In a sense this statement is quite obvious — iterations of validation and improvement will result in better outcome.

In the context of small companies that develop educational games fun and engagement become secondary as primary goal is the educational value of the game. Looking at the Apple App Store's educational games sections it seems that a common way to build educational games is to

take a popular and simple entertainment game and re-use the mechanics to build an educational game out of it. Problem with that is that the games are not built around educational content in mind from the ground up. From one side the educational part of the game is not as effective as it could be and on the other side educational content may destroy the fun and engaging experience the game might have had without it.

## Hypothesis

There is no need for polished playable video game to find out if the core game mechanics are fun and engaging when target group is 6 to 11 years old kids.

## Research Questions

1. How do the results differ when testing the same game mechanic in different prototyping stages?
2. When do signs of flow occur when testing on different prototypes?
3. How much does the game art influence engagement?

## Methods

## Overview of the Methods Used

Methodology used consists of game-testing and pretest-posttest questionnaire. Testing group consists of 3x5 Estonian speaking children from Tallinn. As division by sex would cut off half of the potential customers then authors try to avoid this situation by taking care that both sexes are represented equally in the test sessions.  All the children in test groups are selected from age range 6 to 11.

## Video games and fun, engagement and flow

To evaluate children's engagement while using prototypes in different fidelity levels of the game, authors need to ask children certain questions related and inspired by Csikszentmihalyi's theory "Flow: The Psychology of Optimal Experience" (Csikszentmihalyi, 1990). The same set of questions is asked before and after testing with every prototype.

Additional prototype-based question is asked after testing. After the test session children are asked to guess how much time did they spend on this playtesting session. Opinion and time from the start of the session recording is marked down and compared. If the playtime is bigger than

players opinion then authors consider it as a symptom of the flow. Like Csíkszentmihályi says "A person who is completely absorbed in performing an activity might reach a state of **flow**, a mental condition that is marked among other characteristics by a distorted sense of time" (Csikszentmihalyi, 1990), "time flies!" (Geirland, 1996) adds Csikszentmihalyi.

## Video Game playtesting

Games are prototyped in a lean and iterative manner. The whole process begins with a simple design or idea what the game should be and continues with the creation of more sophisticated prototypes with every iteration. At first the prototype can exist for example on paper or played out in other forms but with every iteration the prototype can become more complex depending what the prototypes purpose is.

Schell has described the process based on Barry Boehm's spiral model of software development like this (Schell 2008, 82):

1. Come up with a basic design.
2. Figure out the greatest risks in your design.
3. Build prototypes that mitigate those risks.
4. Test the prototypes.
5. Come up with more detailed design based on what you have learned.
6. Return to step 2.

At the core of this model is the idea that every prototype has to have a purpose and developer has to assess the risks and then mitigate them. In author's case build a prototype for the purpose of finding out if the game idea is fun for kids in the target group. Building the prototype and testing it can also give the developers valuable feedback what can be improved regarding engagement. As the purpose of the prototype is set minimum effort should be made to develop a robust prototype so the hypothesis can be tested. Hypothesis being that game idea is fun. In the authors' case the purpose is to find out through different prototype fidelity levels when fun can be measured and when it serves its purpose the most.

The whole purpose of the process is to mitigate risks as fast as possible during the design and development cycles. Developing many robust prototypes and quickly testing them is not crucial for developing a better game but also for the scope and resources of the project.

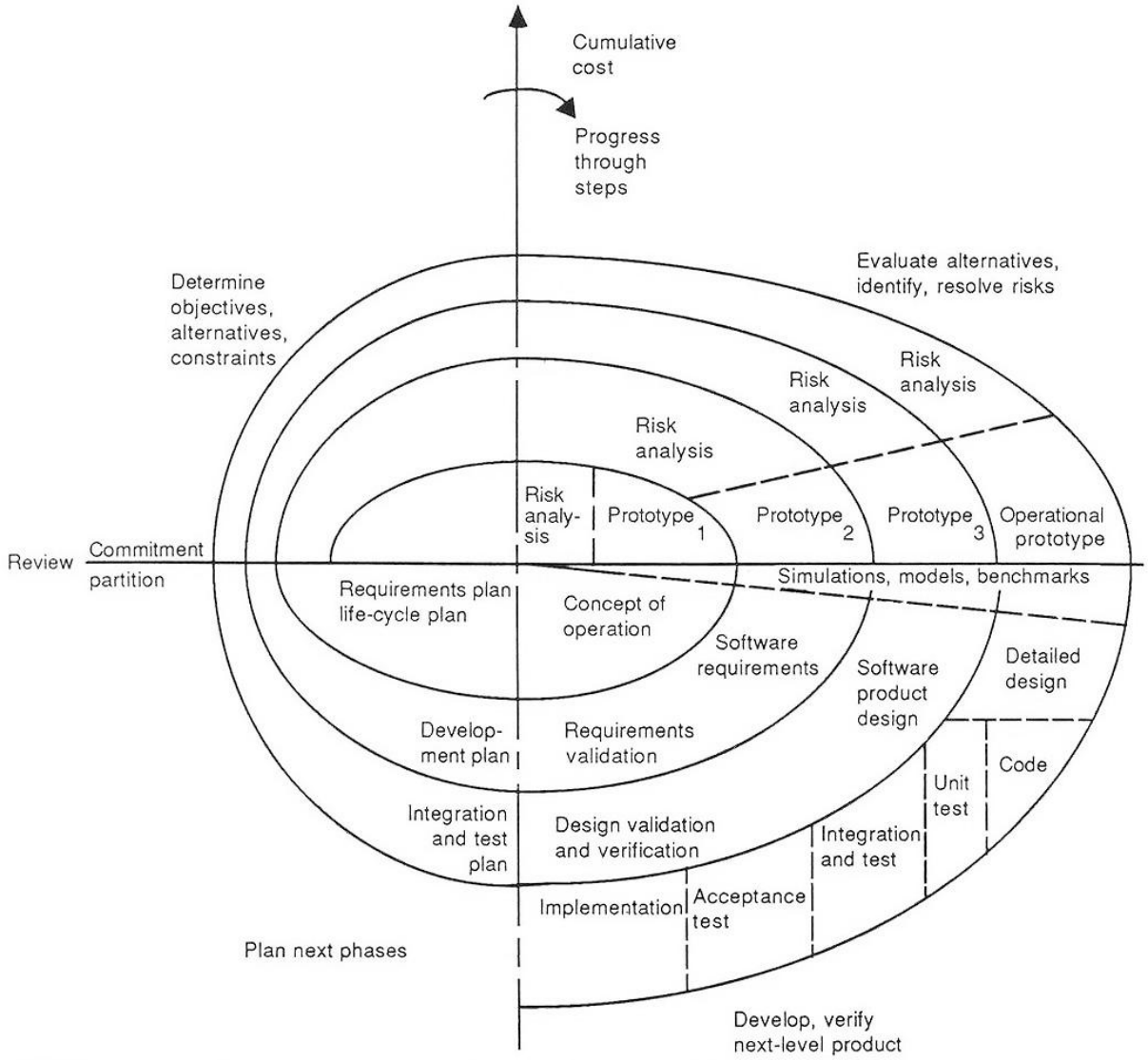


*Figure 1.* The Spiral Model of Software Development (*Source: Schell 2008, 83)*

According to Schell there are two schools of playtesting. First one conducting the playtest without anyone other being in the room but the tester. Pros for this approach are that testers will have undisturbed experience while playing and they will not feel any pressure to act or perform a certain way. Cons are that a lot less information can be gathered during the testing sessions as extra questions can not be asked and testers' emotions and what is going in their heads can not be recorded and explained. Second one having the person conducting the test also in the room. Pros for this kind of approach are that person who is conducting the test can directly observe the user and ask extra questions. It is also a good practice to ask testers to think aloud so that their thought process can also be observed and recorded. Cons for this kind of approach are that person who conducts the test can interfere with the process too much especially if that person is

one of the developers or designers who wants to explain their work or defend it. For our purposes we have to use the second method where testing person has to be present because we are testing on smaller kids.

We based our playtesting session on Fullerton's playtesting rules (Fullerton 2014, 284). There are altogether 25 rules that he suggests. Fullerton himself asks game designers to bend these rules according to their needs.

- Fullerton suggests that game should always be tested before the developers feel comfortable about testing. Idea is that when developers feel comfortable with the game it is already too complete and polished which is not important for playtesting.
- He also suggests to plan the prototyping in different development stages of the game so that minimum effort can be made to start testing. Game should be simplified enough that playtesting could be done on the day.
- Purpose of the prototype can also be simplified by trying to answer one question when developing the prototype. For example if the game idea is fun, should the main character use guns or a bow, how many enemies make the game challenging etc.
- It is also important to be grateful towards playtesters as they have put their time and attention into testing.
- Fullerton also suggests to design the learning experience so that as little as possible should be explained during the test.
- Never should a developer or designer blame the playtester when playtester can't complete a task or gets confused and frustrated during playing. Developers and designers should never make the playtester feel foolish.
- Getting to know playtesters helps sometimes understand why people are playing games in the certain way. People have had many different experiences with games and distinguishing for example casual and hardcore players may have a great impact on the testing results.
- Is also suggested that during the playtest designers and developers should not explain about the history of the project and why any decisions were made. Testing session should simulate the situation when users gets the game in real life and starts using it without any prior background knowledge. "Off the shelf" approach should be followed and as little as possible should be told to the playtesters.
- Notes should always be taken during the playtest. This can be very helpful after the test when data has to be analysed and tester might make an extra effort to give out more

information when they see that notes are taken.

- It is important to be selfish while playtesting. Too much effort shouldn't be made to give testers a good time or the opportunity will be missed to get the hard truth from the players.
- Players should be asked to talk out loud about what they think and what they feel. Only this way observers can see into the player's mind and get a better understanding what is really happening. Players should be reminded to talk out loud when they forget it.
- Everything should be noticed and noted down. Not only the things that observer likes or prefers to see. A lot of new information can be gathered this way that can be used is new designs and prototypes.
- Observers or people who conduct the tests should be as quiet as possible during the test. The main purpose of the playtest is to see how other people think and act and disturbing them while they play will ruin that.
- Fullerton suggests to try to see the big picture. Not to focus only on the game on the screen or table but what is happening around the game. Observers should also notice the body language, emotions, interactions with other people. How the game affects the people who are playing it.
- Questions should be answered with questions so that playtesters can explore the answers further by themselves. Persons who conduct the tests should explain as little as possible during the test and never give definitive answers or the opportunity to find out how playtester overcomes the problem or challenge gets lost.
- Failure should be seen as constructive feedback that can be used to build upon in the future. If something breaks or doesn't work how it was designed it is an opportunity to improve the games.
- After the playtesting session the experience should be discussed. Good questions to ask are for example what was the most challenging part and was the easiest, what did the players most like about the game and what they didn't. Answers should be concrete and specific.
- Feedback should be put into context. Feedback shouldn't be taken literally as people are not always rational about what they think and expect. Also feedback can be misinterpreted.
- Fullerton also suggests that it is a good idea to collaborate with the playtesters. Brainstorm new ideas with them or ask them to think of alternatives that they would like to see. This gives again a very new perspective to the game ideas and the designs that can

greatly help the project further.

- Playtests can be very honest if rules are followed and sometimes things come out that feel very hurtful to designers and developers. This kind of feedback should be embraced and put to good use when developing further designs and prototypes.
- Unexpected things may happen during the testing sessions and this should be also embraced as it may lead to new opportunities and ideas.
- Playtesting process is important because this will help the project further and mitigates the risks.

## Pretest and posttest questionnaire

Authors wanted to measure how this particular combination of game mechanics would influence the player and would the flow (Csikszentmihalyi, 1990) appear. As the testing sessions were made with different prototypes that are not fully featured products, then authors had to find the way to determine if there is a chance of flow in the game for this age group or not. This is why "Design 1: Randomized control-group pretest-posttest" by D. M Dimitrov & P. D. Rumrill (2003, 160) was used as one of the research methods – to detect change in behavior and attitude towards the game by only measuring the important parts of it (see chapter: Pretest-posttest Survey), not the whole combination of aesthetics and function - the game as a product.

The questionnaire consists of six questions that are asked before and after the testing session. Questions are made to fit into both situations. Prototype was demonstrated to the test participants before they had to answer the pretest questions. The difference in answers, that authors wanted to detect, had to come from the playing experience – created by the smaller components like game mechanics, levels with balanced difficulty etc..

## Smileyometer and Ballometer

For pretest and posttest answers to be more comparable and usable in research, authors presented a "Smileyometer" (Read et al., 2002) to children — a selection of smileys. When showing it, authors asked the test participants to select one smiley to present a non-verbal answer. Smileys were covering answers on the scale of one to five, first one being grumpy and last one being happy thus reflecting child's emotions to questions easily.

During the tests, it appeared quickly, that smileys can't be used on questions that do not reflect impression or emotion, but do show children's evaluation on the game. So authors created

quickly a Ballometer , a row of five discs or balls, each bigger than previous (see fig. 6).

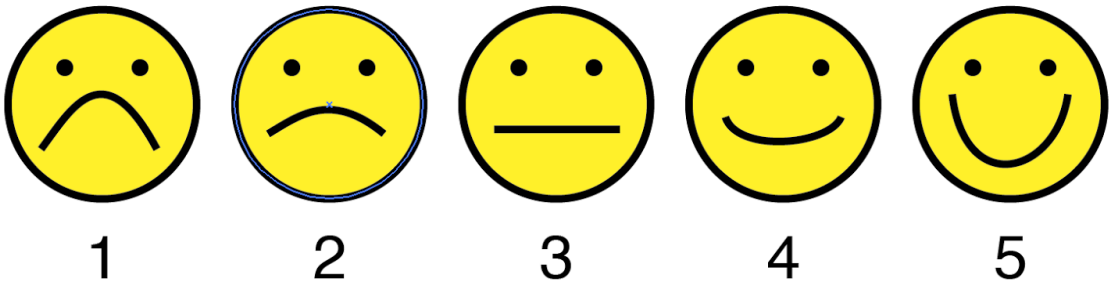


*Figure 2.* “Smileyometer” (Read).

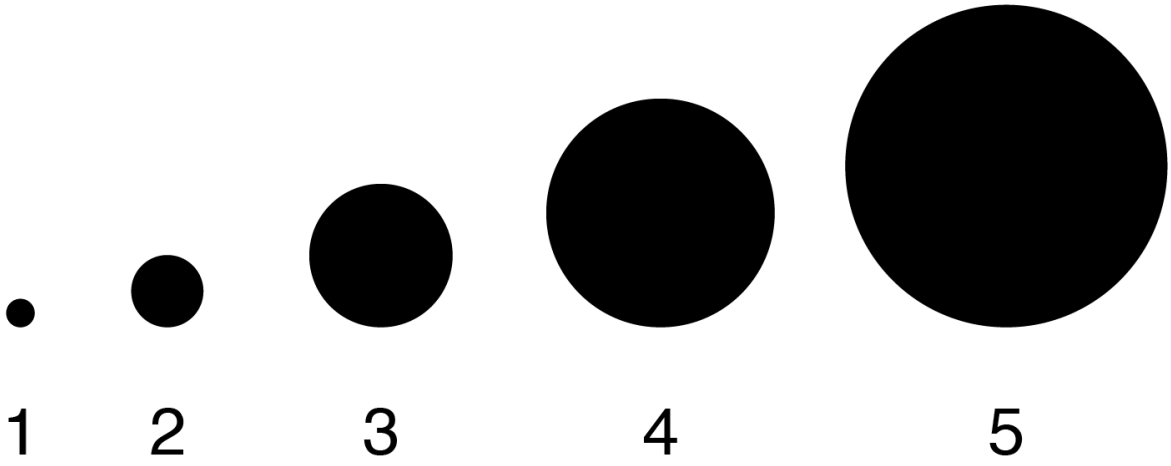


*Figure 3. The* Ballometer - an alternative scale that authors created to get more precise answers to non-emotional questions (Questions 3: How much does the task at hand need concentration?)

## Scope and Limitations

Authors have developed the game prototypes based on Fullerton’s suggestions (Fullerton 2014, 284) which means that:

- only a small number of game design elements and game mechanics will be implemented and tested with each prototype.
- Digital version of the game does not have a tutorial section which has to be done by authors before the game begins when kids don’t understand the rules right away. For the paper prototype more extensive verbal explanation of rules is made;

- game prototypes represent only a very narrow cross-section of the final product;
- The digital prototypes are development builds and can be accessed only on developer's devices. No public access is available to the prototypes.

# METHODOLOGY

## Experiment Using a Prototype

## Procedure of selecting participants into study and conducting hallway test sessions

Children for testing are picked randomly from Gustav Adolf elementary school's (Tallinn) computer class lessons at this very moment when the test takes place. Students will approach testers table alone or in self organised pairs and testing may start. So it is done iteratively pair by pair using the "hallway testing" method. This method is used to gather random people from the "hallway" rather than trained testers from a company for example. Still gathered people should create a cross-section of the target group for the product.

## Playtesting

Playtesting session were carried out either in the hallway of the Gustav Adolf elementary school or at acquaintance's home. The playtesting setup for the testers was the same.

Authors set up the computer with an external webcam that was faced down towards the table so that either paper prototype or both iPads could be seen with kids' hands playing the game. Also kids' voices were recorded during the session for further analysis of the think-out-loud comments and vocal expressions of the emotions. Two iPads were set up with started game apps (prototypes for that session), so that if next children come, they have quick and easy access to start playing. Also the paper prototype was set up so that right levels and the right amount of game pieces could be quickly brought out when needed. There were two observers (authors of this thesis: Rene R. and Kaspar R. ) in the room who carried out different tasks. As test sessions with one child or pair of children are so short, then distribution of the tasks had to be agreed in advance.

When testing with paper prototype only one child was asked in to join the testing session. Main

reason for that was that playing out the rules of the game on the paper prototype needs someone who plays out the game's rules. For example moving the characters when child says "Play" or removing them when they die.

When testing the digital prototypes two kids were asked to come in so they can play together on two iPads and can be interviewed together. The purpose of this arrangement was to reduce the shyness in kids. Playing together with a friend gave kids more courage to talk out loud more and be more open. In these cases kids also make jokes between each other and overall were more relaxed.

The test procedure:

1. Kid is called in from the classroom or other room and cheerfully greeted.
2. Rene asks his or her first name and age.
    a. Kaspar takes notes
3. Rene gives a quick explanation what will happen next during the test session and kids are made as comfortable as possible with assuring that there is nothing to be afraid. Also children are asked to think aloud during the test and represent any kind of emotions they might feel during the playtesting.
4. Either paper prototype or one of the digital prototypes are shown for the first time. For the purpose of giving kids some first impression of the game, they can not yet play the game, only to observe and comment.
5. Kaspar asks the pretest questions from the child in test and writes down the answers. Rene shows Smileyometers or Ballometers to kids depending on type of the question. When needed question is repeated and/or explained so that children would give the maximum quality input to test data.
6. Playtesting begins with giving the kids either the paper prototype or one of the digital prototypes. Some basic rules are explained to kids who play the paper prototype but nothing is said to kids who play the digital versions of the game. During the test Rene writes down the observation comments.
7. During the playtest kids are asked several times to think aloud or asked what they are thinking at the moment. And also when solving the test tasks seem to take a very long time if they wanted a tip so they can quickly continue with next levels so that they wouldn't feel bound to the process. These questions also work well do observe if kids are in the state of flow.

8. When kids finish the levels Kaspar asks them the posttest questions and writes down the answers and shows Smileyometers or Ballometers to kids when needed.
9. Kids are thanked and sent back to the classroom.

## Prototype Design

Authors have developed three Code2Kids game prototypes, which all contain the same game mechanics and levels. These prototypes are based on Tracey Fullerton's descriptions of game prototypes and reflect the different prototyping stages of a game design and development. They have a minimum set of rules or game mechanics implemented (Fullerton, 2008, chapt. 7-9).

These three prototypes are:

1. Robust **paper prototype** that resembles a board game
2. **Low fidelity prototype**: Early development or technical prototype without game art, sounds or animations for iPad
3. **High fidelity prototype**: Early development prototype with animations, game art and sounds for iPad

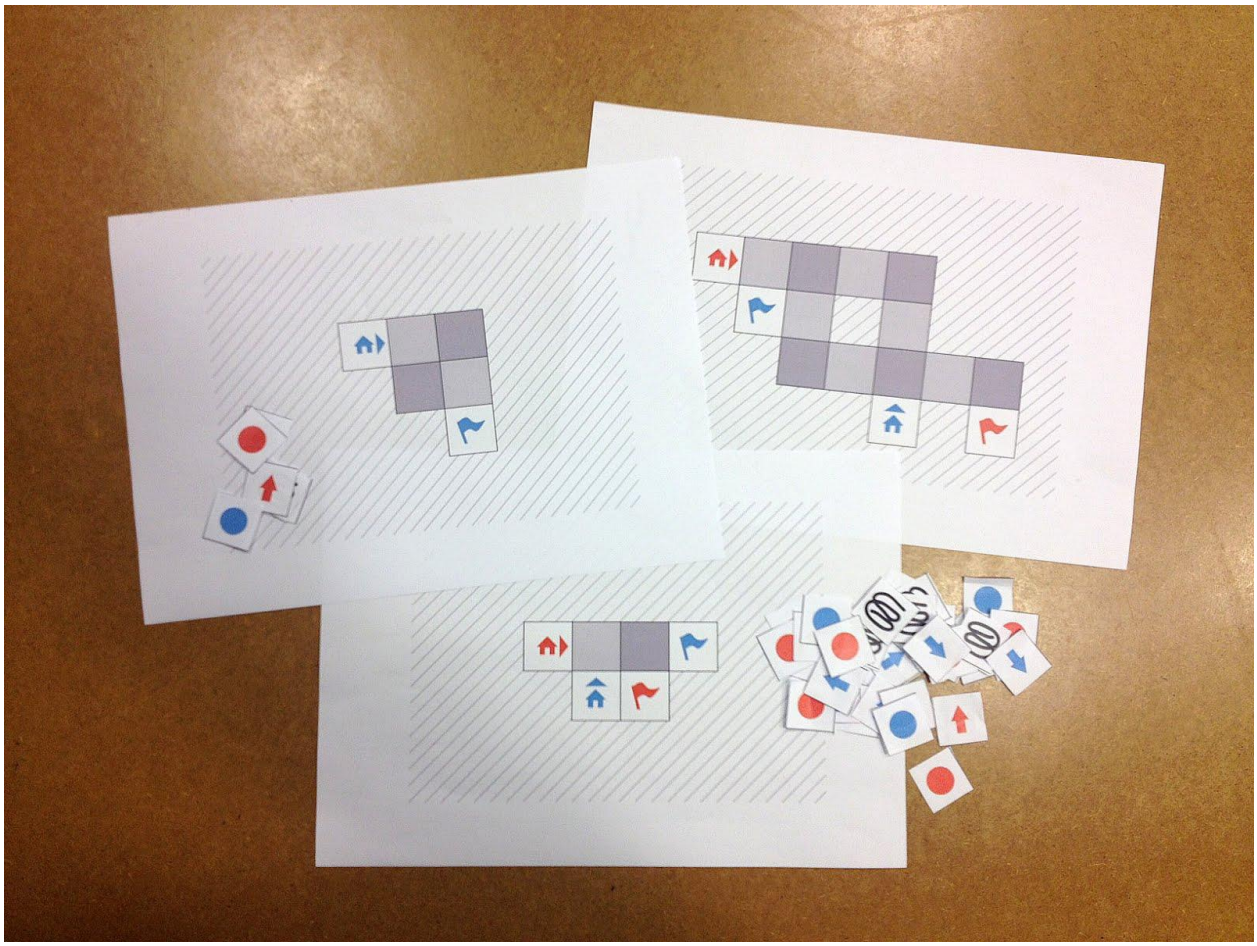

*Figure 4.* Board game as paper prototype

Robust paper prototype as board game was made to reflect one of the earliest prototypes in game development process. The design is based on the selected core game mechanics of our final product. Design of the board game itself is very robust. No or very little attention to aesthetics was given. The design of the paper prototype was led by the idea of keeping it as simple and cost effective as possible, so it can be easily abandoned and redesigned. Paper prototype consists of three main levels that we also test with other digital prototypes. The levels and game pieces were printed out on A4 papers. All the game pieces were cut out into small squares that would fit on the game grid. How the board game set looks like can be seen on the figure 1.1.

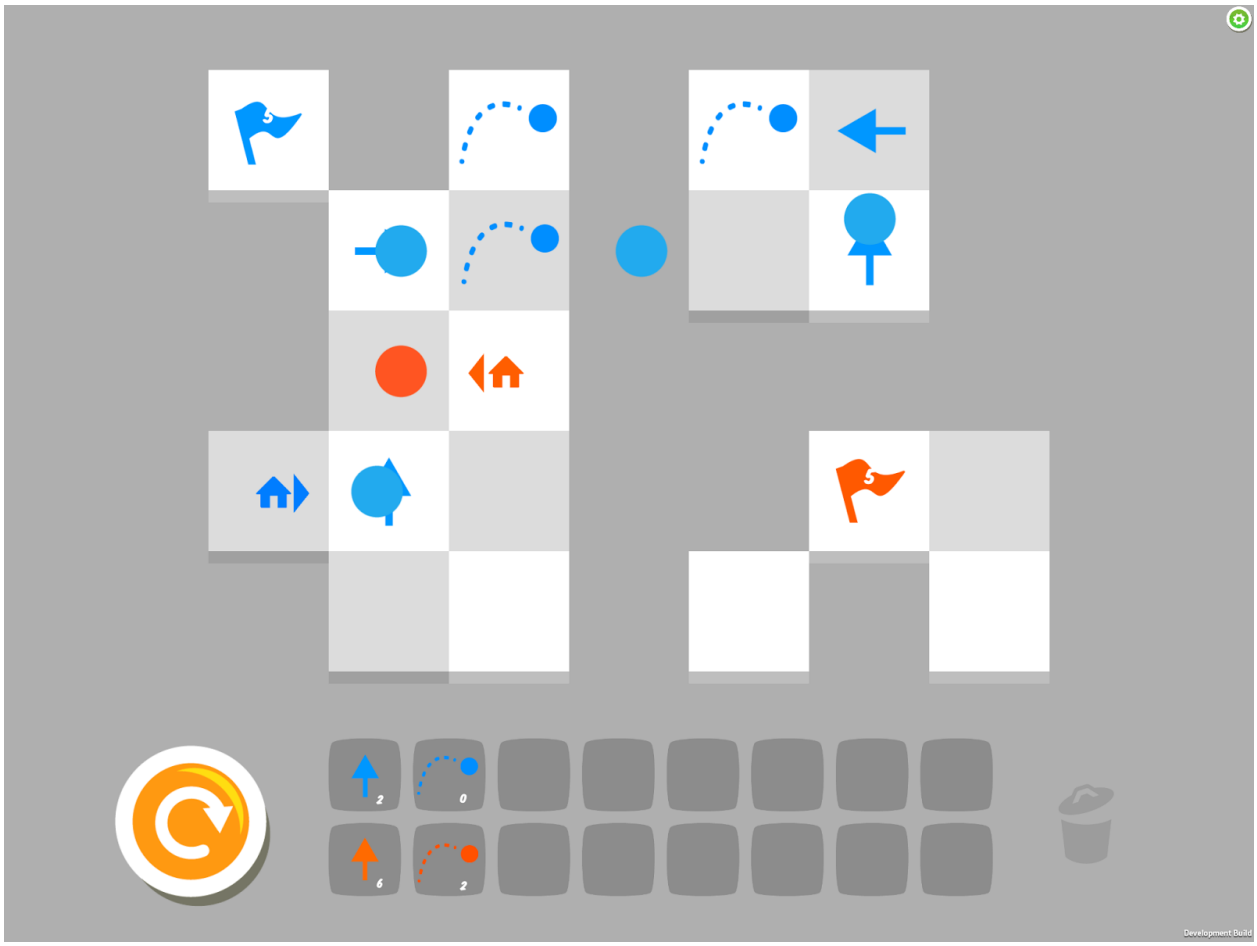

*Figure 5.* Screenshot of the low fidelity prototype

Early technical prototype without graphics, sounds and animations were made to reflect the first playable digital version of the game. The idea behind the early digital prototypes are similar to first paper prototypes — they have to be developed with minimal effort and cost effectiveness in

mind. Our digital prototypes are built with Unity 3D engine. This is also the platform the authors planned to use for the final product. Game elements use the similar robust aesthetics as the paper prototype. The design is based on the selected core game mechanics of the final product.

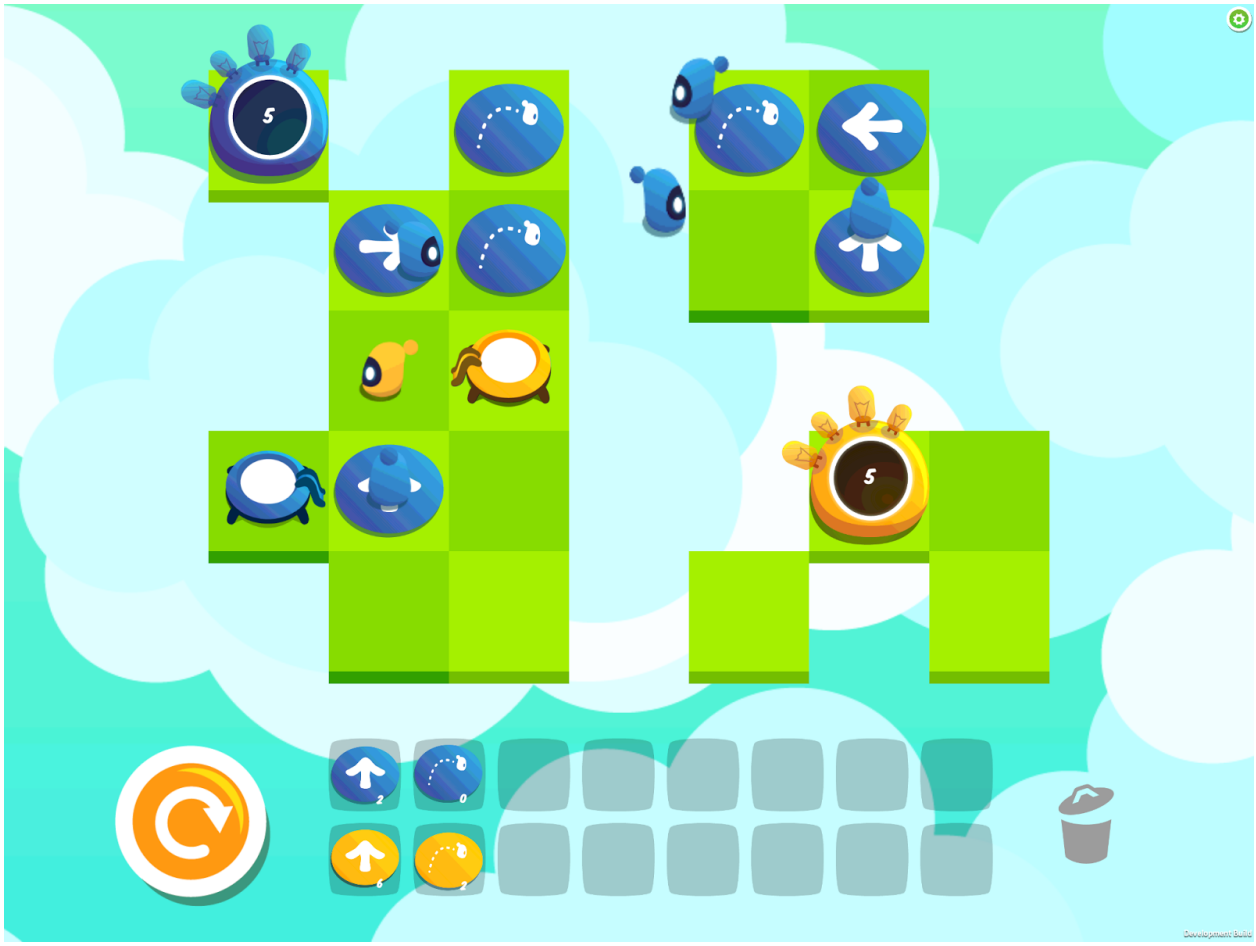


*Figure 6.* Screenshot of the high fidelity prototype

Early development high fidelity prototype with graphics, animations and sounds is closest to the real game. Testing with high fidelity prototype doesn't need much explaining if at all. Animations and graphic elements are self-explanatory to kids in that age group. Children who participated in the test said that it looks like the real games they play.

## Prototype game mechanics and Features

### Goal of the Game

Characters that spawn will have to reach the finish of the same color as they are. All finishes have counters on them, that count the number of correct characters that reach that finish. Finishes also have a number of correct colored characters required for that finish.

The game is over, when all the finishes reach their required count or when all characters are destroyed.

### PCA – Player Controlled Actors

#### Direction change

Changes the direction of game characters movement to the direction shown on the object. Only affects characters of the same color as the object.

**Player Interaction:**

- Can drag the object to the playing field onto an empty square from inventory.
- Can drag the object to another empty square on the level.
- Tapping on the object after it's been placed on the level will rotate the arrow's direction 90 degrees clockwise.
- Dragging the object out of the level boundaries and releasing will remove the object from the level and put it back into the inventory.

#### Color change

Changes the color of game characters to the color shown on the object. Only affects characters who are originally the same color as the object.

**Player interaction:**

- Can drag the object to the playing field onto an empty square from the inventory.
- Can drag the object to another empty square on the level.
- Tapping on the object after it's been placed on the level will change the color that will change the color of the character.
- Dragging the object out of the level boundaries and releasing will remove the object from the level and put it back into the inventory.

#### Jump

Makes the character jump over the next tile on its path of movement depending on which direction the character approaches the object. The direction of movement will not be affected. If there is an obstacle on the tile the character will land on, then the character will be destroyed.

**Player interaction:**

- Can drag object to the playing field onto an empty square  from the inventory.
- Can drag the object to another empty square on the level.
- Tapping on the object will do nothing.
- Dragging the object out of the level boundaries and releasing will remove the object from the level and put it back into the inventory.
- pushing the start/restart button releases fixers from spaceships

### NPCA – Non Player Controlled Actors

#### Fixer

Is an object that moves around the level. It's behavior is defined by other objects and obstacles like:

- direction change object
- color change object
- jump object

#### Start (spaceship)

Start is an object which creates fixers.

#### Finish (home)

Is an object where fixers have to go. All the finishes have counters that count down the numbers of inhabitants when they reach their home. Points will count down when fixer reaches the finish. every finish has it's own colour and it only lets in the same colored fixers.

## Participants in Study

### Testing with six to eleven year old kids

Altogether 18 kids between ages 6 and 11 were supposed to participate in tests. Hallway testing method was used to get a cross-section of the target group. Test group was planned to be gender-

equal — nine boys and nine girls. 15 of the kids were chosen randomly by Code2Kids' educational partner (Gustav Adolf Grammar School). Three children were found from authors' circles of acquaintances.

The selection of kids was based on our target group for the game Code2Kids — children whose parents have already invested into their kids' computer education. Reason for this is that customers of the product will be parents and they have to be willing to invest into means and content of their kids' digital education. Gustav Adolf Grammar School was ideal candidate, because it is one the first schools in Estonia that teaches programming and computer science to kids from the first grade. Our acquaintances were chosen by the same criteria.

For the same reason two kids were left out from our tests because their parents were not interested in investing into their kids computer education nor were their kids familiar or interested in the technology used in the tests. Although they would have been interesting subjects in the tests where we could have tested with kids with no prior knowledge how to use iPads we chose to focus on our target group and potential customers.

As time and budget are very important in a game development studio, we have designed the testing sessions keeping in mind the Jakob Nielsen's user testing with five people principle: "Elaborate usability tests are a waste of resources. The best results come from testing no more than 5 users and running as many small tests as you can afford." (Nielsen Norman Group 2000).

Nielsen also draws out a curve based on his earlier research that illustrates how many usability problems can be found with a number of users (Nielsen Norman Group 2000).

> *In earlier research, Tom Landauer and I showed that the number of usability problems found in a usability test with n users is:*
>
> $$N(1-(1-L)^n)$$
>
> *where N is the total number of usability problems in the design and L is the proportion of usability problems discovered while testing a single user. The typical value of L is 31%, averaged across a large number of projects we studied. Plotting the curve for L =31% gives the following result:*

(Nielsen Norman Group 2000)

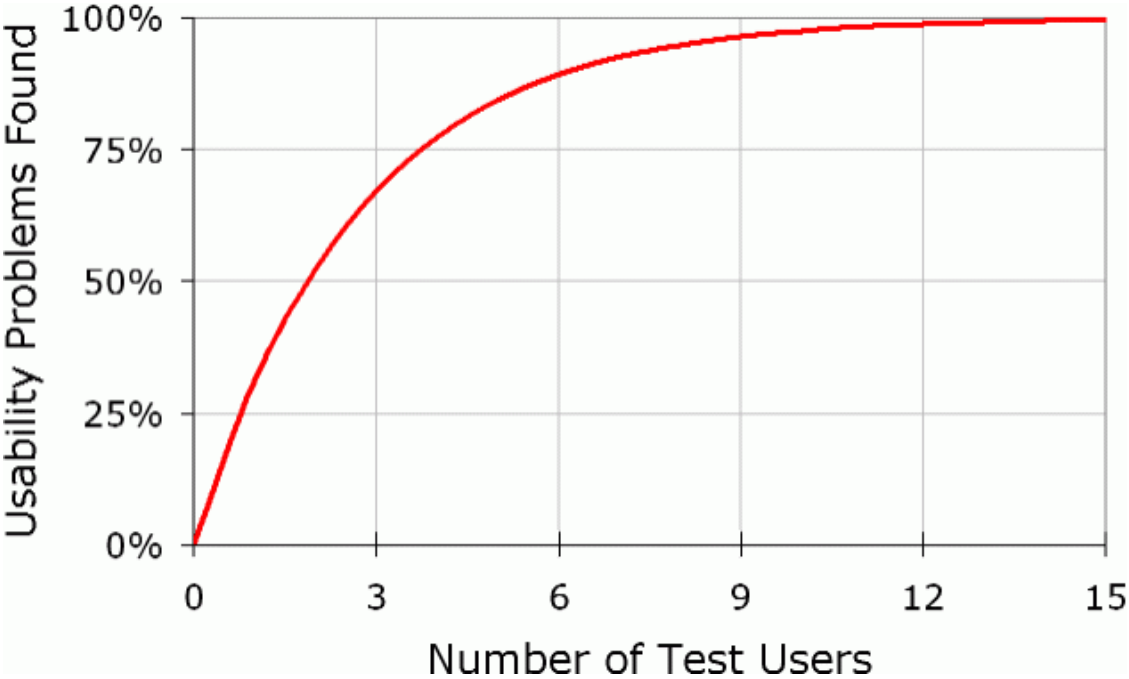


*Figure 7.* Source: Nielsen Norman Group 2000

According to Nielsen the first test gives nearly a third of the insight that can be gathered from all users from all tests. With the second test, information what you can learn has some overlap with the first test. The second user adds some amount of insight, but not nearly as much as the first user when insight grows from the zero to around 30% of the whole test. The third user gives similarly some more insight but a lot of information at this point has been already observed also with the first two users. As people are different some differences occur in users' behavior. As more and more users are added to the tests the less and less new insights can be gathered from the tests. (Nielsen Norman Group 2000)

On the figure 7 it is shown that to find all of the problems 15 tests have to be performed. Nielsen justifies his five user principle with the fact that 85% of the problems can be found with only five users and it will also help to redistribute the budget by doing more of the smaller tests rather than one big one. Basis for this is that prototype development is an iterative process and it is better to get the 85% of the findings with five users, make the improvements and test the prototypes again (Nielsen Norman Group 2000).

As our tests were made for three prototypes with different fidelity levels we used five kids for every prototype. Every test subject was allowed to test only one prototype.

## Pretest-posttest Survey

### Procedure and Context

Different parts of the test session were described and explained to kids at the beginning of every test session, so that they could feel relaxed and know that when the play time and questioning is over, then they would continue with their regular school tasks. One student or a small group of students will start with testing at a time. When he/she/they finish, next person/groups starts right away.

#### Testing session

Testing session consists of

1. Pretest questionnaire
2. Playtesting and observation
3. Posttest questionnaire (same questions)
4. Additional posttest question(s)

### Pretest and posttest questions

Questions were selected according to Csikszentmihalyi's work (1990) and Chen's interpretation of it (2007).

As not all of these Csikszentmihalyi's criterias are needed for experiencing the Flow, authors made a selection and created questionnaire according to previous researches and (Csikszentmihalyi 1990; Chen 2007)

*Table 2*: Description of the components of the flow by Csikszentmihalyi and Chen, aligned with authors' description and pretest/posttest questions to detect it in the prototype.

| Csikszentmihalyi | Chen | Roost & Rebane |
|---|---|---|
| We confront tasks we have a chance of completing; | A challenge activity that requires skills | **The challenge** - How difficult do you think the game is? |
| We must be able to concentrate on what we are doing; | The merging of action and awareness | - |
| The task has clear goals; | Clear goals | **Clear goals** - How clear is the goal of the game? |
| The task provides immediate feedback; | Direct feedback | - |
| One acts with deep, but effortless involvement, that removes from awareness the worries and frustrations of everyday life; | Concentration on the task at hand | **Concentration on the task at hand –** How much does the task at hand need concentration?<br><br>**Control question** – Do you want a tip how to continue? |
| One exercises a sense of control over their actions; | The sense of control | **The sense of control** - Do you understand the interaction and the controls of the game? |
| Concern for the self disappears, yet, paradoxically the sense of self emerges stronger after the flow experience is over; and | The loss of self-consciousness | - |
| The sense of duration of time is altered. | The transformation of time | **The transformation of time** - How much time you have spent on this playtest? |

## Questions that are asked from test participants

To see whether initial levels are designed to be challenging enough authors asked from every participant: **How difficult do you think the game?**

If game is not challenging enough it gets boring fast. On the other hand, those games that are too difficult to succeed, will be dropped too. Although level balancing is not this paper's topic, it is better to know if this would be problem that might offset the test results.

Authors want to be sure weather test participants know the goal of the level and game as whole. Thus the question **"How clear is the goal of the game?"** was asked.
As Csikszentmihalyi has said "Goals transform a random walk into a chase. You need clear goals that fit into a hierarchy, with little goals that build toward more meaningful, higher-level goals." This is one important part of measuring the flow or engagement in game (Geirland, 1996).

Mostly to clarify if high fidelity prototype's design aesthetics is up to elementary schools kids' expectations, a question about visual design, look and feel was added: "**How do you like the look and feel of the game?"** to see in analyses phase if the children's feel about the aesthetics of the prototype correlates with something else asked. This question was added also for business reasons to test cognition of the authors' visual style in target group.

One difficult but important topic to cover is concentration. **"How much does the task at hand need concentration?"** was asked from children and this was the only question that couldn't answer by pointing to the Smileyometer. Thus five black circles are shown, each a bit bigger than previous - these are the possible sizes of the concentration children need to choose one from. It was explained to children that this is the scale between utter maximum and minimum they could possibly think of.

To make things clear for observational experiment of testing, question: **"Do you understand the interaction and the controls of the game?"** was asked. So authors could help children in test to get comfortably started with it and in the posttest questioning authors could estimate, if elements of interaction were intuitive and learnable enough.

After finishing the test and all the previous questions are asked again and one additional question was asked: **"How long do you think that game test lasted?"**. Children's answers were saved and their real time of test is marked down and said to them also.

# RESULTS

## Results from the Playtesting

### Overview of the Responses

Overview is understandable best when presented as line-charts. Lines in charts are illustrative, to emphasize the trend. Only the data point in charts are used in research to make conclusions. Childrens' responses to Smileyometer in pre- and posttest questions are grouped. Dynamics and differences between these two, if exists, is commented below charts.

### Answers to the pretest-posttest questionnaire

#### Questions 1: How difficult do you think the game is?

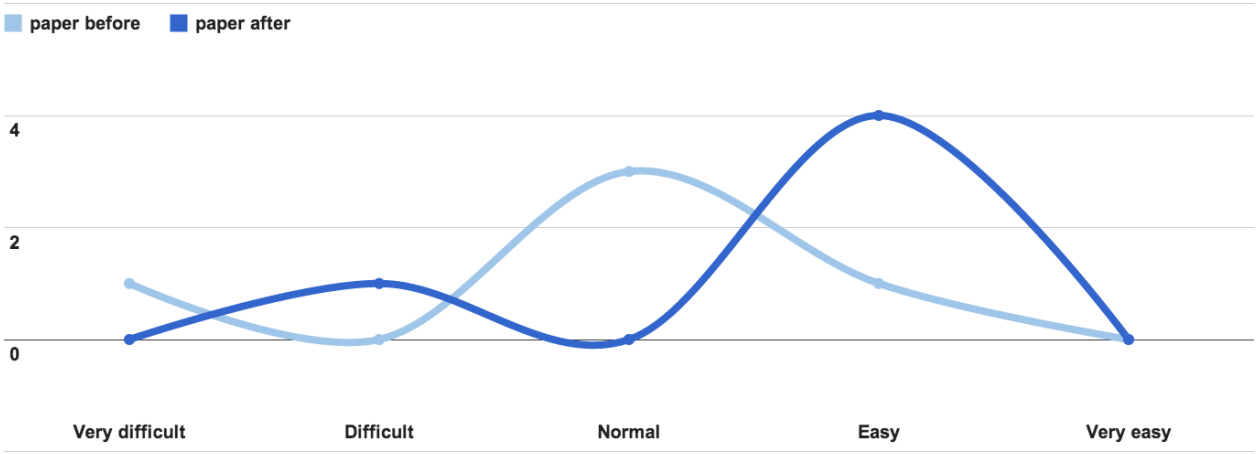


*Figure 8*. Answers to pre- and posttest question: "How difficult do you think the game is?" (Paper prototype)

Most of the children said before test with paper prototype that its difficulty seems to be normal (smiley nr 3), but after playing it, they considered it to be "Easy" (smiley nr 5).

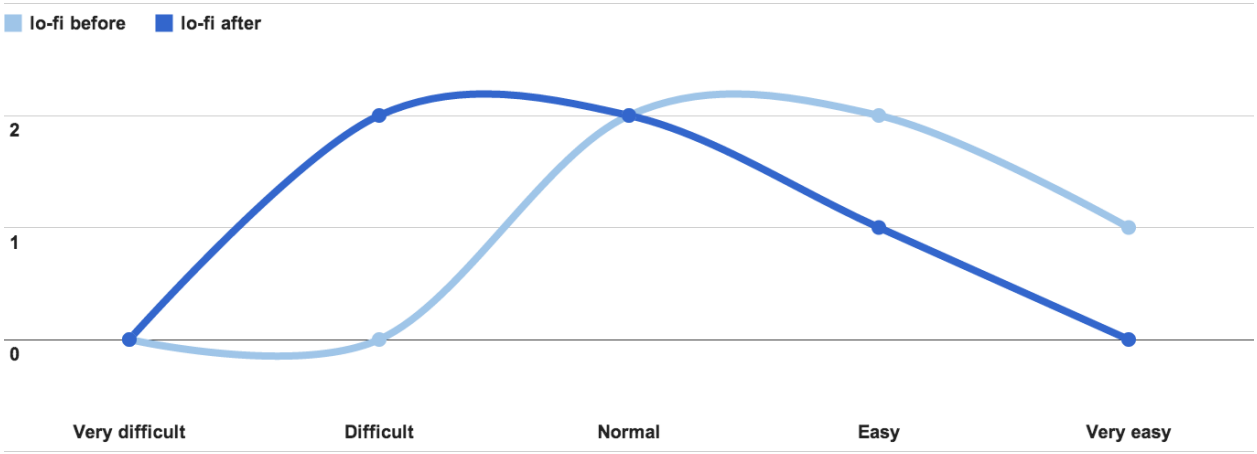


*Figure 9*. Answers to pre- and posttest question: “How difficult do you think the game is?” (low fidelity prototype)

When testing gameplay with low fidelity prototype, most of the children corrected their opinion about difficulty of the game after they had tried it for a while — they had underestimated it.'

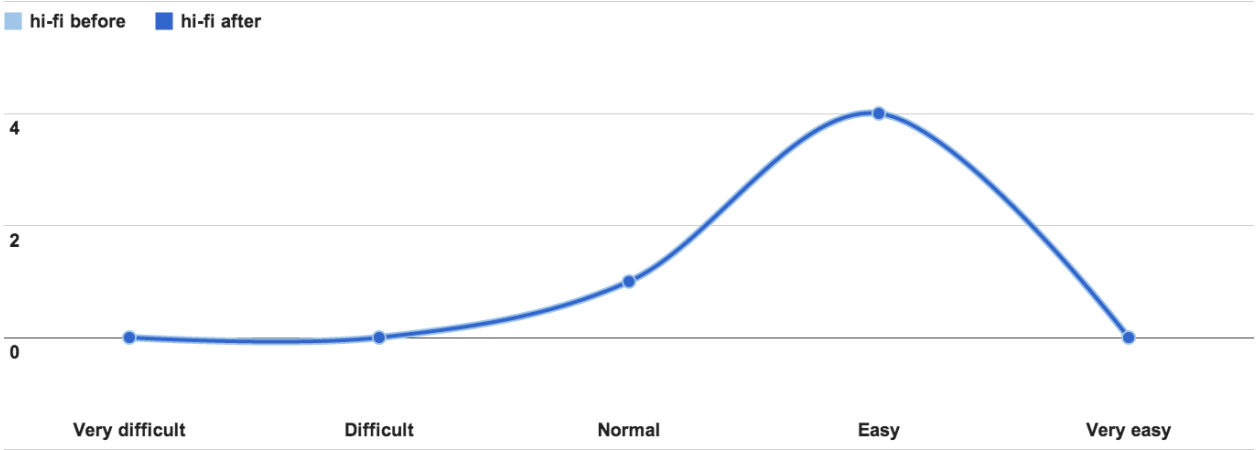


*Figure 10*. Answers to pre- and posttest question: “How difficult do you think the game is?” (high fidelity prototype).

The question about the difficulty got exactly the same answers in pretest and posttest questioning - four out of five in that test group considered it to be easy (smiley nr 4).

**Questions 2: How do you like the look and feel of the game?**

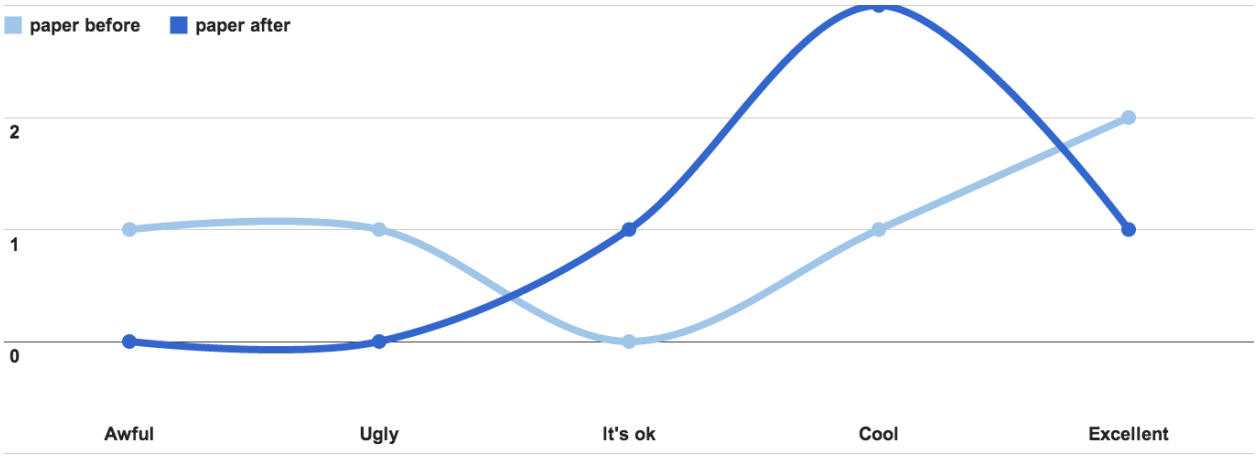


*Figure 11*. Answers to pre- and posttest question: "How do you like the look and feel of the game?" (Paper prototype).

As every child was only testing one prototype fidelity level of the same game, then they didn't have comparable experience and thus so positive results on paper prototype aesthetics - after playing it all of them said that it looks "cool" (smiley nr 4) and only one thought it to be "ok" (smiley nr 3). Before playing it "awful" and "ugly" were also chosen.

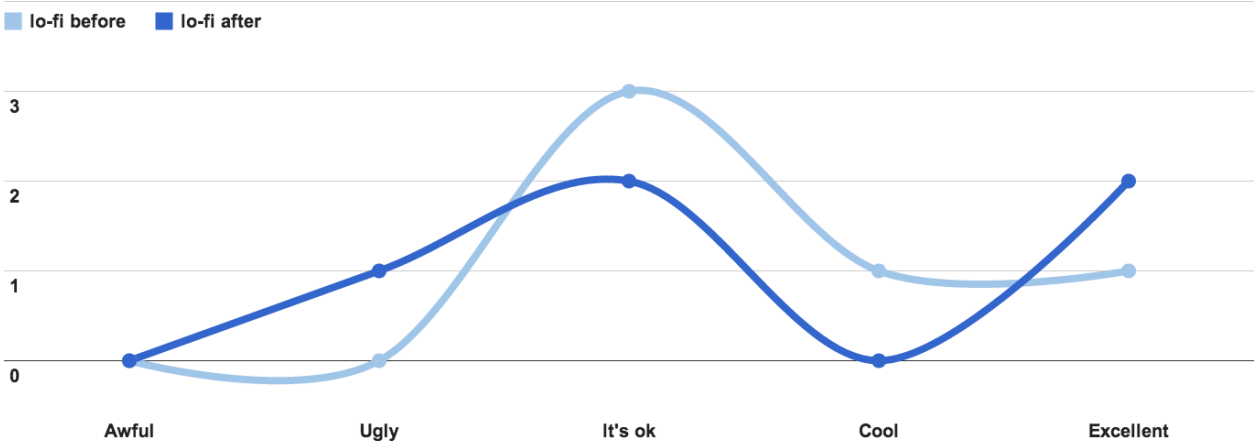


*Figure 12*. Answers to pre- and posttest question: "How do you like the look and feel of the game?" (low fidelity prototype)

When describing the look and feel of the low fidelity prototype most kids chose the medium option (smiley nr 3) before they had the chance to play it. After playing the game, one of them reconsidered and called it ugly, another respondent changed his/her opinion to "excellent" (smileys nr 5).

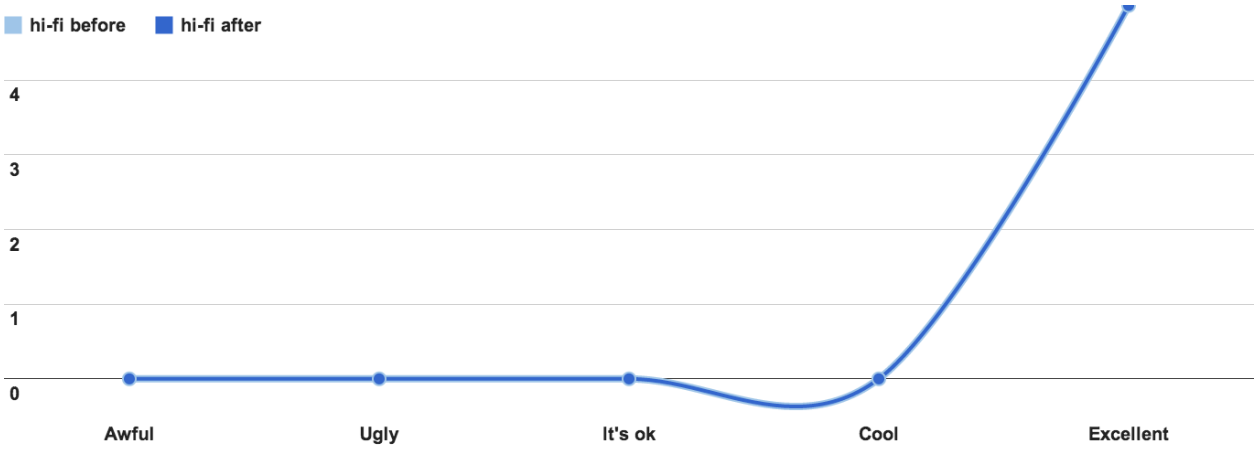


*Figure 13*. Answers to pre- and posttest question: "How do you like the look and feel of the game?" (high fidelity prototype)

When asked about the look and feel of the high fidelity product like prototype, all of the test group members called it excellent, before and after testing it.

**Questions 3: How much does the task at hand need concentration?**

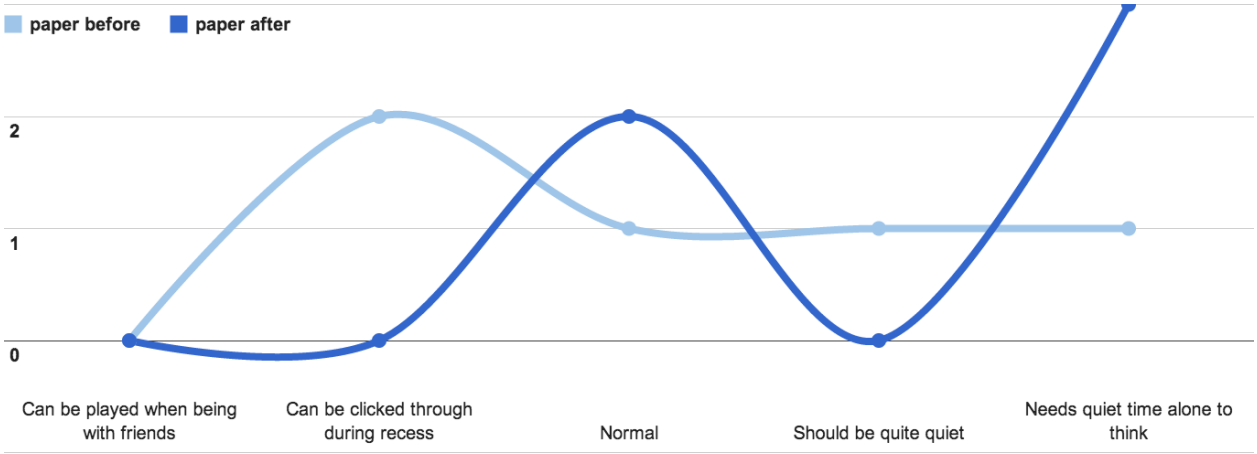


Figure 14. Answers to pre- and posttest question: "How much does the task at hand need concentration?" (Paper prototype, Ballometer)

It appeared that the question about the need of concentration on the task in game was quite difficult to understand for younger kids in test group. Never the less, results in pretest and posttest questions are quite different - from which we can reach to conclusion that using paper prototype in this age group for game testing needs more explaining and thus more time is needed to do the tests.

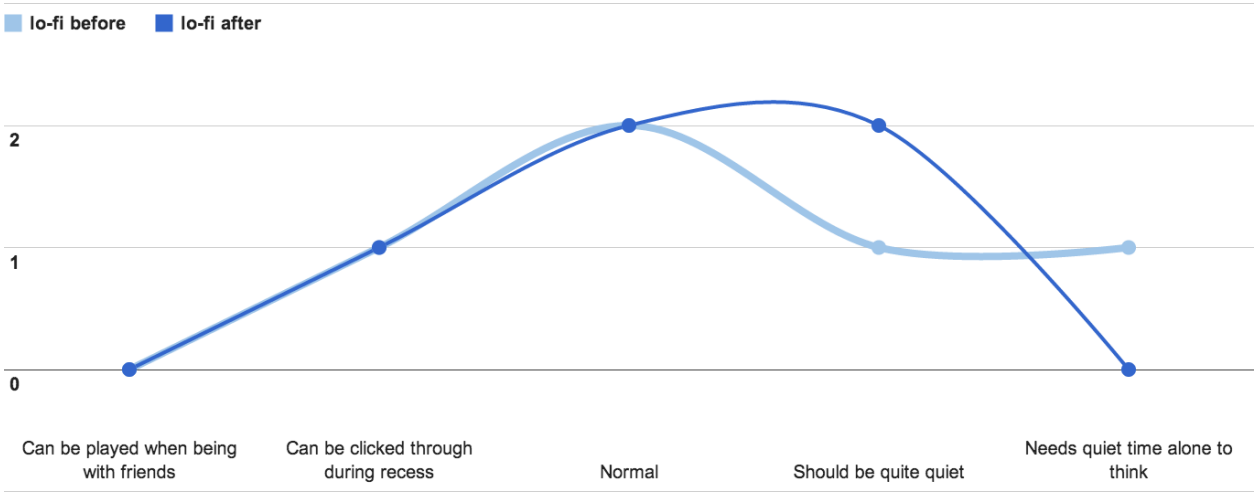


*Figure 15.* Answers to pre- and posttest question: "How much does the task at hand need concentration?" (low fidelity prototype, Ballometer)

Children interviewed didn't answer much differently to this question when they were asked again after the test session, only one respondent changed his/her opinion about needed maximum concentration. One reason why there isn't clear pretest/posttest dynamics, might be that question was not clear enough for them and they chose to say the same thing after the testing session too.

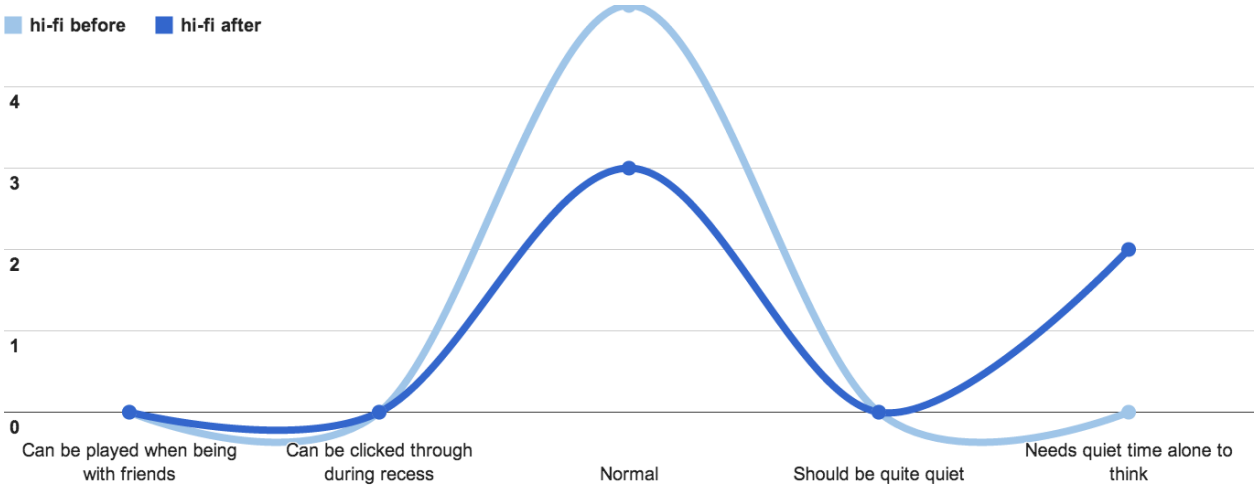


*Figure 16.* Answers to pre- and posttest question: "How much does the task at hand need concentration?" (high fidelity prototype, Ballometer)

All test persons considered the need for concentration to be normal (might be the psychologically secure choice in the middle, meaning "I don't know") when looking at the low fidelity prototype, two out of them changed their answer to maximum, meaning that they considered levels of the game too difficult to be played in context of distractions.

All-in-all authors consider this question about concentration to be too difficult to this age group as it

needed the most explaining and examples to explain kids in test group how they should evaluate the concentration

**Questions 4: How clear is the goal of the game?**

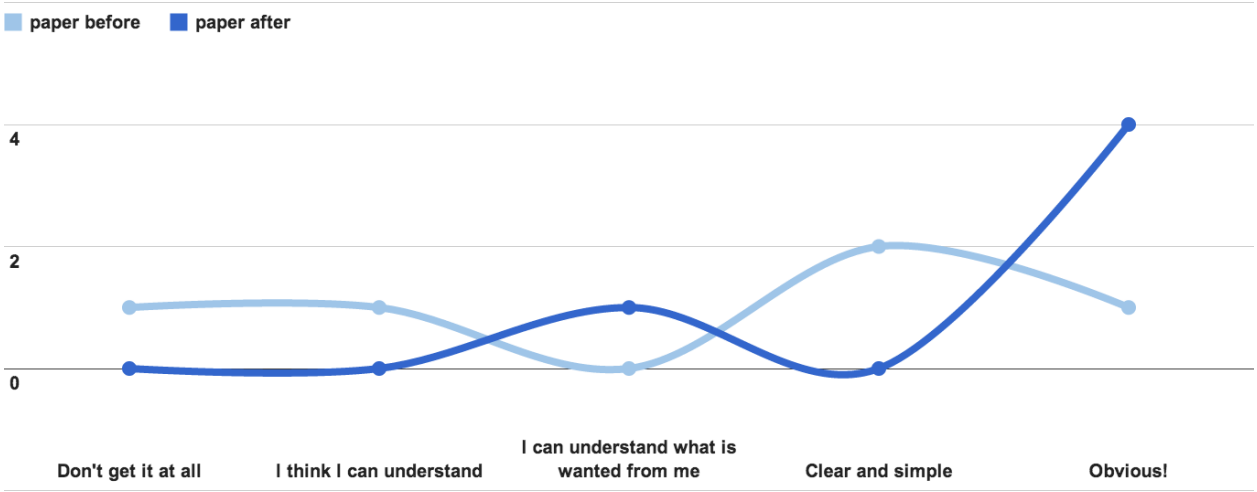


*Figure 17.* Answers to pre- and posttest question: "How clear is the goal of the game?" (Paper prototype)

Some children got the point of the board game really fast but others needed explaining. After test most of them could play it without instructors presence.

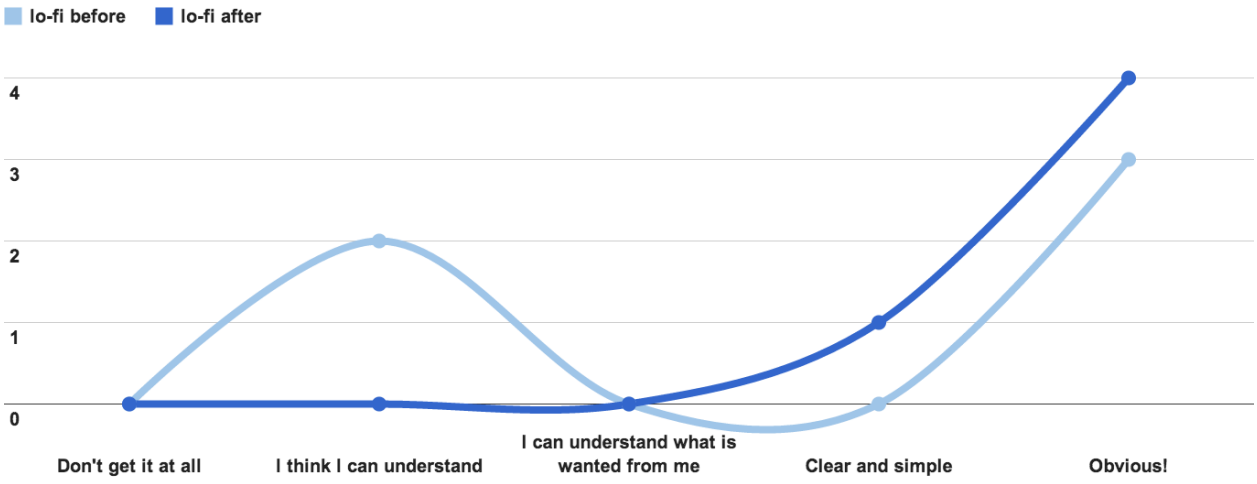


*Figure 18.* Answers to pretest and posttest question: "How clear is the goal of the game?" (low fidelity prototype)

The goal of the game was clear for most of the children before touching the test device, the ones who were doubtful at the beginning, got it really fast after trying to play it a bit. Posttest question got answer "Obvious" in four out of five cases and only one "Clear and simple".

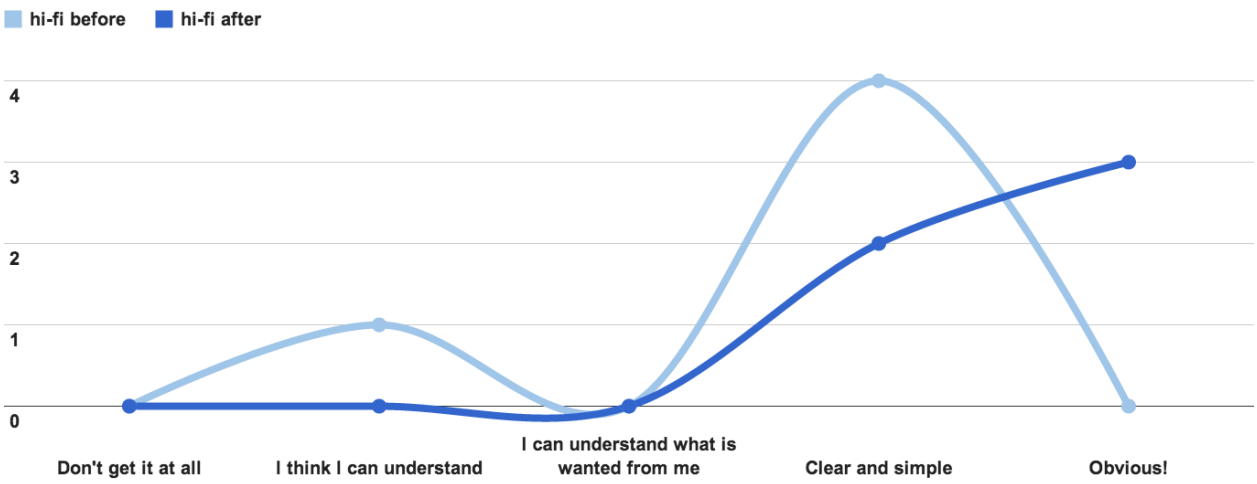


*Figure 19.* Answers to pretest and posttest question: "How clear is the goal of the game?" (high fidelity prototype)

Before playing the game most of the test group members said that the goal of the game is "Clear and simple" and only one child gave it a "2 - I think I can understand". After playing it and/or some explaining three of the test group gave maximum points ("Obvious!") to question about the goal of the game and rest of them chose "Clear and simple" .

**Questions 5: Do you understand the interaction and the controls of the game?**

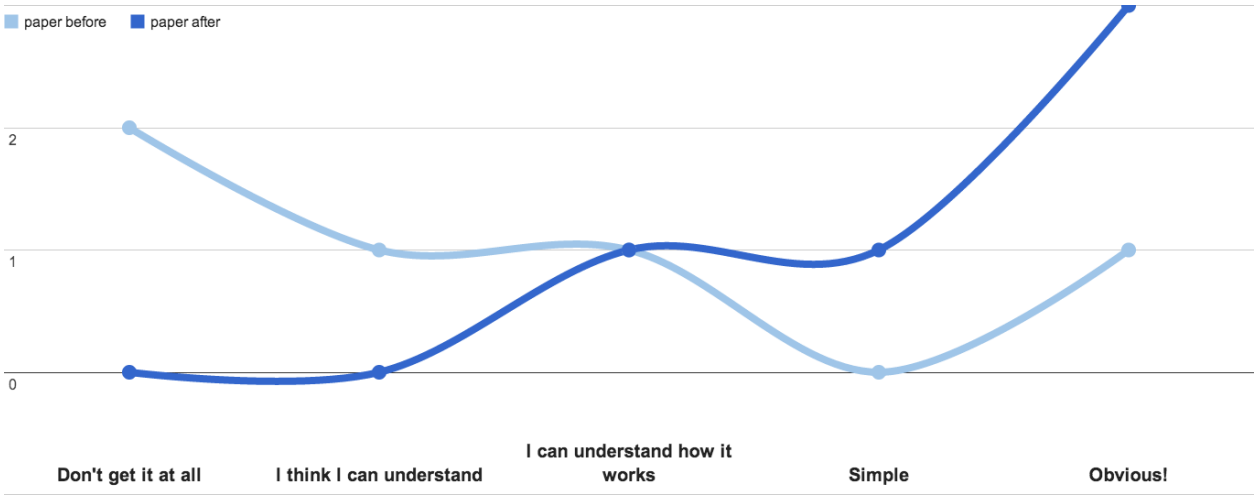


*Figure 20.* Answers to pretest and posttest question: "Do you understand the interaction and the controls of the game?" (Paper prototype)

By only looking at the board game (paper prototype) and it's elements only one said it is obvious, others were more doubtful about how to play the game. After playing the game, most of

tested kids understood it very well.

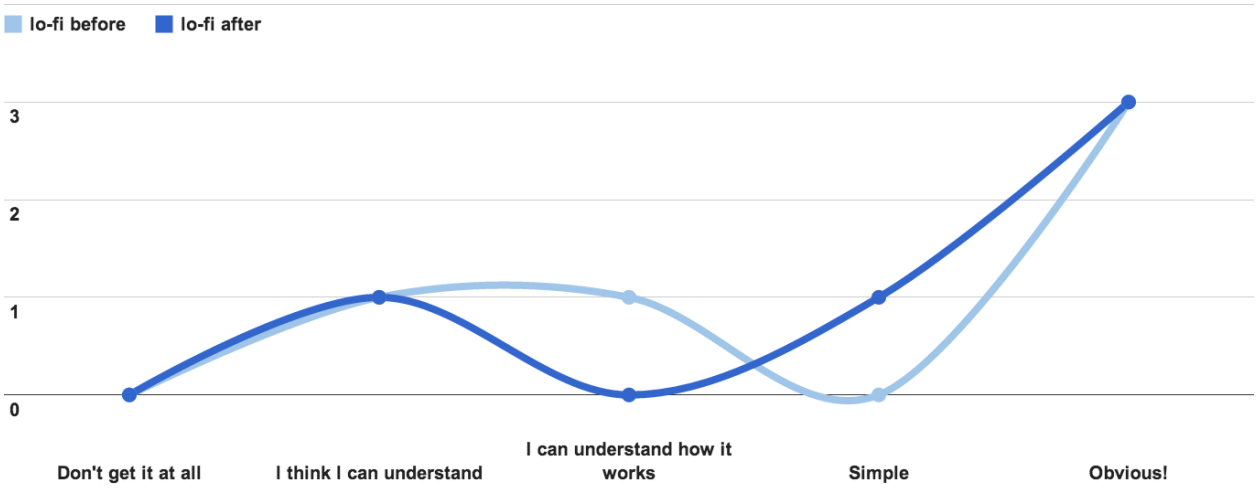


*Figure 21.* Answers to pre- and posttest question: "Do you understand the interaction and the controls of the game?" (low fidelity prototype)

When low fidelity prototype was shown to children, most of them said it is obvious how the game should be played, some said it is simple or at least they can understand it. After playing session with the low fidelity prototype the understanding about this game's interaction was improved or "obvious" as they thought before playing it.

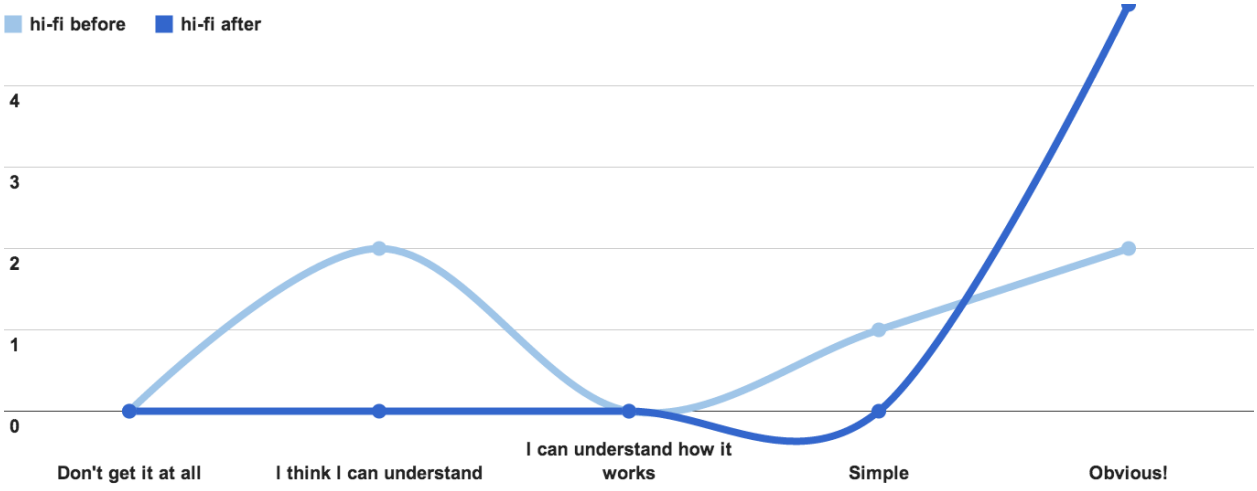


*Figure 22.* Answers to pre- and posttest question: "Do you understand the interaction and the controls of the game?" (high fidelity prototype)

Only one respondent chose answer nr 2 ("I think I can understand") out five before trying out the high fidelity prototype. All the answers about the game controls and interactions were highest (5- "Obvious") after playing the game.

**Questions 6: How much time you have spent on this playtest?**

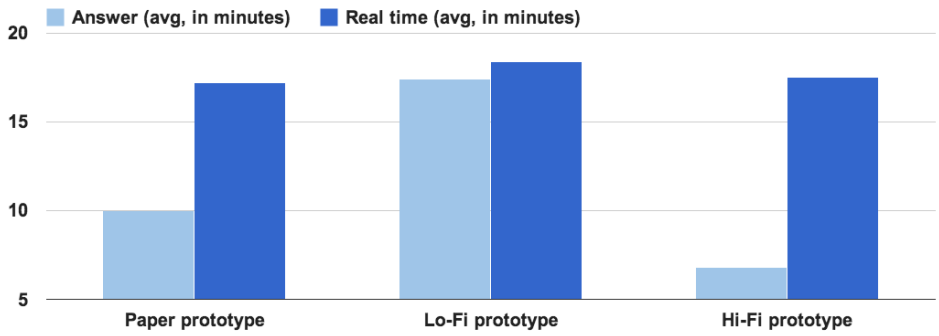


*Figure 23.* Answers to pre- and posttest question: "How much time you have spent on this playtest?" (All prototypes)

Average play times with every prototype are compared with childrens' cognitive estimation of time passed by while playing with prototype. As one of the most important characteristic of defining engagement is inability to assess the amount of time that has passed by while doing particular immersive task (Csikszentmihalyi, 1990).

## Key Findings From the Pretest–posttest Survey

Authors find that the method of pretest posttest survey can be used with children in the age from 8-11 but asking younger children to rate something in the scale from one to five is not so good idea. Questioning sessions showed that children in the age six to eight can not give such feedback or if they are asked to, they just say something randomly or always pick the choice in the centre (see Smileyometer: figure 2).

It was difficult for children to measure or evaluate the time spent on current test - it appeared that the passing of time is not something that children in that age group would feel or would be able to estimate.

Authors' third finding is that if the pretest/posttest survey is going to be used with children, the researchers must prepare questions as simple as possible, otherwise children don't understand what is wanted from them or they just don't remember the first half of the long question by the time researcher reads the second half of it. Bigger number of simple questions makes answering easier and overall less time consuming than having long explanation after every difficult

question.

Preparing pre- and posttest survey as interviews with children in their own environment takes time, experience and one has to be prepared for every kind of surprises. Authors consider the first test session as biased because the lack of knowledge and experience to work with children in their own environment that might include lots of distractors as noise, chatting classmates etc. Because of beforementioned reasons first test person's data is considered as corrupted and this is removed from all the data tables, analyses and calculations.

## Observation notes

During the tests authors observed children to discover how they felt about the prototype they were testing at the moment. Reason for that was to discover all the emotions that can't be asked as interview questions and won't be told also. Kids' facial expressions and emotions were described also in the comments because kids' faces were not allowed to be filmed by school. Some encouragement to think-out-loud was done by authors and some additional questions were asked e.g. what made them feel so anxious; what bothered them in the game; what were they thinking at the moment; what part of the task were they trying to resolve in their head. Only few kids were very shy and afraid to talk out loud, but most of them answered the questions or started talking and sharing their experience. The tone of the conversation was friendly and cheerful and not formal, academical or patronizing. Most of the kids are very happy when they can participate in playtesting session but probably this is mostly due to the fact that they can skip the class and play a video game instead. Some observation notes were also made by analysing the playtesting videos later on. Observation notes about kids' emotions are in appendixes 2 - 4.

### Paper prototype observations

During the playtesting with the paper prototype most of the kids seemed to be somewhat distracted because there was so much extra activity happening while playing with the prototype. Kids themselves had to do only the same actions as they would do in the digital version of the game but all of the other actions and animations had to be played out by the person who conducted the test. For example when kids said "Play" the person who conducted the test started moving the game pieces based on the rules of the game. Some levels had multiple pieces that had to be moved one by one instead of moving them together, which made this particular game element quite time consuming for kids and prolonged the playtime considerably.

Due to these distractions the potential of flow to happen was much lower compared to low

fidelity and high fidelity digital prototypes where there are very little distractions while playing from the game itself.

In some cases kids got confused by the game pieces because they didn't quite understand the rules how the pieces can be used. Small tutorials were made for playing with the paper prototype but this was not always enough. Through all of the tests with all prototypes kids used trial and error to explore the rules and possibilities of the game. While feedback from the game is imminent with the digital prototypes, the game mechanics have to be played out with the paper prototype and this action takes some time and becomes a distraction.

Something very interesting came also out from the observations that surprised the authors or came out as unexpected. When the topic of fun and engagement of games is discussed one would think that kids are cheerful and energetic while playing. Observations proved that this is not always the case throughout the playing session. During most of the tests when challenge increased and more thinking had to be done kids got quieter and more focused. Their faces tended to tense up and some even scrawled or made a face that could be interpreted as angry. Some made thinking noise like "hmm" or sighted loudly — not to be interpreted as signs of boredom but rather signs of concentration, "positive" frustration that meant that levels they were playing created a challenge.  But that changed quickly when kids won the level. Quickly the tension from the face was gone and was replaced with very positive emotions like smiling and cheering.

In some cases kids reacted quite loudly to scenarios like death of the characters. In one instance one 8 year old girl shouted loudly "No" and covered her mouth and face with both of her hands.

Authors also used a test question during the gameplay. When levels got hard and kids had to think the hardest one of the observers asked if kids want a tip how to continue and get unstuck. Purpose of this was the see how focused kids are during the playtest session and if the state of flow can be observed. Kids usually answered quite quickly to this question during the paper prototyping session.

### Low fidelity prototype observations

With the low fidelity prototype it could immediately be observed that kids were much more focused. They didn't talk very much and were concentrating on playing the game and figuring out the solutions. They were also much less distracted by the surroundings.

Digital prototypes were tested in pairs and it could be observed that kids were very focused on their own game. There was very little "peaking" or cheating while playing. Kids wanted to figure out the solutions by themselves.

The low fidelity prototype had some bugs in the game that could be seen visually. For example the direction gates changed their appearance when they were directed left. There was very little confusion or questions about this particular bug while playing. Kids usually saw this and got a little distracted but continued to play the game with the bug and didn't notice it in the later levels.

Again it was observed that kids got cheerful and were smiling when they solved the problems and won the levels.

The test questions was also asked during low fidelity game playtesting. It could be observed that it took some time longer to take their focus away from the game and answer. But all of the kids still answered to question.

### High fidelity prototype observations

During the playtesting of high fidelity prototype kids seemed to be much happier and cheerful. Especially during the first levels of the game. They also made many funny remarks between them and were also much more talkative. Reason for that might be that kids in this testing session were little bit older than in other testing sessions or that the cheerful look and funny looking characters brightened them up.

When levels got harder kids' faces started to turn more serious, focused and tense. Playing session got also a lot quieter with only few remarks to each other or to the observers. With hardest levels of the game it was observed that kids got even quite frustrated or even crossed their hands and took in a defencive position. Still all of them wanted to finish the levels that the game provided. Again mostly positive challenge was observed.

Again it was observed that kids got cheerful and were smiling when they finished the levels and got all of the characters home correctly. After the brief cheering they started playing the next levels and again got more serious and focused.

In some cases it became a grind for the child although they didn't want to accept the failure and carried on playing. Finally one child got very frustrated and stopped playing when asked if she wanted to quit after already several times. It was obvious that she was frustrated and challenge

got too hard for her.

During the high fidelity prototype playtesting session a high level of flow could be observed. When asked the test questions some of the kids didn’t notice the questions. In some cases they didn’t also react to their names at first. But after a while when there was a small break after they pushed the “play” button they were able to answer the questions again.

# DISCUSSION

All the pretest-posttest questionnaire results and observation conclusions about the optimal prototyping fidelity level indicate that authors assumptions were correct. Paper prototype can't be used to detect if the flow appears or not. The process of playtesting itself can be fun for kids but that doesn't mean that the game is fun.

Paper prototypes can be built to get the first feeling of the game — how the game carries out and which kind of problems may rise when using different game mechanics. Also what kind of game mechanics are better to use. Paper prototype gives also the opportunity to change the game rules very quickly. It is also very easy to create extra game elements and pieces to try out different functionalities and rules. As the results show that paper prototype is not enough to test how engaging the games is due to the simple fact that there are too many distractions throughout the playing session that are not present in the digital games. Paper prototypes are very good to be used in early iterations where game mechanics are not fully formed yet. It can be used for quick experimentation and can be good source of "happy accidents" that may give game designers and developers new ideas. From the analysis of responses from the pretest and posttest surveys it can be seen that engagement level was lower with the paper prototype, although kids thought that the they spent less time playing it compared to how long they actually did. Problem here was that kids were quite young and their sense of time may be unreliable.

High fidelity prototypes are expensive and time consuming to build, but there is no doubt that these give the best results when playtesting them. Best results from the testing regarding flow were achieved with high fidelity prototypes. Also these prototypes can answer to more questions regarding art style, animations, game mechanics, functionality and quality. As one of the goals of the thesis was to find the optimal fidelity level for the prototype, that gives usable results regarding engagement and doesn't cost too much, then authors consider the high fidelity prototype being not worth the value invested in it. As explained before these kinds of prototypes

take a lot of effort to make. Every level of fidelity grows the scope of the project exponentially. When comparing the test results then it can be seen that extra efforts like better animations, game art and sounds had some impact on how kids played the games, but somewhat similar results could also be observed in the low fidelity prototype. Thus it can be concluded that the moment when there is a digital playable prototype with placeholder art it can be used to test the engagement of the game. All prototyping efforts before that can be used to experiment with the game mechanics and functionality. All efforts that come after that like adding game art, animations, sounds can be already used for more precise testing of certain game elements. That includes also observing if playtesters are having fun and what elements especially make the game engaging. One point to consider here is also the fact that when state of flow was achieved kids tended to become unresponsive. Due to that talking out loud mentality can become harder to achieve.

Authors consider the low fidelity prototype to be optimal and neutral in the context of game testing and measuring the flow with elementary school kids. There has to be bare representation of testable game mechanics and few visual or semantically correct cues for the users testing it.

It can be also mentioned that research helped authors to continue with Code2Kids development process and is good starting point for next playtesting sessions with the same game or with future developments. Research aims were achieved to the extent that it is a good point for further researches on similar topics, weather in more detailed way or using more quantitative research methods and bigger test groups. See paragraph “Conclusions: further research”.

## Analysis of the Responses

- Engagement level was lower with the paper prototype, although kids thought that the they spent less time playing it compared to how long they actually did.
- Engagement was quite high with the low fidelity prototype.
- Engagement level was the highest with the high fidelity prototype.

Taken all together the analyses show that testing with paper prototype was too much out of test groups expectations of what a mobile games should look and feel like and thus the feeling of time passing by and the actual time spent on it was so different. The medium itself and and processes of using it was interesting yet confusing to them and for this reason the posttest interview question about time expenditure is not considered relevant in measuring the flow.
As we can see from the chart, childrens’ estimates and play time correlate when we compare

averages from testing with low fidelity prototype and estimates do not correlate with play time when prototype has nice user interface with all the decorations, animations, visual and sound effects.

As the result of this paper, authors recommend to build game prototypes that have at least low fidelity graphical interface and functional game mechanics that can be tested in the real game environment / on the original hardware. When developing functional prototype, adding graphical details and audio effects don't add as much to user experience as semantically correct game elements and understandable characters do.

The results after playing the high fidelity prototype of the game were obvious too as it was game well polished and tuned, almost like a real products they have used to play.

# CONCLUSION

As the result of the current research, authors suggest to every game development company or studio that plans to create a game that targets elementary school kids - use a lot of testing with prototypes in any fidelity level from paper prototypes and sketches when in the ideation phase and when user traction is proved. Develop more game-like prototypes, add functionality and fidelity levels to it and keep on testing with kids. Every testing session with higher fidelity prototype gives the studio a lot of new information about the game in development - the more closer the look and feel of the prototype goes to the real game, the more accurate are all the conclusions that team can use not only in development of the game mechanics and design, but also in the business model tailoring.

When test-driven game development process reaches to testing with almost fully functional low fidelity prototype, then authors suggest to slice the testing process into many smaller sessions, so that team can improve the prototype by every session´s results. Small improvements of the low fidelity prototype can make huge impact to how easily children understand the objective of the game and it´s controls. At some point, after adding enough details and functionality to the low fidelity prototype, it is irreplaceable source of information when used in continuous test approved development and at the same time, the effort put in it is optimal - not a line of code is wasted (what might happen, when building the high fidelity prototype).

Authors suggest to game studios to get introduced to “throw-away prototyping” mentality and optimize the development cost according to it. It is not a game development prerequisite to spend time and effort to build a decent high fidelity prototype to test ideas of the game that might turn out to be not attractive and worthless to target audience or with not so much potential to be a successful when developed and launched as a product.

## Further steps for next researchers of the topic

- Draw more universal conclusions by trying the same testing methods with different game

types on kids in same age group;

- See if this testing method applies for older age groups also and what are the differences in evaluating the flow and fun in games;
- Finetune the prototyping fidelity level that helps to test most accurately the needed flow factor in the planned game;
- Repeating the same research to compare the results or doing a follow-up research to see how the results are changed in time and development of these same kids.

## More information is needed

Authors think that more information is needed on kids' previous gaming experience, particularly in educational games area. As kids come from different backgrounds their gaming habits are different, so are rules for gaming at homes or absence of these rules; and also their access to internet, electronic devices and gaming platforms is different. It is not taken account how strong is this kind of contextual influence to children who participated in the tests.

## ACKNOWLEDGEMENTS

We thank the Gustav Adolf Grammar School teachers and specialists Ingrid Maadvere and Kristi Rahn who have given us valuable input while developing the game, and also finding the time and providing the opportunity to do the playtesting at their school. Also our gratitude goes to Indrek Vainu who arranged the playtesting with his kids.

# REFERENCES

## Books and articles


1. **Read, J. C., MacFarlane, S. J., Casey, C.,** 2002, *Endurability, engagement and expectations: measuring children's fun, In:* Proceedings of interaction design and children, Eindhoven, Shaker Publishing, p 189–198
2. **Sim, G., MacFarlane, S., Read, J.,** 2005, *All work and no play: Measuring fun, usability, and learning in software for children*, Preston, Great Britain: Elsevier Ltd
3. **Chen, J.,** 2007**,** *Flow in games (and everything else),* In: Communications of the ACM, New York, NY: ACM
4. **Csikszentmihalyi, M.,** *1990, Flow: The Psychology of Optimal Experience*, New York, NY: Harper & Row, Publishers, Inc
5. **Csikszentmihalyi, M.,** 1997, *Finding Flow*, New York, NY: Basic Books
6. **Schell, S.,** 2008, *The Art of Game design: A Book of Lenses*, Boca Raton, FL: CRC Group, Taylor & Francis Group, 82, 83, 130
7. **Koster, R.,** 2005, *A Theory of Fun for Game Design*, Scottsdale, AZ: Paraglyph press inc.
8. **Fullerton, T**., 2008, *Game design workshop*, Burlington, MA, USA: Morgan Kaufmann publications
9. **Rautiainen, M.,** *et al.*, 2012, *Ubiquitous Framework for Creating and Evaluating Persuasive Applications and Games* — In: Grid and Pervasive Computing Workshops, Berlin: Springer Berlin Heidelberg, 133–140
10. **Ries, E.**, 2011, *The Lean Startup*, London: the Penguin Group
11. **Dimitrov, M., Rumrill, D.**, 2003, *Speaking of Research: Pretest-posttest designs and measurement of change,* In: WORK: A Journal of Prevention, Assessment, & Rehabilitation, 2003, 20, 159–165


## World Wide Web posts, articles, videos


12. **Brown, S.,** *Play is more than fun,* http://www.ted.com/talks/stuart_brown_says_play_is_more_than_fun_it_s_vital.html (05.2008)
13. **Geirland, J.,** *Go With The Flow (interview with Mihaly Csikszentmihalyi)* - In: Wired Magazine, http://archive.wired.com/wired/archive/4.09/czik.html (1996)
14. **Nielsen, J.,** *Why You Only Need to Test with 5 Users* http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ (17.05.2014)
15. **Järvi A., Mäkilä T., Hyrynsalmi S.,** *Game Development Accelerator – Initial Design and Research Approach* http://ceur-ws.org/Vol-1095/iwlcsp2013.pdf#page=47 (17.05.2014)

16. **Floyd, D.,** *Video Games and Learning* https://www.youtube.com/watch?v=rN0qRKjfX3s (17.05.2014)
17. **GSV EDU** *Education Sector Factbook 2012* http://gsvadvisors.com/wordpress/wp-content/uploads/2012/04/GSV-EDU-Factbook-Apr-13-2012.pdf (17.05.2014)
18. **Schutte, W.,** *Newzoo, Global Games Market Report Infographics* http://www.newzoo.com/infographics/global-games-market-report-infographics/ (17.05.2014)
19. **Mojang.com** *Employees* https://help.mojang.com/customer/portal/articles/331367-employees (17.05.2014)
20. **Bristow, E.,** *The Edublogger, Gaming in Education – Minecraft in Schools?* http://theedublogger.com/2013/11/21/gaming-in-education-minecraft-in-schools/ (17.05.2014)
21. **Karmali, L.,** *IGN, GTA 5 Currently Holds Seven Guinness World Records* http://www.ign.com/articles/2013/10/09/gta-5-currently-holds-seven-guinness-world-records (17.05.2014)
22. **French, M.,** *Develop-Online, Inside Rockstar North - Part 2: The Studio* http://www.develop-online.net/studio-profile/inside-rockstar-north-part-2-the-studio/0184061 (17.05.2014)
23. **Pew Research Centre** *E-Reading Rises as Device Ownership Jumps* http://www.pewinternet.org/files/old-media//Files/Reports/2014/PIP_E-reading_011614.pdf (07.07.2014)
24. **Nielsen, J.,** *American Families See Tablets as Playmate, Teacher and Babysitter* http://www.nielsen.com/us/en/insights/news/2012/american-families-see-tablets-as-playmate-teacher-and-babysitter.html (07.07.2014)
25. **Code.org** *Promote Computer Science* http://code.org/promote (01.08.2014)
26. **Lean.org** *What is lean?* http://www.lean.org/whatslean/ (17.08.2014)

# APPENDIXES

## Appendix 1 Pretest and Posttest Questionnaire

First five questions were asked before and after the play testing session, the last question was asked as the last thing when the test session was over. As tests were done with Estonian children, then questions in Estonian are added here also.

1. How difficult do you think the game is?
2. How clear is the goal of the game?
3. How much does the task at hand need concentration?
4. Do you understand the interaction and the controls of the game?
5. How much time you have spent on this playtest?

1. Kui raske see mäng tundub sulle?
2. Kas sa saad aru, mida siin mängus tegema peab, mis on eesmärk?
3. Kas selle mängimine vajab palju keskendumist või pigem vähe?
4. Kas sa saad aru, kuidas seda mängu mängida?
5. Mis sa arvad, kui palju sul kulus aega selle mängu mängimisele?

## Appendix 2 Observation Notes – Kids' Emotions While Playtesting the Paper Prototype

| | | | | |
|---|---|---|---|---|
| 8 years old | 8 years old | 9 years old | 6 years old | 9 years old |
| Boy | Girl | Boy | Girl | Boy |

| | | | | |
|---|---|---|---|---|
| Was very quiet | In a good mood | In a good mood | Very happy and energetic | Serious and thoughtful at first |
| Didn't show much emotion | Face was tense during playing at first | Looks enthusiastic and happy | Seems somewhat intimidated by the testers | Face was tense during playing |
| Face was tense during playing | When wins a level cheers and laughs out loud | When characters die he makes a big O with his mouth and covers it with both hands while saying "Noo" | Goes through the levels very quickly | Happy when wins a level |
| Game became quickly quite hard for the subject | Later on smiles also during play | Makes a lot of "Hmmm" thinking sounds while figuring out the harder levels | Is only one of the testers who wants to move the blue and red balls by herself on the board | Very quiet and focused during playing |
| Focused on the game | Seems enthusiastic towards playing | Harder levels make his face more tense and focused | When levels get really hard she gets more focused and face becomes more tense, but still has a slight smile on her face | |
| | When gets to the third level which is quite hard, gets very focused and scowls | Doesn't want a tip while playing at first but later wants one | | |
| | When asked if she wants to quit during a hard level that she has been working on for some time she says that "I want to figure it out" | | | |

## Appendix 3 Observation Notes – Kids' Emotions While Playtesting the Low Fidelity Prototype

| 8 years old | 8 years old | 11 years old | 11 years old | 10 years old |
|---|---|---|---|---|
| Boy | Girl | Boy | Boy | Girl |
| Serious and thoughtful | Serious and thoughtful | Serious and thoughtful | Thoughtful | Happy and focused |
| Was very quiet | Very focused | Gets loud and angry when "characters" / circles don't go where needed | Has a smile on his face | Thoughtful and tense face during play |
| Very focused | Happy when wins a level | Didn't want a tip and want to solve the harder levels himself | Very quiet and focused during playing | Very quiet and focused during playing |
| Happy when wins a level | Levels seem to be very hard for her | Face was tense during playing | Harder levels take long time to finish, but when gets "characters" / circles to the finish says happily "Yesss" | Comments that she likes the small characters when they hop around |
| Smiles when "characters" / circles start to move in the game | Seems to be somewhat frustrated | Sighs loudly when playing harder levels but doesn't want to stop playing | When manages to finish a level gets very excited and happy | Harder levels take a lot of time but she doesn't want to quit playing although the opportunity is presented |
| The more he plays the more his face becomes relaxed and happy | Face was tense during playing | Gets frustrated during a harder level and screams "Noo" but doesn't want to quit | Wanted to continue playing when asked if he wants to stop playing | |
| There is a small smile on his face half way through the game | | | We allowed him to continue playing during final questions. It was very hard to get answers from him because he was very focused | |

| | | | | |
|---|---|---|---|---|
| | | | on the game and didn't hear or answer to half of them straight away. We had to asked several questions several times. | |

## Appendix 4 Observation Notes – Kids' Emotions While Playtesting the High Fidelity Prototype

| | | | | |
|---|---|---|---|---|
| 11 years old | 10 years old | 11 years old | 11 years old | 10 years old |
| Girl | Girl | Girl | Girl | Boy |
| Looked very happy at the beginning but later on became more serious and focused | Looked very happy at the beginning but later on became more serious and focused when levels got harder | Happy and focused | Happy and focused | Serious and tense face |
| When finishes the level gets very happy | When finishes the level gets very happy and expressis it with loud "Jee" | Thoughtful and tense face during play | Enthusiastic | Seems serious but enthusiastic |
| Face was tense and focused during playing | When asked questions during harder levels doesn't respond and stays very focused on the game | Cheers and claps hands when finishes the harder levels | When finishes the level gets very happy and expressis it with loud "Jee" | When reminded that he can leave when he doesn't want to play anymore makes a comment that he likes the game and adds "You have to think a bit (with this game)" |
| Sighs loudly but happily when playing harder levels but doesn't want to stop | | When asked if she wants to quit continues to play | Tense face during play | Goes very quickly through the levels |

| | | | | |
|---|---|---|---|---|
| playing | | | | |
| When asked questions during play is very focused and doesn't seem to notice the questions | | | Harder levels seem ot be too hard for her. Gets somewhat frustrated and doesn't seem to be so enthusiastic anymore. | His movements are very fast and he quickly tries out several options |
| Got very excited and laughed out loud when trapped some characters in a loop so they couldn't escape | | | Crosses her hands several times during harder levels | When levels got harder made a loud thinking sound "hmm" |
| | | | Gets angry by making an angry face when characters don't get to the finish | |
| | | | Didn't want to continue playing because game got too hard for her in the end | |

# SUMMARY IN ESTONIAN

**Videomängude kaasahaaravuse testimine prototüüpidel: videomängude prototüüpide võrdlemine erinevates arendusfaasides testides neid kuue kuni üheteist aastaste lastega.**

Magistritöö eesmärgiks on uurida, milline peaks olema videomängu prototüübi viimistuslik tase, et sellega lastel testides oleks võimalik veenduda, kas mäng on lõbus ja kaasahaarav ja kas selle arendusse tasub raha ja aega panustada. Uuringu vajadus on pärist elust enesest - autorite loodud mängustuudio töö optimeerimiseks on vaja võimalikult vähese panusega teada saada, kas parasjagu planeeritud mäng on kaasahaarav ja kas sellel on potentsiaali sihtrühma kasutajate seas. Samal põhjusel tegelevad testimisega ka kõik teised mängustuudiod - tavaliselt luuakse erinevaid prototüüpe ja mudeleid, nii digitaalseid kui ka paberprototüüpe, et nii ise kui ka kasutajate peal läbi mängida peamised mängumehhaanikad ja tuvastada, kas planeeritav on piisavalt lõbus ja kaasahaarav või mitte.

Erineva tasemega prototüüpide loomiseks kulutab ettevõte kordades erineva arvu tunde ja eurosid. Näiteks paberprototüübi loomiseks võib kuluda paar tundi aga sama asja primitiivne digitaalne versiooni (low fidelity) arendus võib võtta päevi või isegi nädalaid, mis on ettevõtte jaoks riskiga tööjõukulu - kasutajatega tehtud prototüübitestide tulemusena võib selguda, et see idee ei ole kõlbulik ja kõik tuleb ringi teha või lausa minema visata.

Senise kolmekümne aasta jooksul on mänguarenduseks vajaliku testimise (*playtesting*) teemadel väga vähe akadeemilisi uurimistöid tehtud, veelvähem lastele suunatud hariduslike mängude osas.

Selles töös kasutatud kasutati autorite mängustuudios loodud haridusliku mängu Code2Kids prototüüpe. Väikese meeskonnaga iPadile loodav hariduslik mäng on plaanis turule tuua peale selle töö valmimist ja töös leitu sisseviimist high fidelity prototüüpi, millest seejärel, peale arendustööde lõppu, 2015 aasta alguseks saabki mängu esimene avalik versioon.

Uuringuks viidi läbi mängutestimisi Tallinna Gustav Adolfi gümnaasiumi 6 kuni 11 aastaste

lastega. Nii testimise korraldamisel kui ka eeltöö juures olid abiks ka sama kooli haridustehnoloogid ja algklasside õpetajad, kes sarnaseid mänge juba praegu oma õppekavades kasutavad.

Testimisel sai iga laps proovida läbi mängida mitmeid tasandeid talle etteantud prototüübil. Et lapsed saaksid olla võimalikult objektiivsed, siis ei näidatud neile teisi prototüüpe. Testimise viidi läbi klassikalisel kasutatavuse testimise meetodeid kasutades. Lisaks mängutestimisele tehti lasetga ka eel- ja järelintervjuud, kasutades Mihaly Csikszentmihalyi *flow* teooriast tuletatud küsimusi. Autorite jaoks olulisel kohal on ka mängutestimise ajal läbiviidud obersvatsioon, mille märkmeid kasutatakse kindlasti mänguarenduse planeerimisel ja läbiviimisel. Testitaval mängul peab testimiseks olema vähemalt töötav mängumehaanika ning ning võivad olla lisatud väga algelised semantiliselt korrektsed visuaalid. Mängumehaanikal on kogu mängimise kogemusele suurem mõju kui visuaalidel, animatsioonidel või helil.

**Non-exclusive licence to reproduce thesis and make thesis public**

We, **Kaspar Roost** and **Rene Rebane**

_____,

(*authors*)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:
1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and
1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

**Video game engagement testing with prototypes:**

**Comparing video game prototypes in different development**

**stages by testing these on six to eleven years old kids**

_____,

(title of thesis)

supervised by _____**Ander Tenno**_____,

(supervisor's name)

2. I am aware of the fact that the author retains these rights.
3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Viljandi, **19.08.2014**