

Performance, motivation and immersion within a suite of working memory games

Hanne Fagerjord Karlsen



Master's Thesis
Master of Applied Computer Science
30 ECTS
Department of Computer Science and MTDMT Technology
Gjøvik University College,

Avdeling for
informatikk og medieteknikk
Høgskolen i Gjøvik
Postboks 191
2802 Gjøvik

Department of Computer Science
and MTDMT Technology
Gjøvik University College
Box 191
N-2802 Gjøvik
Norway

Performance, motivation and immersion within a suite of working memory games

Hanne Fagerjord Karlsen

26/11/2014

Abstract

Almost 20% of Norwegian children and youth struggle with behavioural and cognitive disability. Working memory deficiency is especially common among children with ADHD. Recent advances in developmental psychology suggest that people with ADHD might benefit from games designed to train working memory abilities. The motivating factor from computer games can be especially strong to those with ADHD, as they respond strongly to motivational reinforcement.

This thesis investigates performance, motivation and immersion within a suite of working memory games. A user study was conducted with a group of 27 Norwegian children. The participants played a game with a suite of working memory training minigames and answered a questionnaire about the game experience. Observations, game event data and questionnaire results were analysed in order to explore performance, motivation and immersion in game experience. The analysis showed that participants improved as they played the games and that the new skills appeared to transfer to the other minigames as well. There is a significant correlation between immersion and wanting to play the game again. Participants who had higher immersion rates were more likely to want to play the game again. The positive findings in improvement and immersion support the idea of using games as a source of motivation in a serious context.

Acknowledgements

The completion of this thesis concludes a chapter in my life and allows a new one to begin. It has been quite the journey. First and foremost I want to express my gratitude and appreciation towards Simon McCallum. I could never have asked for a better supervisor and motivator. Thank you so much for all your time and good advice. I want to give big thanks to Charlotte Lunde, who allowed me to use her game for this research project. I would also like to thank the science center, Vitensenteret Innlandet, for being optimistic and helpful during the experimental sessions. Additional thanks goes to all the children that volunteered for this project. I want to thank Mariusz Nowostawski for helpful advice and criticism during the revision process. Thanks to existing and new friends whom I've have shared lunches and conversations with during times of much work and stress. Last but not least, I want to express my gratitude and love for my partner, Even, who has been my biggest support and source of laughter. Oss to i Oslo<3

Hanne - November 26th

Contents

Abstract	i
Acknowledgements	ii
Contents	iii
List of Figures	v
List of Tables	vi
1 Introduction	1
1.1 Motivation	1
1.2 Contributions	2
1.3 Delimitations	2
1.4 Research Questions	2
1.5 Thesis Structure	3
2 Serious Games	4
2.1 Serious Games	4
2.2 Game Experience	4
2.3 Playability	6
2.4 Motivation and Games	7
2.5 Games and Learning	7
2.6 Games and User Acceptance	8
3 Working Memory	10
3.1 Short Term Memory and Working Memory	10
3.2 WM Training	10
3.3 Computerized WM Training	12
3.4 ADHD, Working Memory and Games	12
3.5 Effectiveness of Working Memory Training	13
4 Game Analytics	15
4.1 Analytics	15
4.2 Telemetry	15
4.3 Metrics	16
4.4 Mixed Methodologies in Game Analytics	16
5 Methodology	18
5.1 Scientific Point of View	18
5.2 Variables	18
5.3 Game Description	19
5.4 Game Workshop and Game Design Sessions	23
5.5 First User Test - Summary	23
5.6 Second User Test - Questionnaire Pre-Study	25

5.7	Questionnaire Coding	27
5.8	Avoiding Bias	28
5.9	Third User Test - Questionnaire Pre-Study	30
5.10	Fourth User Test	31
6	Results	35
6.1	Inclusion Criteria and Data Cleanup	35
6.2	Results from Questionnaire	35
6.3	Results from logged Game Behaviour	38
6.4	Observational data summary	39
6.5	Results and statistical tests related to each hypothesis	41
6.5.1	Research Question 1: Can we measure and observe improvement in game-play?	41
6.5.2	Research Question 2: What makes children want to play a Working Memory Game?	44
7	Discussion	50
7.1	Hypothesis Discussion	50
7.1.1	Research Question 1: Can we measure and observe improvement in game-play?	50
7.1.2	Research Question 2: What makes children want to play a Working Memory Game?	51
7.2	Participant Bias	53
7.2.1	Questionnaire	53
7.3	Game Design	53
7.4	Validity of Questionnaire	54
7.5	Weaknesses of this study	54
8	Conclusion	56
9	Future Work	57
	Bibliography	58
A	Appendix 1 - Informed Consent	63
A	Appendix 2 - Questionnaire	65
A	Appendix 3 - Game Event Logging	72
A	Appendix 4 - Pearson Correlations - Time and Level	74

List of Figures

1	The Flow Model	6
2	Memory Span Tests	11
3	N-back Test	11
4	Minigame 1	20
5	Minigame 2	20
6	Minigame 3	21
7	Minigame 4	22
8	Minigame 5	22
9	First User Test - Setup	24
10	Smileyometer	26
11	Research Setup	31
12	Research Environment	32
13	Research Environment Entrance	33
14	Self reported frequency of playing.	36
15	Immersion Metrics	37
16	Flow Metrics	38
17	Mean playtime of each game	39
18	Distribution of fail	40
19	Progression - Minigame 1 and 2	42
20	Progression - Minigame 3 and 4	43
21	Progression - Minigame 5	44
22	Graphs: Mean mistakes and order of play	45
23	Bar chart: Mean mistakes and order of play, sorted by minigame	46
24	Clustered bar chart: Difficulty progression and time spent playing	48
25	Self reported difficulty and time played	51
26	Example of game log code	72
27	Example of saved XML structure of lick events.	73

List of Tables

1	Prestudy: Time each group spent playing	25
2	Frequency Table: Do you want to try it again?	37
3	Statistical tests: Intent to use, flow and immersion	47
4	Statistical test: Time spent playing and reported level of fun.	48
5	Statistical test:Time spent playing and reported level of difficulty	49

1 Introduction

1.1 Motivation

Many children with behavioural disabilities go through daily struggles related to their deficiency in various cognitive tasks. The Norwegian Institute of Public Health revealed in a 2014 report how 15-20% of Norwegian children and youth are struggling with behavioural disorders, ADHD being prominent [1]. Children with ADHD exhibit significant impairment in cognitive abilities such as working memory [2]. The motivational approach to ADHD suggests that these patients respond strongly to motivational reinforcement and are reported to concentrate more when playing computer games [2]. Therefore, there is potential for massive benefit if a computer game is proven effective for minimising the effects of these disabilities. Recent research suggests that it could be possible to train working memory abilities through games [3]. Digital games are easily distributed to players and can be brought everywhere through handheld devices. The Norwegian Media Authority reports that 83% of children aged 8-16 have access to smartphones, making touch screen devices a well known technology that's likely to grow [4].

The "2014 Sales, Demographics, and Usage Data" report from the ESA (Entertainment Association) shows how the Games Industry in 2013 grossed a total of 21.53 billion US Dollars [5]. The Global Games Market Report 2014 predicts that the Global Games industry will hit a total revenue of 102.9 billion US Dollars by 2017 [6]. Additionally, the revenue of mobile games is predicted to double by 2017. In a 2014 report, The Norwegian Media Authority states that 94% of 9-16 year old Norwegians play digital games weekly [4]. This is a clear increase from the 2012 report where 85% of the same age group play games weekly [7]. The report states that Minecraft [8] is the most popular game among the 9-11 age group and among the top 5 games for the 12-14 age group (both genders included).

With a game industry growing rapidly, it is hard to get noticed as children are met with a plethora of games through their devices. As children are prone to stick to the most engaging options, it is increasingly more difficult to keep their attention. Existing research demonstrates how games use psychological aspects of the human mind to keep players engaged and motivated to play the game an extended amount of time [9] [2]. Most studies on games for cognitive training focus on performance and effectiveness of the game as a treatment [10] [3] [11]. While effectiveness is extremely important, it is also important to note that without a fun experience, players might not be interested in the game long term. There are still few studies that focus on user experience and motivation in the current games made specifically for cognitive training [12] [2]. It has been suggested that using familiar gamelike tasks might promote motivation and performance in participants with ADHD [13], creating an interest in how concept familiarity might affect experience.

1.2 Contributions

The goal of this thesis is to measure performance, motivation and immersion in children playing a cognitive training game for working memory. This thesis seeks to explore parts of the ambiguous questions “*what makes a game fun?*” and “*what keeps the player engaged?*” in the context of children’s games for health. Through quantitative experiments and qualitative observation of a working memory game, this project aims to explore how performance, motivation and immersion is linked to game experience and ultimately the continued use of an application. Personal contributions include adding event logging to an existing game, attending workshops and game design sessions. Minigame 5(p. 22) was conceptualized in one of these sessions and I was a part of the creative design process. Other contributions related to the game include involvement in early playtests, observing and reviewing play.

1.3 Delimitations

The main purpose of this thesis is to find out what makes children interested in playing a working memory game and if it is possible to measure improvement in gameplay. Parts of the suite of minigames presented in this thesis were previously developed for Charlotte Lunde(MD) as part of an on going doctoral dissertation. Game event logging was added to the already existing game code. Sample code of how events were logged can be found in the Appendix, Section A. The game uses graphics well known from Minecraft. It was of interest to test if concept familiarity would create a better game experience. Due to its popularity among Norwegian children, Minecraft was deemed a good choice to test this. The game is yet to be validated and still can’t in any way be seen as treatment to medical conditions. ADHD and other behavioural deficiencies present complex situations where trained professionals must assist the treatment. This thesis discusses how games have been used to train working memory and if it is feasible to use this method as a treatment. The thesis does, however, not attempt to prove or disprove if working memory training works. The questionnaire developed for this study is an adaptation of the Game Experience Questionnaire(GEQ) [14, P.265] and has not been independently validated. This study did not ask for the medical status of the participants, assuming everyone to be typically developing. There was not enough time to test the atypical participants, that became available only at a very late stage in the project. Therefore, the results from this study only stem from participants without known disabilities.

1.4 Research Questions

This thesis asks two main questions with following hypotheses:

Research Question 1: Can we measure and observe improvement in gameplay?

Hypothesis 1. *Time spent playing can predict progression.*

Hypothesis 2. *There is transfer in learning across the minigames.*

Research Question 2: What makes children want to play a Working Memory Game?

Hypothesis 3. *Reported immersion and flow can predict player intent to use.*

Hypothesis 4. *Participants who play the game longer find it more fun.*

Hypothesis 5. *Reported level of difficulty can predict time spent playing.*

Hypothesis 6. *Concept familiarity will affect experience positively.*

Through out the thesis, the null hypothesis will be marked by " H_0 " and the alternative hypothesis will be marked " H_1 ".

1.5 Thesis Structure

Chapter 2 presents the game-related background material, focusing on factors of game experience and playability. This Chapter also discusses how games can be used as a successful learning aid. Chapter 3 explains the underlying theory on working memory and short-term memory. The Chapter presents the standard memory tests which the game is based upon and includes a presentation and discussion on current working memory game systems. Chapter 4 shows how Game Analytics is an increasingly popular branch of intelligence gathering in the game industry and game user research. Examples of how mixed methods in game analytics can provide a better understanding of games are presented [15] [16]. The methodology in Chapter 5 explains how the project was organised and how several user tests were organised in order to increase validity of components of the methodology. Chapter 6 presents descriptive statistics from the questionnaire and game logging, qualitative data from user observations and statistical calculations related to each research question. The discussion in Chapter 7 discusses the results related to each hypothesis. In addition, the participants, questionnaire and game design are also discussed. The validity and weaknesses of the study is also discussed. Chapter 8 concludes the findings in relation to the research questions and the overall goal of the thesis. Chapter 9 suggests how future work can extend and continue the work started by this thesis.

2 Serious Games

Humans have always always played games to learn. As the Dutch historian Johan Huizinga famously said in his book *Homo Ludens* [17]:

Play is older than culture, for culture, however inadequately defined, always presupposes human society, and animals have not waited for man to teach them their playing.

This Chapter is focused around the term *Serious Games*, experience, motivation and evaluation. Section 2.1 discusses the term *Serious Games* and other popular terms used by industry and researchers when describing games used for more than entertainment. Section 2.2 and 2.3 presents terms and concepts in game experience and playability. Motivation is briefly discussed and a game example of motivation is presented in Section 2.4. Section 2.5 shows how games can be used as a tool for learning and Section 2.6 presents how Technology Acceptance Models can be used in game evaluation.

2.1 Serious Games

Using games for useful purposes is not a new phenomenon. It has been proven that games have been used as deliberate tools since ancient times, either for training or entertainment. The serious use of games throughout history is mostly prominent in military tactics. From the ancient sand tables used by the Roman empire [18] to chess and modern day war simulation games such as VBS3 [19].

Serious Games can be defined as "full-fledged games for non-entertainment purposes" [20]. In 2002, The *Serious Games* Initiative was launched to focus on uses for games in exploring management and leadership challenges facing the public sector [21]. Since the launch, the term *Serious Games* has become an umbrella for several emerging branches of research and the games industry. A part of this initiative is The Games for Health Project, which seeks to explore the intersection of videogames and health.

Alongside *Serious Games*, a plethora of new terms are being used to define the use of games and game thinking. The term Gamification has gained some momentum the recent years and have been used to describe "the use of game design elements in non-game contexts" [20]. Due to polarising debates regarding the practice of gamification, some researchers are suggesting the term *Gamefulness* and *Gameful Design*, as introduced by Jane McGonigal [22], as an alternative for academic use [20].

For the sake of consistency, this thesis will use the term *Serious Games* with the before mentioned definition.

2.2 Game Experience

When we seek to understand player experience, we also need to understand three different concepts: *immersion*, *presence* and *flow* [23].

Immersion

Immersion is described as "a powerful experience of gaming". The term immersion is widely used when describing the game experience, but it is not used consistently with the same definition. A qualitative study by Brown & Cairns [24] suggests *immersion* as a term to describe the level of involvement with a game over time. Three stages of *immersion* were suggested: *engagement*, then greater involvement in *engrossment* and in the end, *total immersion*. Brown & Cairns also describe how each stage has certain barriers that must be removed before the player can reach it. The first stage, *engagement*, describes the initial involvement with the game, the player's interest and investment in the game. The next stage, *engrossment*, is described as emotional investment in the game. The third stage, *total immersion*, is described as *presence*. This refers to the game being the only thing affecting the player's thoughts and feelings. The term *presence* is here used interchangeably with immersion.

Presence

There is still an ongoing debate on how to define the term *presence*. *Presence* is often referred to as a state of mind rather than an experience in time [23]. Jennet et al. [25] presents a distinction between *presence* games (roleplaying games, first person shooters, etc) and *non-presence* games (abstract puzzles). Tetris is an example of a game used to exemplify non-presence and high immersion. Even if the player doesn't feel like they are in a world made of blocks, they can still feel drawn into the game by not noticing their surroundings, experiencing time loss [25]. Additionally, a game can include a high sense of *presence* without *immersion*. One could imagine a very realistic simulation game where the player feels present, but does not experience interest, emotional investment, lost sense of time or other factors of *immersion*.

Flow

The psychological concept of flow was first proposed by Mihaly Csikszentmihalyi [26] to describe the experience of being fully focused and engaged in an activity. Flow was described by Csikszentmihalyi as: "being completely involved in an activity for its own sake. The ego falls away. Time flies. Every action, movement, and thought follows inevitably from the previous one, like playing jazz. Your whole being is involved, and you're using your skills to the utmost.". Csikszentmihalyi also argued that flow occurred when the challenge in an activity was balanced with the skills of the individual.

Csikszentmihalyi presented several attributes that flow is said to consist of:

- A balance between challenge and skill
- Clear goals
- Clear feedback
- Lost sense of time
- Loss of self-consciousness
- Feeling of enjoyment and control
- Feeling of intrinsic reward

The flow model shown in Figure 1 is based on this sentiment. This model was first presented by Csikszentmihalyi [26]. The model contains three regions of experience where the flow chan-

nel is the region where skill and challenge matches. This model has become popular in games research and the games industry, despite it containing several shortcomings. This model of flow argues that skills must be stretched in order to maintain the flow state and that activities that do not require much skill cannot provide flow. Csikszentmihalyi and Nakamura [27] reports how experiential data from participants did not match up to the original mappings of the model. They went on to redefine flow as "the balance of challenges and skills when both are above average levels for the individual".

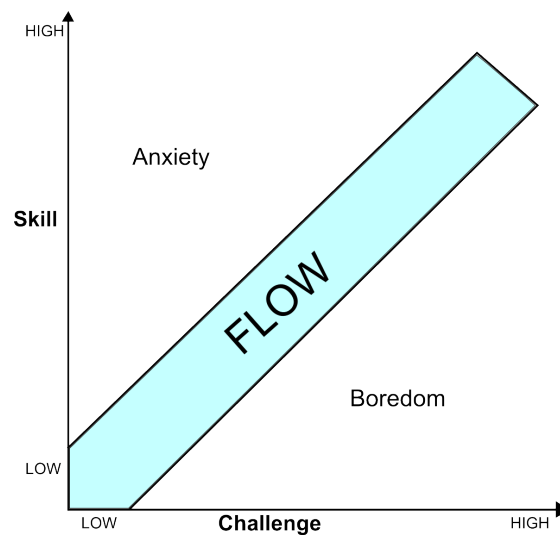


Figure 1: The flow model mainly describes the relationship between challenge and skill. If skills are significantly higher than the challenge, the model proposes boredom as a result. If the challenge is significantly higher than the skill, anxiety is predicted. The flow channel is in the middle, showing the area where challenge and skill is balanced or very close.

2.3 Playability

Playability is defined as the "evaluative process directed toward games, whereas player experience is directed toward players". Playability seeks to measure the game in order to find areas of improvement. Player experience research focused on measuring the player and how to improve the gaming. It is important to separate between the two in a research setting as the focus greatly impacts which methodologies to use.

Nacke [28] argues that good playability should be a prerequisite for all game evaluation. Game metrics have become a popular means of gathering intelligence for playability research. Game metrics refers to data collected directly from the game, usually various game events, such as time spent doing a specific task. It is an effective way to get a large mass of data on specific parts of the game. Metrics can tell researchers exactly what is happening during gameplay without any player feedback. As the data only tells *what* happened, it is important to pair it with other measures in order to understand *why* it happened. These measures allow researchers to correlate the subjective and objective experience to actual game events. A game that appears

too hard and unbalanced when looking at the metrics might in reality be perceived as fun and addictive by the player. It is suggested that researchers should apply many measures in order to get a fuller understanding of the game experience. More details on game metrics and analytics can be found in Chapter 4.

2.4 Motivation and Games

When we want to understand why players are motivated to continue playing games, we need to understand two types of motivation, intrinsic and extrinsic motivation. Extrinsic motivation is described as "the performance of an activity because it leads to instrumental rewards" and refers to external sources of motivation [29]. Intrinsic motivation refers to "the performance of an activity for its inherent enjoyment other than a separable outcome" [29]. Extrinsic motivation can be tricky to keep sustainable if not managed. Incentives are commonly used as extrinsic rewards. This can be compared to a child being paid to take out the trash. If the parent decides to stop the money as an incentive, the child might decide that the task is no longer worth doing. When designing game experiences depending on extrinsic reward, it is important to pay attention to the possible fragility of this motivation.

That said, it is possible to keep a person interested by use of extrinsic reward if the loop is crafted well enough. A great example of successful extrinsic motivation can be found in modern massive multiplayer online (MMO) games. In most MMO games, the player is often faced with quests (Tasks) where certain objects need to be obtained or actions are required in order to obtain a reward. The reward serves as the motivator and extrinsic reward for completing the quest. Once the quest is completed, the reward loop doesn't stop. The reward is more than often a tool to complete even bigger tasks, such as getting better equipment that allows the player to attempt even harder challenges. Improving the equipment isn't the end of the game, far from it. The improved equipment makes it easier to obtain bigger goals, defeat bigger enemies and gain even bigger rewards. Every reward leads to new challenges and goals, revealing new motivating factors to keep the player motivated and engaged. The bigger rewards are often recognisable and give the player a certain status among other players, positively setting them apart. Overall, this creates a motivation cycle of everlasting extrinsic reward that sustains itself.

The challenge of *Serious Games* and gameful systems is recreating this kind of sustainable motivation. A common flaw is to stop the engaging elements too quickly. This is often seen in attempts to gamify systems, where mechanics such as points and badges are added to an already existing system. The problem with this design emerges once the player has acquired a momentum of points. At this point, the player might expect being able to use the points for new goals and rewards. Quite often, the points and badges have no function other than being an initial reward. Once the elements of motivation no longer provide a new reward, the player becomes disinterested in the game. This defeats the purpose of using a game for real world problems.

2.5 Games and Learning

Serious Games are being used by schools and companies alike to assist educational goals. The Norwegian military has used games as a platform for teaching cultural awareness [30] and

Norwegian schools are increasing their use of games as motivating learning aids [31]. Games have also been used as a method for spreading global awareness through the Global Conflicts series [32]. This game series makes it possible for players to explore and learn about different conflicts around the world, such as the Israeli/Palestinian conflict or child labour.

A meta study on educational games conducted by SRI International summarised the findings of 69 studies with a total of 6868 unique participants [9]. All studies were to include at least one comparison of a game versus a non-game condition, pre and posttest results of learning had to be available and all participants needed to be aged between 6-25. This meta-analysis argues that current research should not ask *if* but *how* games can compliment learning. The study warns against generalising all *Serious Games* approaches and traditional non-game approaches in learning. *Serious Games* for learning and traditional learning methods are complex media that one should be careful to simplify. Instead of comparing a serious game to a non-game condition, different types of game design could be compared in order to reveal differences in learning outcome due to design differences. This sentiment is shared by Wouters et al who indicate how *Serious Games* can boost learning outcome and retention when used complimentary to traditional training methods [33]. These findings are interesting as they suggest how *Serious Games* can be thought of as part of a battery of educational tools instead of a competing method that seeks to replace the old.

2.6 Games and User Acceptance

User acceptance has been defined as “demonstrable willingness within a user group to employ information technology for the tasks it is designed to support” [34]. The most known and discussed acceptance model is the Technology Acceptance Model (TAM). TAM was presented by Davis Et.al [35] in 1989 as a tool to predict if new technology will be accepted and used within a group or organisation.

The TAM presents several variables as possible predictors that influence whether or not the user will use the system. Perceived Usefulness and Perceived Ease of Use are presented as the main predictors to attitude and intention to use the technology. Perceived Usefulness describes the users’ perceived probability of a system improving their performance. Perceived Ease of Use describes the degree to which the user will expect the system to be used effortlessly [36].

Perceived Usefulness has been shown as a significant determinant to behavioural intention to use technology, Ease of Use is secondary [29]. Perceived Ease of Use also affects Perceived Usefulness. If a system is found to be too inconvenient, this can affect the Perceived Usefulness negatively as extra time and effort is needed to learn the system.

Modified versions of the TAM have been applied with success in *Serious Games* contexts. A study by Ibrahim et al. [34] shows how a modified TAM was used to evaluate an educational game. This study also shows a significant relationship between ease of use and intent to use and points out the importance of understanding what makes people want to play games. Yusoff, Crowder and Gilbert [37] suggest a modified TAM for *Serious Games*. Their model includes more game specific variables such as incentive for motivation. However, these modifications rarely directly include game engagement variables in the model. Little work has been done to create new acceptance models for *Serious Games* in relation to user engagement and motivation. The

EGAM (Educational Games Acceptance Model) is an attempt to include these perspectives into the acceptance models [38]. The proposed EGAM lists enjoyment as a variable and is one of the few proposed acceptance models for games to do this.

3 Working Memory

3.1 Short Term Memory and Working Memory

Working Memory (WM) has been defined as "a brain system that provides temporary storage and manipulation of the information necessary for...complex cognitive tasks". WM is considered one of the most important components of human cognition [11].

Working Memory can be described as post it notes or buckets of information. Short term information is stored in these spaces and retrieved later. Distracting information is, in successful cases, stored in different spaces. For others, the distracting information is stored in the same space that the original information was stored in, making them forget the original information. WM is often confused with Short Term Memory (STM), but the two are separate concepts. STM is thought of as the ability to retain information readily available for a short amount of time and can be considered an ability used by WM. This ability is usually measured by how much information a person can successfully recall. STM is often tested through Verbal and Visuo-Spatial **Simple Span tests**. In these tests, the participant is shown a number of Visual (numbers, words etc.) or Visuo-Spatial (images or placement of an object etc) items. The user has to recreate the items in the correct order. The test usually ends when the user is no longer capable of recalling the order of items. Figure 2 shows examples of verbal and visuo-spatial simple span tests.

The **Complex Span Test** has been used as a reliable predictor of cognitive ability [39]. This test is similar to the simple span test with the added feature of having to solve a task between each presented item. The user still needs to recall the items in the correct order. Figure 2 shows an example of verbal and visuo-spatial complex spans. A sequence of letters is shown. Between each letter, the user needs to solve a simple math puzzle. Afterwards, the user needs to recreate the letters in the correct order. The Visuo-Spatial Test shows the sequential placement of an image. Between each image, the user is shown another image and needs to decide if the image contains a cat or not. The user is also asked to repeat the sequence of images in this task.

The Stroop Task asks the user to repeat the colour of a written word. The written word is often different colour names. The challenge of this task is related to the mix of Verbal and Visual input, causing users to focus on the written word rather than its actual colour.

The N-back test can appear similar to the simple and complex span test, but requires different recall. The user needs to remember a n -number of presented items (often 3) and must evaluate if the current presented item was also presented n times ago. The challenge of this task is to constantly update the list of n items. Figure 3 shows an example of the n -back test.

3.2 WM Training

Training of Working Memory is in most cases based on the theory of neuroplasticity, where it is claimed that the neural pathways in the brain can change throughout life and by training. This opposes the traditional view that the human brain does not change. A well cited example

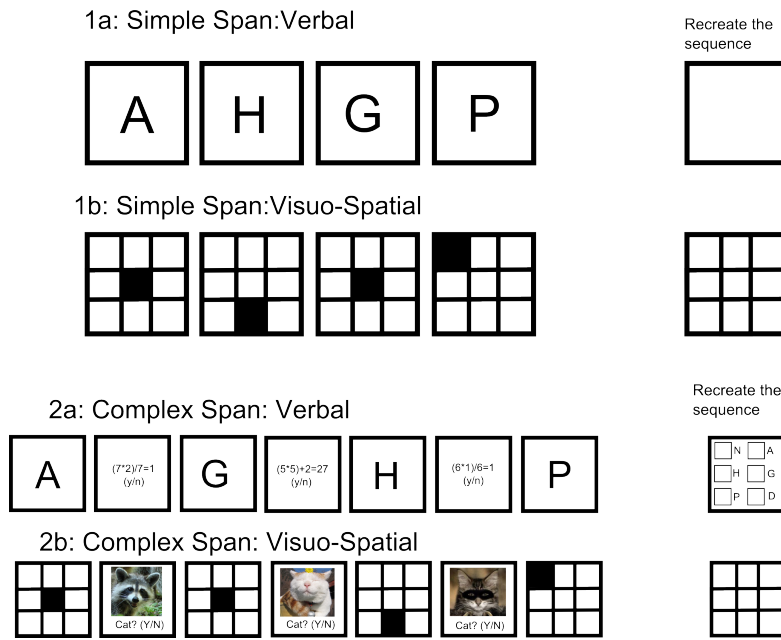


Figure 2: Span tests. Time is represented from left to right. Number 1 represents simple span tests and number 2 complex span tests. 1) The participant must repeat the same sequence after seeing them. 2) The participant must repeat the same sequence. The participant must complete a task between each item.

of neuroplasticity explains of how newly acquired juggling skills showed up as changes in brain structure after three months of training [40]. The introduction of neuroplasticity laid the foundation of modern day training of working memory ability. If WM training has positive effect, it has great potential to alleviate symptoms in patients with cognitive impairment.

Many studies on WM training question the effect of narrow-transfer and far-transfer. The idea of transfer, which is based on neuroplasticity, hypothesizes that training tasks related to working memory can transfer over to real world tasks such as improved attention in school. This is often done by training with traditional cognitive tests such as the complex span test and then testing for improvement in WM related impairments such as attention. Far transfer could be proven by

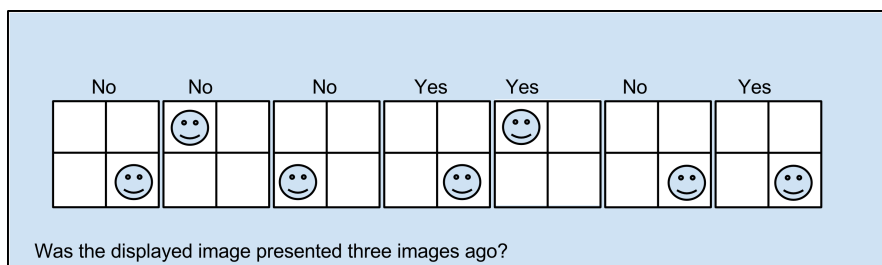


Figure 3: n-back test

effect in untrained tasks while narrow transfer can be seen in improvement of tasks very similar to the trained task.

3.3 Computerized WM Training

Recent years there has been a flux of digital training programs and games claiming to train the brain and improve working memory skills. Examples of these are Cogmed [41], Lumosity [42], Cognifit [43], Jungle Memory [44] and MindSparke [45]. All of these programs are similar in the way that they offer traditional WM training tasks translated into games and game like tasks. To this day, most WM game research is focused around Cogmed and Lumosity. **Lumosity** [42] is a system that provides a battery of games inspired by tasks that measure cognitive ability. Lumosity claims to improve cognitive ability such as memory, attention and problem solving. **Cogmed** is among the most well known training program and claims to "change the way the brain functions to perform at its maximum capacity." [41].

Shipstead et al [3] presents 21 studies where 6 included children with ADHD and 3 studies with typically developing children. The majority of these Cogmed studies report a significant transfer of various cognitive tasks. Shipstead still argues that the overall results are modest at best and that many claims made by Cogmed are based on findings that aren't replicated or haven't been studied thoroughly. This sentiment is also shared by Roche and Johnson in a review of Cogmed Research [46] where they conclude that whilst the program has potential and is among the better WM training programs, there still isn't enough evidence to claim any substantial treatment effects.

3.4 ADHD, Working Memory and Games

Children with Attention Deficit Hyperactivity Disorder (ADHD) exhibit significant impairments in WM and response inhibition [2]. The latest ADHD research suggests two approaches to ADHD, executive functioning and motivational. It is assumed that self-regulation lies at the core of ADHD syndrome and is related to executive abilities such as working memory. Self-regulation plays an important part in everyday life for children, from paying attention in school to waiting for their turn to speak. The motivational approach is based on how ADHD makes a person prone to respond strongly to motivational reinforcement such as rewards and feedback. Prins et al. [2] goes on to describe how, when playing computer games, children with ADHD are reported to concentrate for longer periods of time and behave less impulsively. Computer games are well known to include a wide variety of incentives and intrinsic rewards, which can lead to a stronger and more motivational effect with ADHD patients.

WM Training Games and motivation

Not much research has covered the role of engagement in WMT games. Game based learning has been correlated to higher levels of enjoyment and engagement in learning [9]. There is more information about this in Chapter 2. If one seeks to use games as a platform for WM training, it is worth investigating how game engagement could affect the effectiveness of the program.

A 2011 study tested if computer based WM training with game elements enhanced motivation and training efficacy in children with ADHD [2]. 51 children with ADHD, aged 7-12, were

randomly assigned to regular WM training and WM training in a gaming format over three weeks. The motivation level was assessed by time spent during playing and questions from a post playing questionnaire. The study found that children who had game based WM training were more motivated than the control group. The game group had lower absence time from the tasks, completed more tasks, made fewer errors and did better on an untrained WM task. These improvements were not observed in the control group. While the study can't claim that the game elements contributed to the improvements, it is a promising result for future studies.

3.5 Effectiveness of Working Memory Training

There has been some criticism directed at working memory training, some claiming its impact is too small and insignificant to be useful [10] [11]. It has, however, been argued that we don't know enough about which groups of people might benefit from such training and how it is best executed. There is a dissonance between reported effect by commercial companies and independent research. Cogmed advertises being research based and proven, yet independent studies of the system are modest in their results [47] [3] [46].

A 2012 meta analysis of twenty-three studies [10] concludes that "...these training programs give only near-transfer effects..." and that "... there is no evidence these programs are suitable as methods of treatment for children with developmental cognitive disorders or as ways of effecting general improvements in adults or childrens cognitive skills or scholastic attainments.". This meta-analytic review compares twenty-three different studies which tests WM training. All studies had to be randomized controlled trials or quasi-experiments without randomization, have a treatment , and have either a treated group or an untreated control group. As it is a meta analysis, it is important to remember the possibility of important data being left out of the equation. Additionally, because of a lack of studies, diverse groups of participants of neuro typical and atypical participants were merged. The study can still be seen as indicative of the general findings of the field, where there is little evidence for far-transfer and clear evidence for near-transfer effects.

There is a concern that WM training is too similar to WM tests. As suggested by Sprenger et al [11], there is a change that learning effectiveness does not transfer to tasks significantly different from the training. Sprenger suggests that it is possible for WM training to provide process specific results, claiming that "...narrow training yields narrow transfer.." [11] [48]. If WM tests are too similar to the training, there is a possibility that the user simply improves this one skill and therefore excels the test. The transfer of trained skills is one of the promises of WM training. However, if only the trained skills are tested, it might give a faulty conclusion regarding the reach of transfer. WM training could possibly be compared to weight loss. While one person could benefit from losing 20 kg, it could kill another. Similarly, people are coming to WM training with different baselines. For someone with excellent cognitive abilities, WM training might not make a big difference. For someone with a clear deficiency in certain abilities, there could be more room for improvement. Consistency might still be an important component to continued success. If a successful dieter quits eating healthily, they will quickly gain back their weight. Similarly, there is a possibility that WM training is in need of consistency in order to keep the result [10]. Lervåg and Hulme points to how short term effects are more prevalent than long term effects, which are nearly non existent.

Shipstead et al [39] goes on to describe non blind raters as a serious problem and points out a predictable pattern relating to blind/non-blind raters. Four studies were pointed out with the following tendency: When raters are aware of children receiving WM training, they report improvement in behaviour. Raters that are blind to assignment of WM training do not report any behavioural changes. Results from another meta study show how nonblind raters reported larger benefits than blind raters and objective tests [49]. There are exceptions to the rule, however. Shipstead points out two studies where how nonblind teachers and parents reported positive changes in inattention, but no change in hyperactivity. Regardless, these findings suggest that measuring of WM training is vulnerable to bias. When testing effectiveness of WM training, it is imperative to be aware of systematic error, also referred to as experimenter bias [50, P 57-63]. This can be countered by keeping control groups with blind raters.

4 Game Analytics

Game Metrics and *Analytics* have become widely used buzzwords in the game industry the recent years. These terms refer to the strategic collection of intelligence from games in order to improve gameplay and increase revenue. The success of social online games and the Free-to-Play business model has greatly pushed the need to gain better knowledge about users [15].

This Chapter will explain the various terms used about *Analytics: Telemetry, Game Metrics* and *Game Analytics*. Game Analytics refers to the science of data gathering and analysis in order to support decision-making in all aspects of the game. Telemetry refers to raw data gathered at a distance. This can, for example, be quantitative game logs about how a user interacts with a game, which is saved and moved to a server. Game Metrics refers to quantitative data turned into something that can be interpreted. Examples of this can be "Amount of points" or "Time spent playing". This is data that means something to us, as opposed to raw data which could be a long list of timestamps or other measures. Telemetry data is often the source of metrics, but metrics can be obtained by other measures as well.

Section 4.1 of this Chapter will describe the process of analytics and how it has come to affect the game industry. Section 4.2 will narrow down and define the term telemetry. A practical example of telemetry gathering will be provided. Section 4.3 discusses the term metrics and how it differs from telemetry. Section 4.4 describes the use of mixed methodologies in Game Analytics and how this can be beneficial. Industry case studies of mixed methodologies are presented.

4.1 Analytics

It is important to note that analytics is not the same as analysis. Analytics refers to the entire methodology related to finding and presenting important patterns in data, which leads to problem solving in the real world. Analytics in the games industry is in many ways similar to Business Intelligence (BI), where the goal is to turn business data into valuable information. In the ICT sector, BI is widely used to assess market reports, QA reports (Quality assurance). Game Analytics is still a new field with few standard methods and terms.

4.2 Telemetry

The word telemetry stems from the Greek: tele=remote; metron= measure. Telemetry has traditionally been associated with transmission of radio waves or information sent over an IP network. Game Telemetry has been defined as *...data obtained over distance, which pertain to game development or game research* [51]. An example of game telemetry can be a log with user behaviour or time spent in the game. Game telemetry data is usually digital and transferred from the game client to a server for further processing. Although telemetry is typically digital, the term can be used to refer to any type of data recorded outside the game client. Examples of non digital telemetry data could be time recorded with a stopwatch, printed questionnaires or diaries.

This could manifest in the following scenario: Players keep quitting a particular game level

and the developer wants to understand the reason behind this. The game client gathers raw data about how the players navigate through the game and what they are doing. The game client sends the coordinate data and behaviour data to a collection server. This is the telemetry data. The developer or game researcher pulls out the raw data in order to work with it.

4.3 Metrics

Once the telemetry data has been gathered or pulled from the game server, it is possible to transform the data into metrics. Drachen, Seif El-Nasr and Canossa suggest the following tentative definition to Game Metrics: "A game metric is a quantitative measure of one or more attributes of one or more objects that operate in the context of games" [51].

Metrics can be explained by expanding the example from the telemetry section: After the telemetry data on player behaviour and position is pulled from the collection server, the researcher can clean the data if necessary and create meaningful reports from it. The telemetry data on player position reveals in a heat map that players are stopped at a specific part of the game level. The telemetry data on player behaviour reveals the players being repeatedly defeated in combat. This is metric data. The metric tells the researcher that players are stopped at a specific location and are constantly defeated. This does not, however, give insight into how players experience the game or the real reason for players being defeated. The metrics can tell you something about what is happening and is a crucial tool in analytics.

Although metrics are usually derived from telemetry-data, this does not mean it's the only way. Metrics can also be created from non-telemetry sources. Hazan [16] points out the strength in using a mixed-methods approach in order to gain stronger results.

4.4 Mixed Methodologies in Game Analytics

The field of social science has a long tradition of mixed-methods. Using more than one methodology can help in giving more depth to results where the human factor is involved.

A case study from the development of the game "Prince of Persia: The Forgotten Sands" shows how a mixed-method approach can help developers make the right decisions about changes to a game [16]. The game is a 3D platformer set in a third-person perspective where the player can directly control the camera with the controller. There were usability concerns regarding this feature as the camera controls could be cumbersome to some players. Quantitative data about camera control was recorded in the playtest. The data revealed that players used the camera controls frequently. Additionally, the players had to fill out a questionnaire after the playtest. Some of the questions were specifically asking about the players experience of the camera controls. The questionnaire revealed that the players did not dislike the camera controls and some did not even realize how much they had used the feature. Had the researchers only had quantitative data, they could have been led to believe that the camera controls were disruptive to gameplay as they were used so much. On the other hand, if the researchers only had the qualitative data, they might have believed that players did not use the camera controls much as they didn't pay much attention to it. Both these scenarios could have led to changes for the worse in gameplay and shows how quantitative and qualitative data provides higher certainty that the correct measure has been made.

A case study from the development of Assassin's Creed 2 also shows how mixed-methods were applied with success. Players rated game missions very differently on fun and the researchers wanted to find out why. Game Metrics revealed how few players were climbing rooftops and how those who climbed more rated the missions higher. By changing the speed of climbing, more players climbed rooftops and rated the mission more fun. Without the quantitative data it would have been difficult to understand why some players rated the missions less fun.

5 Methodology

5.1 Scientific Point of View

The scientific method is a structured method which helps us collect and analyse data in a reliable and reproducible way. Most, if not all scientific disciplines have their own preferred set of methods.

This thesis focuses on the evaluation of performance, motivation and immersion in a game for WM training. The scientific point of view used is multidisciplinary. The core of the methodology draws upon the field of Human-Computer Interaction (HCI). As a relatively new field of approx. 30 years, HCI is a blend of computer science, sociology, psychology, communication, human factors and many others. Principles from HCI, Usability and Interaction testing are utilized.

The methodology of this thesis follows two different perspectives, objective game performance and the subjective self reported experience. Logging of game events was added to measure individual performance in the game. This gave an objective perspective on how the players behaved in the game. Subjective measures were done by a questionnaire that logged self reported experience of the game. These measures were later processed for possible correlations according to the hypotheses. The measures from the game and the questionnaire offered quantitative data to the study.

The following list shows which measures were used to answer each hypothesis:

- Hypothesis 1: Time spent playing can predict progression.
Game Events: Logged time, logged difficulty progression.
- Hypothesis 2: There is transfer in learning across the minigames.
Game Events: Logged wrong clicks, logged order of minigames played.
- Hypothesis 3: Reported immersion and flow can predict player intent to use.
Questionnaire: Again Again Table, Answers related to immersion and flow.
- Hypothesis 5: Reported level of difficulty can predict time spent playing.
Game Events: Recorded time spent playing
Questionnaire: Answers related to perceived difficulty
- Hypothesis 6: Concept familiarity will affect experience positively.
Questionnaire: Answers related to Minecraft and positive/negative affect.
Observational data.

Observational data was in majority used as supplementary information to help validate or invalidate results.

5.2 Variables

Independent Variables

An independent variable is something the researcher can control and may manipulate in order to answer specific questions.

- Platform (tablet)
- Game
- Age (Children 8-12)
- Available time to play

As the game is made in Unity, it can be available on many platforms such as Windows, OS X, Android and iOS devices. The touch tablet was chosen as platform for this project. The Samsung Galaxy Note touch tablet was chosen because of its compatibility with the gameplay, which is based on tapping and dragging. Touch tablets are also considered as familiar and easy technology for children.

Dependent Variables

Dependent variables are the outcomes that researchers can't control and need to measure. Lazar et.al [50, p. 26] describes how dependent variables in HCI can be categorized into five groups: efficiency, accuracy, subjective satisfaction, ease of learning and retention rate, and physical or cognitive demand. These categories and the research questions laid the foundation for identifying dependent variables.

The dependent variables of this study were game performance and questionnaire response. The game performance was based on difficulty progression and wrong clicks made. The questionnaire was a modified version of the Game Experience Questionnaire and other variables such as gender, familiarity with games and computers were added.

5.3 Game Description

The game used in this thesis was produced for Charlotte Lunde, MD by Øyvind Byhring. The game resembles well known game systems like Lumosity [42] and CogniFit [43], using established WM tests as inspiration for gamedesign. The game uses Minecraft [8] related graphics as a way to catch the attention of children. The Minecraft references are, however, not crucial to gameplay and it is not required to have Minecraft knowledge in order to play. Once the game is started, the player is met by a map similar to a Minecraft world. The player can control a character over to various objects in order to enter the different mini games.

Minigame 1 - Arrows

In this minigame, the participants were gradually presented with an array of numbers displayed on wooden display boards. The sequence disappeared after being displayed a short while. The participant needed to remember the placement of each number in the sequence by tapping the boards in the correct order. Game is based on simple span tests.

Timer

Unlimited.

Positive Feedback

On all correct: Points are awarded, with animation of points in green text, "Correct" sound effect. Gained star

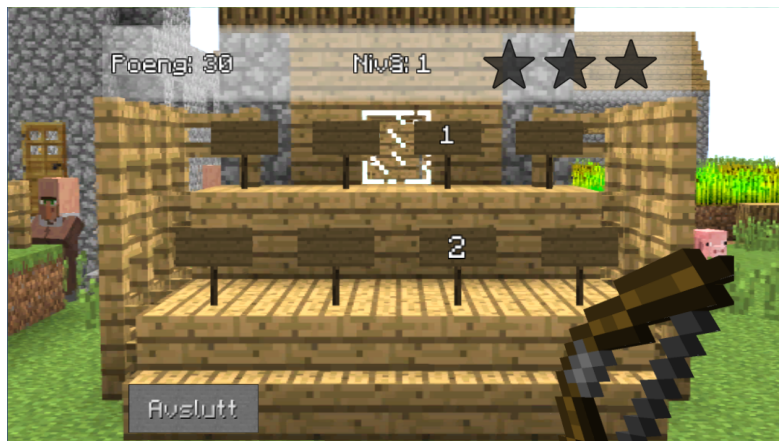


Figure 4: Minigame 1

Negative feedback

On wrong click: Error sound. A star is lost if there are errors when the participant finished the sequence.

Minigame 2 - Gems

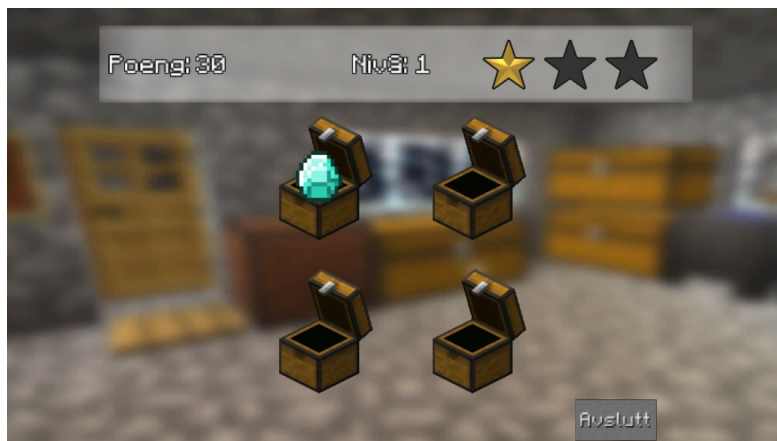


Figure 5: Minigame 2

Participants saw a number of gems move into chests that eventually shut close. The participants tried to remember the placing of all gems by tapping the chests. Participants could click in any order. The amount of gems increased as the participants progressed. Game is based on visuo-spatial simple span tests.

Timer

Unlimited

Positive Feedback

On correct clicks a "Correct sound" is played. When a player finds all gems they are awarded with points and a star.

Negative feedback

On wrong clicks an "Error sound" is played. A red cross appears over the wrong chest. The player lost points and star.

Minigame 3 - Labyrinth

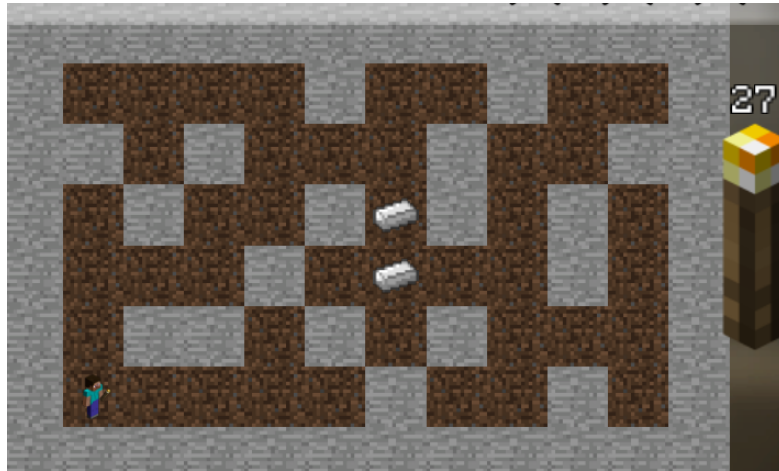


Figure 6: Minigame 3

The participants were shown a labyrinth with pieces of iron scattered about. When ready, the participants could chose to start the game. The light within the game was turned off and the participant could only see what was close to the avatar. The participants needed to remember where the iron bricks were placed and find their way through the labyrinth the the dark to pick them up. This game is based on the complex span test as the participants needed to remember the labyrinth layout while also remembering the placement of items.

Timer

Countdown that increases at each level.

Positive Feedback

Correct sound + points (with green number graphics). Extra sound when all collected.

Negative feedback

None. Level is restarted with new generated labyrinth.

Minigame 4 - Zombie

Zombies peek out from the bushes. The participants needed to remember their placement in order to tap bushes and blow up the zombies. The zombies were displayed for shorter amounts of time as the participants progressed. Inspired by simple span tests.



Figure 7: Minigame 4

Timer

No time limit. Wrong bushes became unclickable after a wrong click.

Positive Feedback

Explosion, zombie falls out. The participants were awarded with points and stars.

Negative feedback

Error sound is played a wrong clicks. Points and star are lost on wrong clicks.

Minigame 5 - Animal Sounds



Figure 8: Minigame 5

The participants were presented with animals and animal sounds played. The participants needed to only click the animal when it moved and the corresponding sound was played. If a pig moved and a cow sound was played, the participants were to hold back and not click. More animals were displayed as the participants progressed. The game was based on the stroop effect. Assumed to train impulse control.

Timer

Countdown

Positive Feedback

On correct clicks, the “Correct” sound was played and a green “check” symbol appeared. The participant was rewarded with a star and points

Negative feedback

The participants lost points and a star on wrong clicks. The “Error” sound was played and a red cross appeared over the animal.

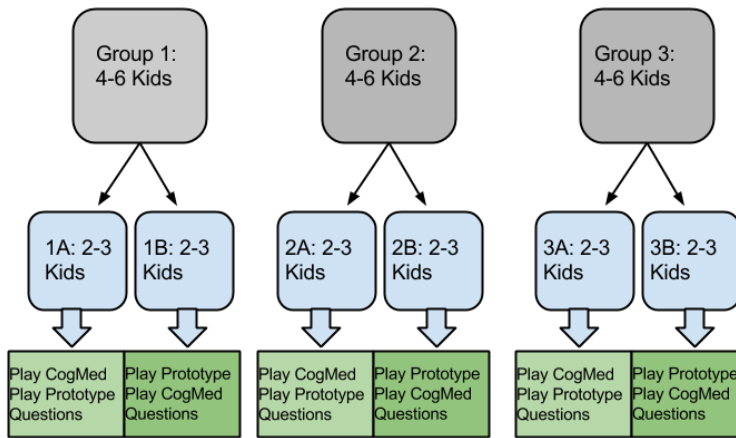
5.4 Game Workshop and Game Design Sessions

During the early stages of the project, a game workshop was arranged. The workshop team consisted of various professionals: designers, researchers, game programmers and a medical doctor. The main objective of the workshop was getting familiar with the project, user testing and discussion of observations and future development. The workshop group was split in two groups and a three children were used as game testers. After briefing, each group observed the children playing. After a brief period of game observation, each group discussed their findings regarding usability and ideas of improvement. The children were asked for their opinion and ideas as well. Both groups then met up as a team and discussed their combined ideas. The main findings of this workshop were related to improvement of feedback in the game and future development.

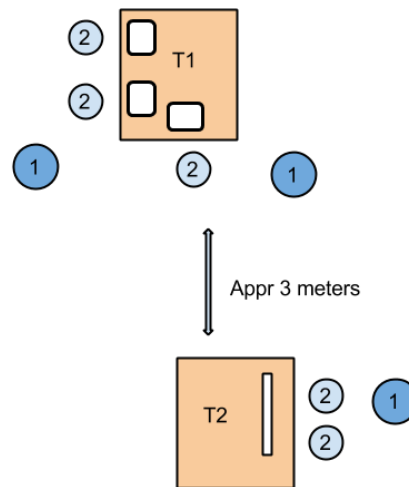
After the workshop, game design sessions and user testings were held. The ideas from the workshop were further discussed in the game design sessions. Minigame 5, which is based of the stroop test was designed during a game design session.

5.5 First User Test - Summary

An earlier version of the game was tested in a kindergarten with 3 groups of 4-6 children. The purpose of this test was to uncover any usability issues that could be missed by adults. The user test was conducted in a kindergarten in an environment familiar to the children. The employees sent in a new group of children when each group was finished. Three groups in total were sent into the room. Each group of children was split in two new groups where one played an early version of the game and the other played a similar cognitive training game, CogMed. Each game was placed on its own table in the same room. The children played the games simultaneously around each table. See Figure 9 for setup of groups and environment. The children were allowed to play the entire duration or stop earlier if they wanted to. Game Sessions lasted 14.5 minutes on average. The groups switched places and played the other game before a new group was sent in. The testing was conducted in a casual observational manner. The children were given help if they didn't understand what to do in the game. Positive feedback was also used to encourage the children to play the game. Afterwards, the children were asked about their opinion about the game and which of the two games they preferred (this data was not recorded).



(a) First User Test - Group setup



(b) First User Test - Setup: T1=Table with Minecraft Game on three tablet devices. T2=Table with CogMed on one PC. 1=Observers. 2=Kindergarten children.

Figure 9: First User Test - Setup of groups and environment.

First User Test - Results

It was observed that having more than one child playing the game caused a distracting effect. Some children started comparing their performance with others and commented if someone performed better than they did. Some children were distracted by seeing others playing different mini games. In one case, the rest of the group followed when one child switched mini game.

In all 6 groups, it was observed that the children started rapidly switching between games after appr. 10-15 minutes. Some started requesting more games or if they could do something different.

All groups of children reported positive feelings towards the game, describing it as “cool” and “fun”. Some of the children enjoyed the games even if they didn’t perform well. Particularly the game where you have to remember the way in the dark was appealing to the children, despite the fact that most of them had trouble with the controls. The older children had few problems understanding the games and playing them. The younger children needed more guidance to understand the game.

When asked if they preferred CogMed or the Minecraft game, the children seemed to favour the last game they played. Similarly, when asked about which mini game from the Minecraft game they liked the best, the children responded with the last mini game they tried.

From this experience it was clear how impressionable the children were, both by each other and observers. This was particularly useful for later testing of the game.

	Time
Group 1	20
Group 2	18
Group 3	8
Group 4	15
Group 5	12
Group 6	14
Average	14,5

Table 1: Prestudy: Time each group spent playing

5.6 Second User Test - Questionnaire Pre-Study

In order to test the usability of the questionnaire tools and question style, an early version of the questionnaire was tested on a sample of 11 children aged 10-13. The questionnaire included use of the Smileyometer (Illustrated in Figure 10), and The Again Again table (Illustrated on page 65). The main purpose of this pre-study was to answer the following questions:

1. Are children able to easily and quickly understand and fill out the questionnaire without assistance?
2. Do children understand the developed Smileyometer?
3. Do children understand the developed Again Again tables?
4. Do the children have an interest in Minecraft and Minecraft-related games?

It was desirable to observe whether or not the children expressed any difficulty in understanding the questionnaire. It was also of interest to test the universality of the questionnaire by testing it with different games. During a research evening at a local science center, children were allowed to try out different games through the Oculus Rift, a highly immersive virtual reality head-mounted display. The children were allowed to play for a couple of minutes, getting used to the sensation, before they had to take the headset off and answer the questionnaire. All children were asked if they would like to answer a few questions about their experience. No children were forced to answer the questionnaire. Each child was told to ask if there was something they did not understand. This early version of the questionnaire contained six questions. Few questions were chosen in order to make the process quick and efficient. As understanding of the questionnaire tools was to be tested, it was not deemed necessary to include more questions. The three first questions were Smileyometer coded 5-point likert scales where the very sad face equals “Disagree Strongly” and the very happy face equals “Agree Strongly” (See Figure 10). These questions were related to immersion and flow, such as asking the players if they felt like they were another place.

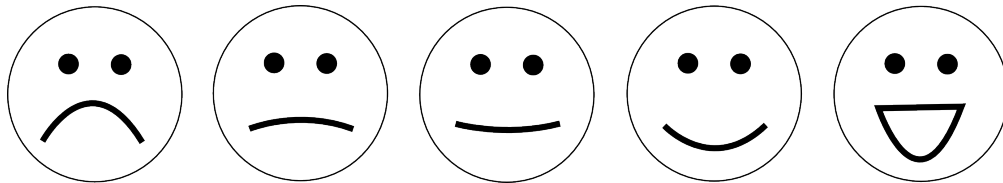


Figure 10: The developed Smileyometer

The last three questions asked if the player wanted to try the game again and their familiarity and feeling towards Minecraft and Minecraft-related games. These questions were coded according to the Again Again table, asking the player to cross out the square that fit. When asking if the player had played Minecraft before, only Yes/No answers were possible in order to force a boolean result.

Second User Test - Results

As with the early user testing with kindergarten children, it was clear how children affect each other. The game testing was situated in a science center open space where children could roam freely. The children that finished the game stayed around watching others play, often commenting with vigor and enthusiasm. This was expected behaviour and was deemed to not jeopardise the questionnaire as it was understanding of the questionnaire that was important, not the validity of self reported experience. Because of this, the measures of experience in this questionnaire cannot be seen as valid. The children were also able to try different games for the Oculus Rift. Some of the games were more active than others and the environments were very different. The questionnaire did not record which game the participants were playing, therefore it is impossible to say if variance in self reported experience was dependant on which game was played.

None of the children expressed any difficulty with the questionnaire and none asked for help with the questions. All filled out the questionnaire quickly and at approximately similar times. Several children did not notice the age field and had to be reminded to fill in their age. Some

children had to be reminded that there were questions on both sides of the paper handed out.

Eight of the 11 children had played Minecraft before and six of these expressed wanting to play a game that looks like Minecraft. No one answered "no" to playing a game that looks like Minecraft. All children who had not played Minecraft answered "maybe" on "Would you want to play a game that looks like Minecraft". Two out of the eight children that had played Minecraft also answered maybe to this question. This supports that the children were easily able to understand how to use the Again Again tables.

The distribution of answers to the various questions suggested that the children were able to understand the nuances between the smiley faces. As the children played games in a novel and highly immersive platform for the first time, the feedback was distributed in the positive part of the scale. It must be noted that there is a possibility that the children filled out the questionnaire quickly in order to proceed to other science center activities.

As the pre study suggested that children were able to understand the questionnaire style and fill out the form quickly, it was decided to include more questions in the final questionnaire. The final questionnaire kept the formatting and theme from the Game Experience Questionnaire, but cannot be counted as such.

5.7 Questionnaire Coding

In order to measure the player experience, it was desirable to use The Game Experience Questionnaire (GEQ) as it is a validated questionnaire for game experience [14, P.265]. This questionnaire measures seven aspects of player experience: immersion, tension, competence, flow, negative affect, positive affect and challenge. The questionnaire is based of focus group research and was developed by a group of experts through the EC-funded project Fun in Gaming (FUGA). The original GEQ contains questions that could be too difficult for children to answer. Since this thesis deals with children, there was a need for a modified version for children. The questionnaire was also previously only available in English, German and Swedish, making it necessary to translate it into Norwegian. The finished questionnaire was heavily inspired by the GEQ, but cannot be regarded as such as the questions were translated and altered for children. Therefore, the final questionnaire does not hold the same validity as the GEQ. The questionnaire kept the original categories since the questions still were related to the respective categories. The original 42 questions were shortened down to 8 and the language simplified for use with children. In cases with difficult words, the question was changed into a description of the word (Example: "it was aesthetically pleasing" was changed to "the game looks good". Questions about game story were removed as there was no story in the tested games.

The questions were organised in the following order based on the GEQ categories:

- **Immersion**

- It was exciting.
- It looked nice.
- I liked that it looked like Minecraft.

- **Flow**

- I felt like I was somewhere else.
- I forgot everything around me.

- **Competence**

I felt good at it.

- **Challenge:**

I found the game.. (Very hard, hard, Just right, Easy, Very easy).

- **Positive/Negative Affect**

I thought the game was.. (Very boring, boring, okay, Fun, A lot of fun).

The questionnaire used a 5-point likert scale. Using this kind of scale with children can possibly be a source of confusion if they don't understand the logic between the questions and the answers. To help overcome this, the Fun Toolkit was utilized. The Fun Toolkit is described as an instrument for gathering the opinions of children about technology and has been validated by a number of empirical studies [52] [53] [54]. The toolkit consists of three instruments: The Smileyometer, The Fun Sorter and The Again Again table.

The Smileyometer uses smiley faces and supporting text coded into a 5-point likert scale to represent the answers. This technique is widely used in research studies with children as it is easy to understand, quick to complete and requires no writing. As the sample of participants were of varying age, it was important to code the questionnaire in a manner that would be understandable to all. A possible weakness with the Smileyometer is that it could be priming for positive bias. The smiley faces communicate positive or negative emotion where positive is used where "Agree" usually is written in traditional likert scales. Agreeing isn't always a positive reaction. The questionnaire asked the participants if they found the game boring. Agreeing that something is boring isn't usually done with a smiling face. This could cause confusion on the meaning of the positive smiley. Especially if the positive smiley is agreeing in one question and disagreeing in another. However, if one were to ask if the participant found the game fun, the Smileyometer would immediately make sense, as agreeing with this can also be associated with a smile. This slight positive bias promotes writing questions with positive words. A possible solution to this problem is formulating neutral statements and using the supporting text to clarify the meaning. This solution can also prevent situation where a child has given conflicting answers such as reporting that a game was both boring and fun, both easy and hard. If one pays close attention to how the question is being asked and how the alternatives are coded, it might be possible to use the Smileyometer for clear and precise answers from children. The Again Again table is made specifically to ask the child if they want to do an activity again by asking them to tick the box for "yes", "maybe" and "no". The final questionnaire can be found in the Appendix, page 65.

5.8 Avoiding Bias

Systematic error, or bias, can severely affect the reliability of an experiment. ([50, p. 57-63]) The 5 main sources of bias are the following:

- Measuring Instruments
- Experimental procedures
- Participants themselves
- The experimental environment
- Experimenter behaviour

With this list in mind, measures were taken in order to avoid bias:

Survey

A possible bias questionnaires can create is that of leading questions. If the participant gets the impression that they are expected to answer in a positive manner, there is a higher risk that they will. This can destroy the validity of the questionnaire. When coding the questionnaire, each question was translated and created with this in mind. It was important to communicate a neutral question that encouraged the participants' honest opinion.

User Testing

Similarly to question bias, experimenter behaviour is described as a factor that can heavily affect the outcome of an experiment. This is especially important when the researchers are involved in the development of the systems/games being used. When conducting experiments, it is preferable to use a neutral researcher with no personal attachment to the games used. As experimenting and user testing is time consuming, it isn't always possible to find a qualified professional for this purpose. In these cases, it is crucial that researchers are aware of their language and general behaviour. In this case, it was important to never mention personal involvement to the participants as it might cause them giving better feedback in order to not hurt any feelings. A script was developed to ensure that each case was as similar as possible.

Environment

Participants might behave differently if the test environment is significantly different to the natural setting. Some might perform worse due to the nervousity of being watched by an observer. When it was important to make the participants feel relaxed and comfortable in the situation. As the test took place in a science center, it was assumed that the participants were ready to explore new settings. The lab environment didn't appear too different to other science center activities, which made it feel more natural to the participants. The participants were ensured that no one would know how they performed or what they thought about the game. This was done to promote a setting where the participants would feel comfortable about sharing their honest opinion. As all data is recorded within the game and through pre/post questionnaires, it was possible with casual observation during gameplay. This was done to make sure the participants could enjoy the game alone as much as possible, without feeling stress about performing well for an observer.

Participant bias

There was a possibility that some participants were familiar with Minecraft and therefore would be positively inclined toward the game. The gameplay does, however, not require any previous knowledge of Minecraft.

Data Logging

To be able to track player performance in the game, data logging was added to the existing game. This data made it possible to track the gradual progress of each player throughout their gameplay session. The game logged data based on player performance. The following list shows the quantitative data log:

- Time spent: Game Session
- Time spent: Each Mini Game
- Time spent: Leveling up in each Mini Game
- Starting and quitting a minigame
- Wrong clicks
- Correct clicks
- Difficulty level

5.9 Third User Test - Questionnaire Pre-Study

The third user test was conducted to test the procedure structure and final questionnaire before the main experiment. The testing was conducted in a separate and neutral room in a local science center. The same research environment was utilized in the main experiment and this test acted as a verification of the environment suitability as well as the procedure. The group of participants consisted of four children with ages varying from 8-14. The children were recruited when visiting the science center. This is a possible bias as the children were already primed to try out new things with enthusiasm. One could also argue that children who go to science centers are already prone to be enthusiastic about new things and technology.

The test followed the following procedure:

1. The participant is welcomed to the experiment. S/he is shown the facilities and explained how the experiment will be conducted and what they can expect.
2. Give participant pre questionnaire to fill out.
3. Show game to participant and explain how to start and switch between games. Let the participant know they can switch games and stop at any time. Make sure the participant knows they can ask questions at any time.
4. Let the participant start playing the game freely.
5. Stand back to the observers table and casually observe the participant's behaviour. Do not stand behind participant and stare at their gameplay.
6. Preferably, make sure the participant has tried all five mini games.
7. Give participant the post game questionnaire and tell them to ask if there is a question they don't understand.
8. Control that all pages of questionnaire has been filled out.
9. Thank participant again for participating.

Observations and results of Third User Test

This test provided useful input how the research environment should be handled. The parents of one child wanted to observe the gameplay. The parents were informed that the child should be allowed to play undisturbed, yet continued commenting on the child's performance during gameplay. After the test, the scenario of intervening parents was added to the procedure list of extraordinary scenarios. The importance of thanking the participant and taking notes of any unforeseen events was added to the procedure.

5.10 Fourth User Test

Participants

The sample consisted of voluntary participants recruited at the local science center in Gjøvik, Oppland county, Norway. The sample consisted of 27 children, mostly aged 7-12. Two participants marked themselves as 12+. 8 female, 19 male. All participants had played Minecraft before. All participants reported playing video games on a regular basis. 14 participants reported playing games every day and 12 participants reported playing a few times per week. One participant reported playing games less than every week. 52% of all participants reported higher proficiency using computers.

Research environment

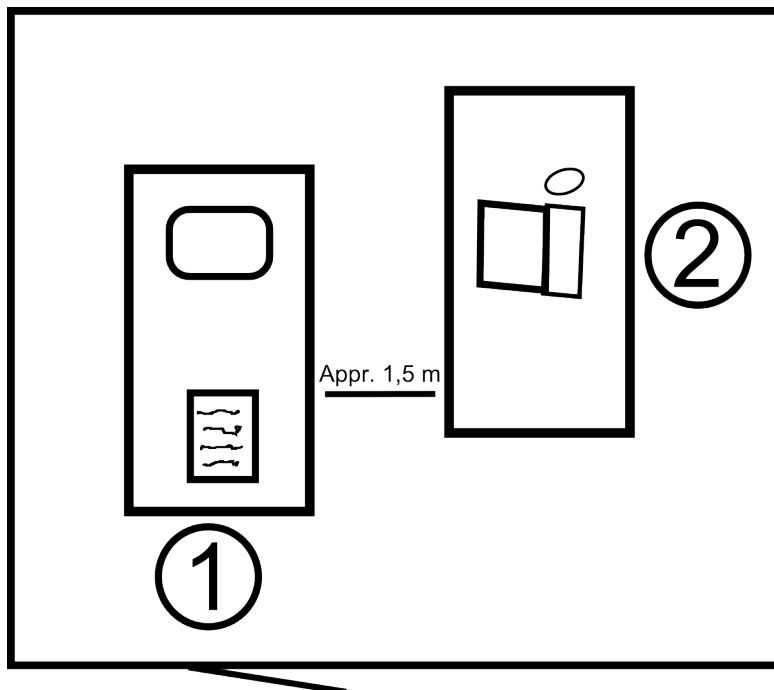


Figure 11: 1: Participant and table with tablet and questionnaire. 2: Observer and table with laptop.

Protocol

Setup

The experimental room was as neutral as possible, without distracting factors. Two tables and nearby power supplies were the minimum requirements for equipment within the room (See Figure 11). The science center didn't have chairs high enough for the tables, but the tables were designed for standing children (See Figure 12). The room was separated from other science center activities in order to avoid people coming in and distracting the subject. It is common for children to roam freely in a science center. The children could be recruited without parental



Figure 12: Research Environment

consent, thus it was decided to leave the door open to ensure that parents could spot their child if they had not seen them enter. There was only one researcher available for the session. Project posters were set up outside the door to make sure there was enough information for approaching parents or guardians (See Figure 13). In most cases parents were nearby and could be informed before any testing started.

Procedure

1. The participant is welcomed to the experiment. S/he is shown the facilities and explained how the experiment will be conducted and what they can expect.
2. Ask if the subject has any questions.
3. Give participant pre questionnaire to fill out.
4. Show game to participant and explain how to start and switch between games. Let the participant know they can switch games and stop at any time. Make sure the participant knows they can ask questions at any time.
5. Let the participant start playing the game freely.
6. Stand back to the observers table and casually observe the participant's behaviour. Do not stand behind participant and stare at their gameplay.
7. Preferably, make sure the participant has tried all five mini games. Take notes if they don't.
8. Thank the participant as gameplay ends.
9. Give participant the post game questionnaire and tell them to ask if there is a question they don't understand. Help the participant only if asked.
10. Control that all pages of questionnaire has been filled out.
11. Thank participant again for participating.



Figure 13: Research Environment Entrance

Extraordinary Scenarios

"Subject had to leave" Example: Subjects has forgotten an important appointment and needs to go immediately in the middle of the recording session.

1. Do not ask what business the subject needs to attend etc.
2. Inform the subject that it is not a problem to leave early and that it does not compromise the study.
3. Ask if the subject has time to repeat the session.
4. If possible, try to arrange time for a new session.
5. Note the event down for the analytic process.

"Parental supervision" Example: Parent or guardian wishes to observe the session.

1. Inform the parent about the experiment and offer them the project description. Hand over consent form(Found in Appendix A).
2. Politely inform about the importance on non disturbance during gameplay and questionnaire fillout.
3. If parent still talks to child during gameplay, do not disturb, but note it down for analysis.

Technical Issues

"Subject fails at game" Example: For some reason, the subject fails at playing the game to such an extent that no actual gameplay is being recorded. Subject accidentally quits the game.

1. Ask the subject, observe and outline what the problem is.
2. Tell the subject it is not a problem and that it does not compromise the study.
3. Try to accommodate the subject in eliminating the problem. Example: Explaining a game mechanic which hinders the subject. Restarting the game.
4. Try to finish the session as normally.
5. If it is not possible to continue, ask if the subject could attend another session.

6. Note the event down for the analytic process.

"Game/Device Crashes" Example: The game or device does not respond.

1. Assure the subject that it is okay and will not affect the final results.
2. Investigate the problem and try to fix it as quickly as possible.
3. If the problem is fixed quickly, resume the game session.
4. If the problem takes too long to fix, try to arrange a new session as quickly as possible.
5. Note the event down for the analytic process.

6 Results

This Chapter will present general observations and tendencies recorded during the user testing. Section 6.1 will describe the analysis strategy and data selection. Section 6.2 presents general results from the questionnaire. Section 6.3 shows the results from logged game behaviour. Section 6.4 will present a summary of observations made during the user tests. Section 6.5 presents results from relationship tests related to each hypothesis.

6.1 Inclusion Criteria and Data Cleanup

The study had 27 participants in total. Before data cleanup and analysis, each participant was screened with additional inclusion criteria. In order to be included in the data analysis, the following assumptions had to be met:

- The participant needed to have played all five minigames
- The participant needed to have completed the entire questionnaire
- The participant needed to have recorded game logs

These assumptions were added to ensure that no cases of null data could skew calculations. Due to these requirements, 7 participants were excluded from the final sample. Participant 6 and 25 did not finish the questionnaire and were therefore excluded even if game logs had been recorded. Participant 12-15 had played the game and finished the questionnaire, but were excluded due to lack of game log. Participant 23 was excluded due to only playing 4 of the minigames and not answering the questionnaire question about the lacking minigame. After exclusions, the final sample consisted of 20 participants. This was deemed enough to proceed with the planned analysis.

Questionnaire results were coded in an ordinal way of 0-4 where 4 represented the most positive answer. For example, 0 represented strongly disagree or "very boring" and 4 represented strongly agree or "a lot of fun".

The Again Again table was coded as suggested by Sim and Horton [52] with Yes coded as 2, Maybe = 1 and No = 0. This allowed one to create a game score based on the feedback of each of the 5 minigames. The game score was calculated for each participants by adding together the value of each minigame answer. This made it possible for each participant to give the game a maximum score of 10. The final game scores were calculated using the same procedure.

The game telemetry data was saved as XML-files which were imported to a spreadsheet format and analysed through SPSS.

6.2 Results from Questionnaire

Almost all participants were able to fill out the questionnaire without assistance. Only two of the younger participants needed help to read the questions. A few of the participants asked for clarification of the "it looked nice" question. In this case, the following examples were given as

clarification: "If you thought it was pretty or cool to look at, that you liked the graphics or the look of the game.". This seemed to be sufficient explanation.

50 % reported playing digital games every day and 45% a few times per week (See Figure 14). Only one participant reported playing digital games a few times per month. 100% of all participants had played Minecraft before.

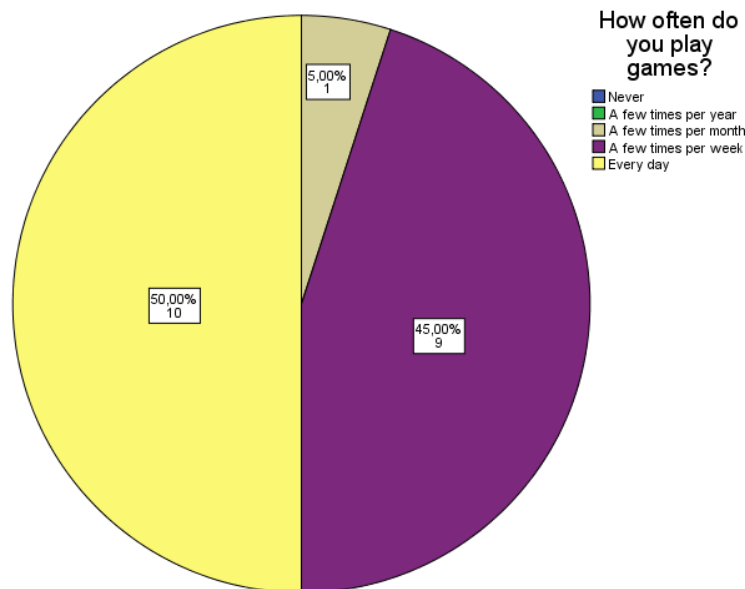


Figure 14: Self reported frequency of playing.

70% of the participants reported the difficulty in the game to feel "just right". No one described the difficulty as difficult or too difficult. The other 30% reported the difficulty to feel easy or too easy. Two participants called the game "too easy". 65% of participants agreed or strongly agreed to feeling skillful when playing the game. 10% reported to disagree/strongly disagree to feeling good at it(the game).

The last question in the questionnaire asked the participants to rate the overall game experience from boring to a lot of fun. 65% rated the game as "a lot of fun" and 15% as "fun", making up an 80% positive result. 15% rated the game as "A little bit of fun" and only one person labeled the game as "boring". No one chose "Very boring".

Again Again Table

Through the Again Again Table, the participants could give the game a score from 0-10 by scoring each minigame. Answers was coded as follows: "Yes"=2 points, "Maybe"=1 point and "No"=0 points. A score of 10 would imply that the participant wanted to try all minigames again. The final table revealed a slight preference of the Gem and Labyrinth minigames (See Table 2). The Gem and Labyrinth game also had the fewest participants answering "no" to try the minigame again. The Arrow minigame had the highest distribution of uncertain answers with 9 participants rating it as a "Maybe" and 5 as a "No". The mean score of the overall game was 6.4/10. The

	Yes	Maybe	No
1: Arrow Game	6	9	5
2: Gem Game	12	6	2
3: Labyrinth	14	3	3
4: Zombie	9	5	6
5: Animal	9	5	6

Table 2: Again Again Table results: Do you want to try it again?

answers were distributed at a mean of 22% "No", 28% "Maybe" and 50 % "Yes".

Immersion Metrics

There were three questions in the questionnaire related to immersion:

- It was exciting.
- It looked nice.
- I liked that it looked like Minecraft.

The immersion response was positive, with 50% reporting to strongly agree and 30% agreed. Participants reported to like the appearance of the game with 90% positive feedback, where 20% reported strong agreement, 70% agreement and 10% slight agreement.

90 % of participants reported positively to the game looking like Minecraft. Only one participant reported to dislike the resemblance. Figure 15 displays the full distribution of answers to these questions.

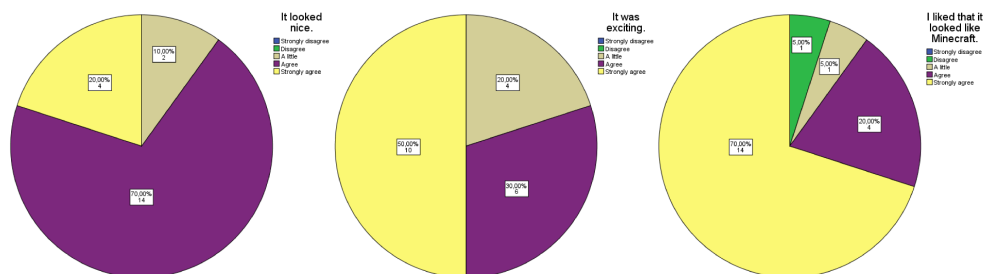


Figure 15: Reported immersion: From left to right: "It looked nice", "It was exciting" and "I like that it looked like Minecraft". Percentage and count of participants marked in each category.

All three questions garnered a general positive response, where 80-90% of answers were distributed on the positive side of the scale (Agree and Strongly Agree).

Flow Metrics

There were two questions in the questionnaire related to flow:

- I felt like I was somewhere else.
- I forgot everything around me.

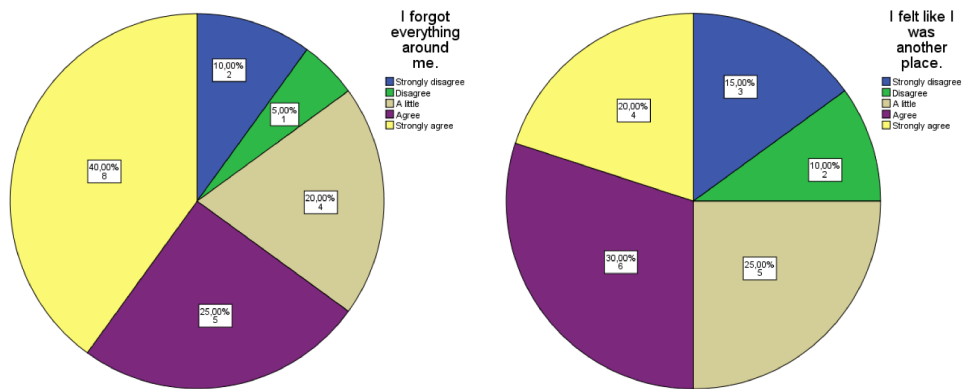


Figure 16: Flow Metrics: From left to right: "I forgot everything around me", "I felt like I was another place". Percentage and count of participants marked in each category.

The results from these questions were slightly more fragmented than immersion related questions. 15% reported not forgetting their surroundings and 25% disagreed to feeling like they were another place. The majority of the answers were distributed among the positive answers. 65% agreed or strongly agreed to forgetting their surroundings and 55% reported feeling like they were another place. Figure 16 displays the full distribution of answers to these questions.

6.3 Results from logged Game Behaviour

All participants were allowed to decide for themselves how long or short to play. If the duration exceeded 15 minutes and the participant had tried out all minigames, they were asked to proceed to the questionnaire. Durations of active gameplay ranged from 4 to 17 minutes with an average of 9 minutes. Figure 17 shows how the Labyrinth and Gem game were the games the children played the most when allowed to decide for themselves.

Measuring Success

Success in each minigame is different based on the gamedesign and overall style of the tasks. There were some limitations due to the data collected from the game. In addition to difficulty progression, success in each minigame was based on the following criteria:

- **Labyrinth**
Finding all items before time runs out
- **Gems**
Measured correct and wrong clicks.
- **Animals**
Measured correct and wrong clicks.
- **Arrows**
Measured correct and wrong clicks.
- **Zombie**
Measured correct and wrong clicks.

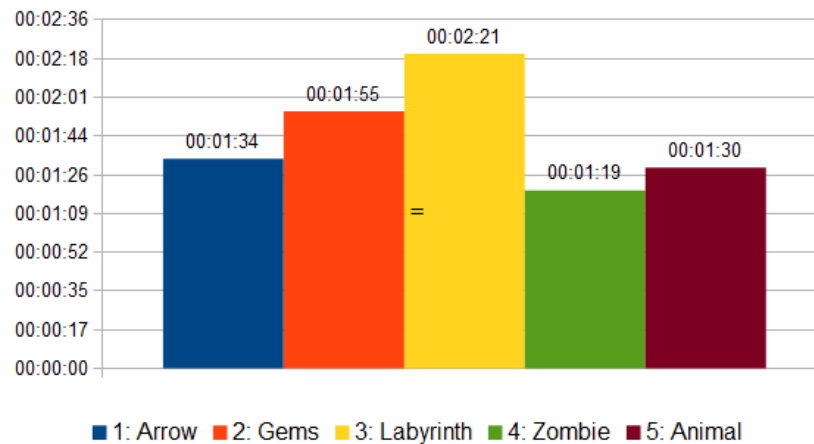


Figure 17: Mean playtime for each minigame.

Based on the recorded success/fail events, the participants had a mean of 12.7% fail rate. There was a vast difference in fail rate. Where some participants had a low fail rate at 1-7%, other participants were as high as 34%. Figure 18 shows how the distribution of failed actions still fits in a standard normal distribution.

6.4 Observational data summary.

This section summarises qualitative findings from observing the children playing the game.

General observations

The majority of the children appeared concentrated during gameplay and expressed that they found it enjoyable. As the door out to the science center was left open, there was occasional noise coming from the outside. Especially from a stand that measured the decibel level of the visitors screams. These noises did not appear to distract the participants away from the game. All participants were told to ask if there was anything they did not understand. Since everyone had been given this message, it was decided to not help unless asked.

One child was observed throwing their hands up in the air in celebration of completing a difficult task. In most cases in the game, only single taps were necessary. Three children would tap furiously instead of single tap. It was attempted to ask every child if they thought something could be improved about the game. Many children were already on their way out immediately after finishing the questionnaire. It was not deemed important enough to pursue the child for their opinion. Most children asked for their opinion about the game said that there was nothing to improve, even if they had struggled with one or more of the minigames. It is possible that the children gave this answer to quickly finish the user test. A participant that did very well in all minigames expressed that the difficulty should increase even more as one progresses.

No participants had trouble switching between the different minigames.

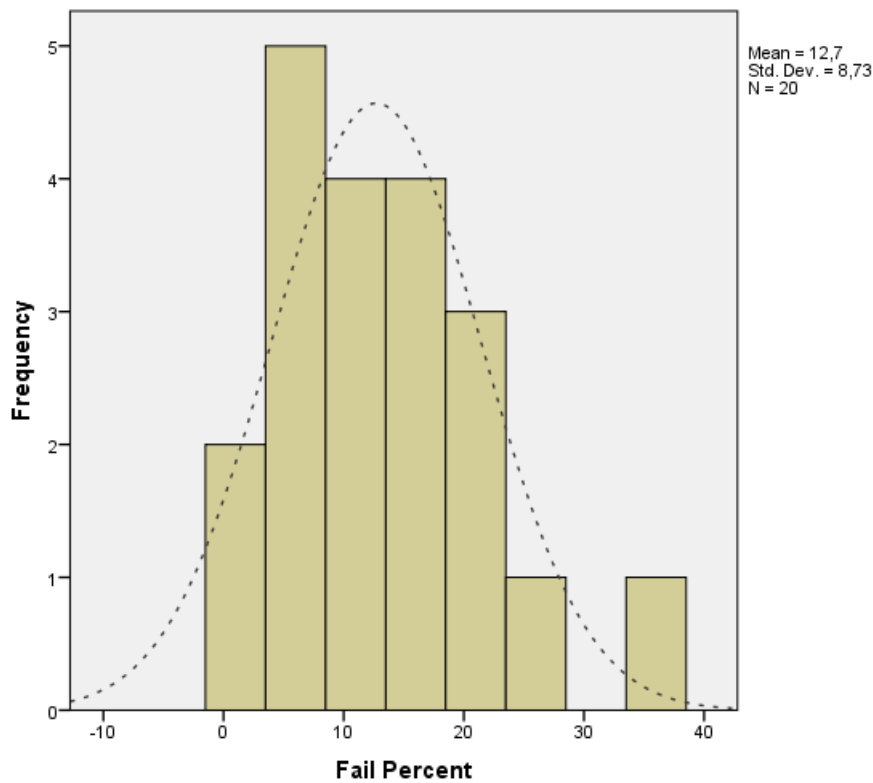


Figure 18: Distribution of fail failed actions for all participants

Arrow Minigame

10 participants struggled with understanding the Arrow game. The behaviour varied from tapping the signs for too long and tap dragging from the bow to the signs. Only one participant asked for an explanation of the game.

Gem Minigame

The vast majority understood the Gem minigame immediately. Only one participant asked for an explanation. Participants only failed at this minigame at harder levels.

Labyrinth Minigame

Three participants struggled with the controls of the Labyrinth minigame. One participant did not understand that "Start" had to be pressed to start the game and others had trouble moving the character around the labyrinth by using many small strokes instead of dragging. The majority expressed interest and concentration when playing the Labyrinth game.

Zombie Minigame

Four participants struggled with understanding the Zombie minigame. Many were observed tapping the correct spots before the zombies had disappeared. The participants who did this kept tapping furiously until the animation sequence finished and the icons became clickable. One

participant clicked the icons for too long.

Animal Minigame

All participants were given an explanation to this game, which at the time wasn't intuitive enough. The explanation "Click when the animal makes the right sound" was given to all participants and the vast majority appeared to understand this quickly. Only two participants asked for help when they didn't understand something in the game. Several children were observed as impatient when the game didn't progress immediately after they pressed the correct way.

6.5 Results and statistical tests related to each hypothesis

6.5.1 Research Question 1: Can we measure and observe improvement in gameplay?

Improvement and success in gameplay was measured by difficulty progression and wrong/successful game events. Observational data on success and failure was also recorded.

Hypothesis 1: Time spent playing can predict progression.

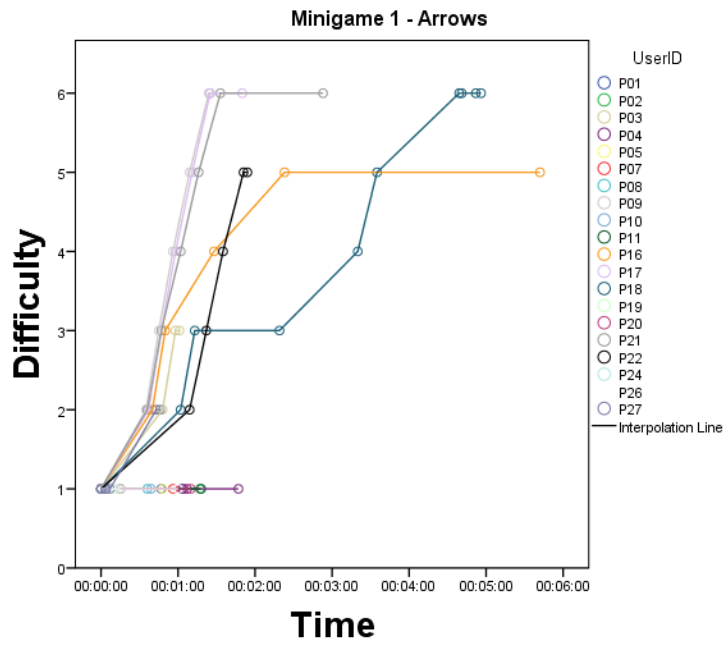
The scatterplots in Figure 19, 20 and 21 describe the gradual difficulty progression of all participants with recorded game log. The scatter plots show how there appears to be a linear relationship between time and difficulty. The further test this, a Pearson product-moment correlation coefficient was calculated to test for correlation between time spent and difficulty. This was done to test if there was high variance in player progression and if players were able to continue their progression through the minigame. Pearson correlation was chosen as the data was continuous, came out positive from normality tests and had a linear scatterplot without too significant outliers. There was a strong correlation between time spent and difficulty level. Pearson calculation details are included in the appendix, Section A.

Most of the minigames appear to have clusters of participants stopping at level 1.

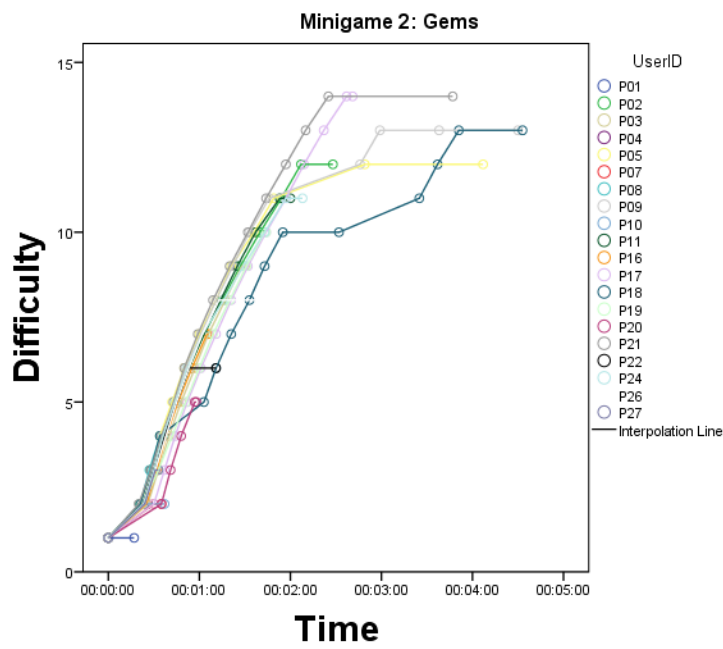
Minigame 1 shows some clustering of participants at level 1 as max difficulty. 8 participants reached levels higher than 1, with 6 of these reaching level 5-6. Level 5-6 appears to stall the progress curve of participant 16, 17, 18 and 21. Participant 16 reached 14 wrong clicks after level 5, but kept playing level 5 for 3:18 minutes before quitting the minigame. Participant 17 got an error cluster at lvl 6 and quit shortly after. Participant 21 had only two errors after level 6. Minigame 2 shows a small cluster of participants stopping at level 1, all playing under 1 minute. The progress curve is linear with little difference until level 10-15 where participants appear to spend more time. Minigame 3 also displays consistent progression, with some progression stalling after level 6. Minigame 4 shows relatively similar progression, with some breaks in the graphs. The breaks correlate with recorded clusters of wrong clicks in participant 10, 17 and 21. Minigame 5 shows some clustering at level 1 since many of the participants did not reach a higher difficulty level. None of these participants played for more than one minute. Only 5 participants reached levels above 1, where only 2 of these got above level 10. All participants above level 1 shared a similar progression curve.

Hypothesis 2: There is transfer in learning across the minigames.

In order to test for a learning effect across the minigames, it was of interest to see how the mean of mistakes made changed throughout the game session. During testing, participants had

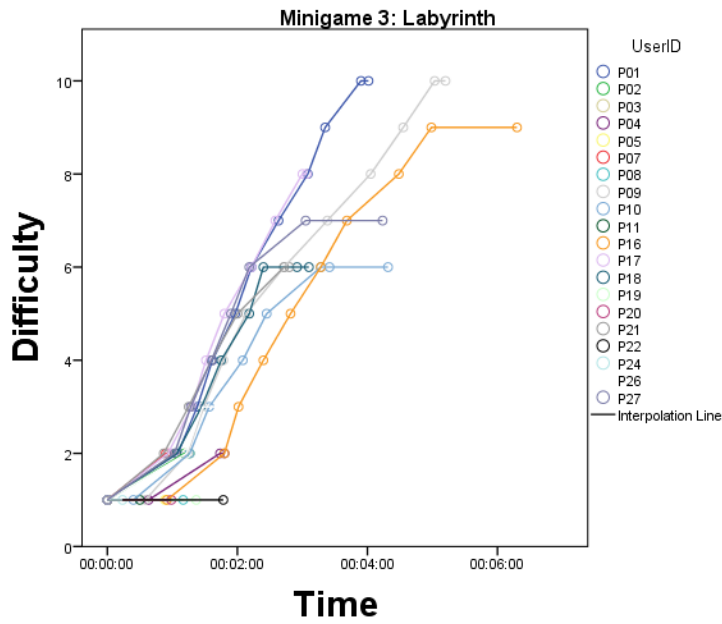


(a) Minigame 1

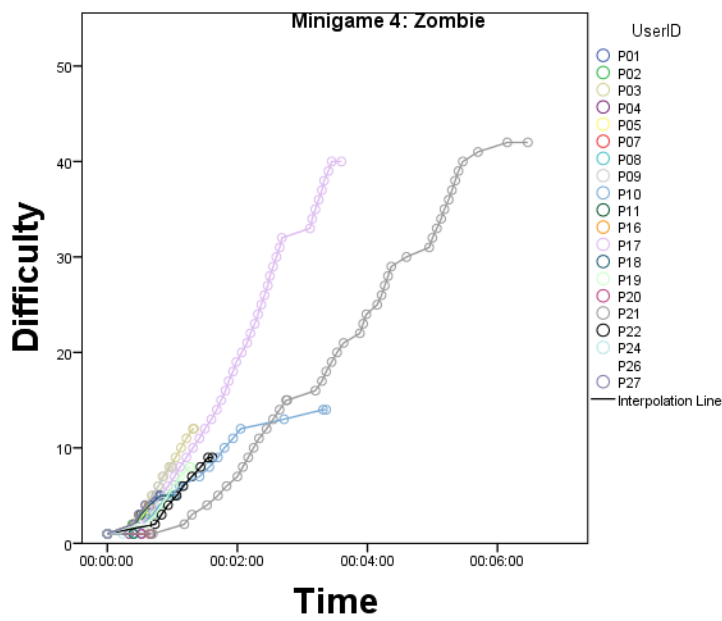


(b) Minigame 2

Figure 19: Progression of difficulty levels over time spent playing. All participants with recorded game logs included.

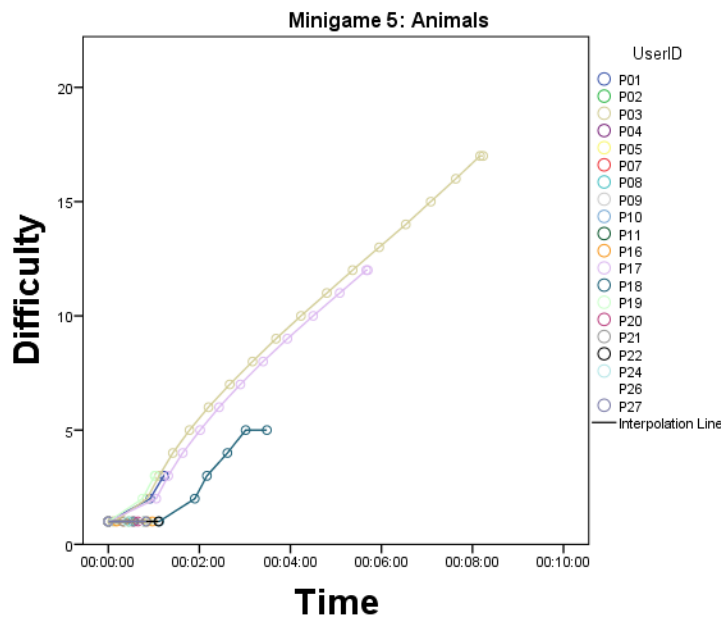


(a) Minigame 3



(b) Minigame 4

Figure 20: Progression of difficulty levels over time spent playing. All participants with recorded game logs included.



(a) Minigame 5

Figure 21: Progression of difficulty levels over time spent playing. All participants with recorded game logs included.

been allowed to play the game freely, therefore some had gone back and forth between the minigames. For conformity, participants who had gone back and forth between the games many times were excluded. Only participants who had played through all the minigames once were included. A sample of 13 participants were selected based on these criteria. Figure 22 shows how there appears to be a spike in mean mistakes made during the middle of the game session. The error y-axis refers to the mean of mistakes made in any minigame played in the game session. The "Order of play" x-axis refers to the order a person played the minigames where 1 refers to the first game played in the session and 5 refers to the last game played.

Figure 23 displays how each minigame had different means of mistakes depending on the order they were played. Three minigames had no mistakes when played as the last minigame.

6.5.2 Research Question 2: What makes children want to play a Working Memory Game?

Hypothesis 3: Reported immersion and flow can predict player intent to use.

It was of interest to test the relationship self reported immersion and flow could have with intent to use. In order to test the relationships, the measures of intent to use, flow and immersion were grouped into "high" and "low" measures. "Intent to use" was split into "High intent" and "Low intent". The basis for these categories were the measures from the Again Again Table (presented in Section 6.2). In order to be ranked as "High intent", one of the following assumptions needed to be met:

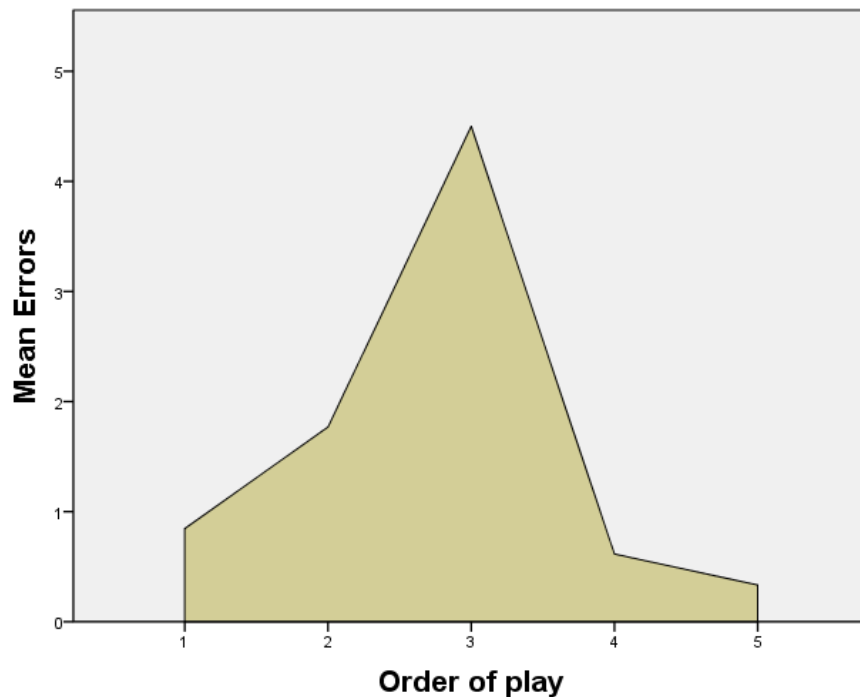


Figure 22: Graph showing how the mean mistakes made develop according to the order or played minigames.

- At least three "Yes" answers.
- At least two "Yes" and two "Maybe"

After applying these assumptions to the sample, 6 (30%) participants were marked as "low intent" and 14 (70%) as "high intent". This matches relatively well with the raw values from the Again Again Table where the mean values of "Yes" and "Maybe" answers made up 78% of answers. Three different questions were related to immersion. In order to be marked as a "High immersion" participant, each participant needed to pass the following two assumptions:

1. No disagreeing answers
2. At least two "Agree" or "Strongly agree" answers.

Participants who did not pass these assumptions were placed in the "Low immersion" category. After categorising the participants, there were 18 participants (90%) with "High immersion" and 2 participants (10%) with "Low immersion". This distribution appears to reflect the metric results displayed in Section 6.2 where 80-90% of answers were positive. Similarly, there were two questions related to flow. Participants needed to pass the following assumption to be ranked as "High flow":

1. No disagreeing answers.
2. At least one "Agree" or "Strongly agree" answer.

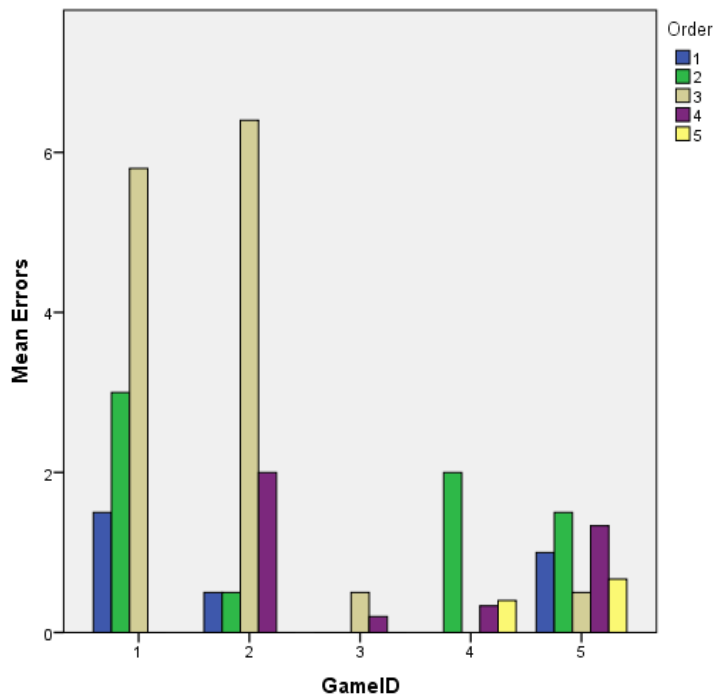


Figure 23: Clustered bar chart showing the mean mistakes in each minigame, based on the order they were played. All participants in this graph had played all minigames, a non-existent bar signifies no mistakes made.

Participants who did not pass these assumptions were marked as "Low flow" participants. After categorising the participants, 13 participants (65%) were marked with "High flow" and 7 participants (35%) were marked as "Low flow". This measure also reflects the flow metrics displayed in Section 6.2 where 55-65% of answers were positive. A 2x2 Chi-Square test was chosen to test the relationship between Intent to use and immersion. The Chi-Square test was chosen as the variables were categorical and ordinal. Each variable included two independent categorical groups. The Chi-square test revealed that there was a relationship between intent to use and immersion ($X^2(1) = 5.2, p(.023) < .05$). There was a statistically significant relationship between the two variables as the p-value was less than .05. This shows how the majority of participants with high intent also had high immersion. No participants with low immersion had high intent. Based on these results, the H_0 could therefore be rejected. A Fisher's exact test was also calculated, as the expected value of two cells were below 5. The Fisher's p-value (.079) was slightly above .05 and suggested no significance.

As the data was similar, the chi-square test was also applied to test for relationship between Intent to use and flow. The test showed that there was no relationship between flow and intent to use. ($X^2(1) = 3.7, p(.052) > .05$). The p value was slightly above 0.05, making the relationship very weak at best. The H_0 could not be rejected.

An additional Chi-square test between immersion and flow revealed how there was a relationship between these two variables ($X^2(1) = 4.1$, $p(.04) < 0.05$). High flow could only be linked to high immersion as no participants with low immersion had high flow.

	Chi-Square Value	df	Sig. (p-value)	Sig. (Fisher's exact p-value)
ITUxImmersion	5.2	1	.023	.79
ITUxFlow	3.7	1	.52	.12
ImmersionXFlow	4.1	1	.04	.11

Table 3: Statistical tests: Intent to use, flow and immersion. ITU= Intent to use, Df=Degree of freedom, Sig= Significance. Level of significance at .05.

Hypothesis 4: Participants who play the game longer find it more fun.

In order to answer this question, the null and alternative hypothesis was tested for relationships.

- H_{40} : There is no difference between time spent playing and reported levels of fun.
- H_{41} : There is a difference between time spent playing and reported levels of fun.

The independent variable used as a predictor was *time spent playing*. Participants were split in two groups, those that played the game up to 10 minutes and those who played above 10 minutes. Each participant had rated the overall game experience from boring to fun. As the data was vastly positive, the participants were split into two groups, those who found it fun and those who found it very fun. This was done to eliminate the lack of negative data. Only one participant (P27) had marked the game as boring. This participant response was considered a significant outlier and was excluded from the calculation.

The hypothesis was tested with a 2x2 Chi Square test. For SPSS calculation, the following assumptions are listed [55]:

1. The variables should be measured at an ordinal or nominal level
2. The variables should consist of two or more categorical, independent groups.

The collected data passes these assumptions. Lazar et al [50, Ch. 4] also mentions that sample sizes should be at least 20 in order to acquire a robust chi-square test. The sample size included 19 participants after excluding participant 27, but the chi-square test was still deemed appropriate.

The Chi-Square test was calculated and no relationship was found between time played and reported level of fun ($X^2(1) = 0.65$, $p(0.42) > .05$). There was no statistically significant relationship between the two variables as the p-value was above 0.05. The H_0 could therefore not be rejected.

The Chi-Square expected count of frequency turned out to be less than five. The expected count of frequency for each category is often recommended to reach at least a count of 5. In these cases, the Chi-Square test might not be counted as strong and the results could not be as meaningful. It has been argued that this requirement is too strict and that the 2x2 chi-square test still is robust as long as the sample size exceeds 10 and the expected count of frequency is no less than 1 [56]. In this case, the sample size was 19 and the minimum expected count of

frequency was 2.2. Therefore it was decided to keep the Chi-Square results and calculate other tests in order to validate or invalidate the findings.

The Fisher’s exact test is recommended when the expected count of frequency is below 5 as this test does not have the criteria of sample size. This can make the Fisher’s exact test a good choice for a smaller sample size. The Fisher’s exact test showed that there was no significant difference between overall time spent playing and reported level of fun. The calculated p-value was 0,61. The p-value needs to be 0,05 or lower in order to reject the null hypothesis. In this case, the p-value was above this level and the null hypothesis could not be rejected, supporting the findings from the Chi-Square test. Results displayed in Table 4.

Additionally, a clustered bar chart in Figure 24 helps illustrate and validate the result of no significant difference between time played and reported level of fun.

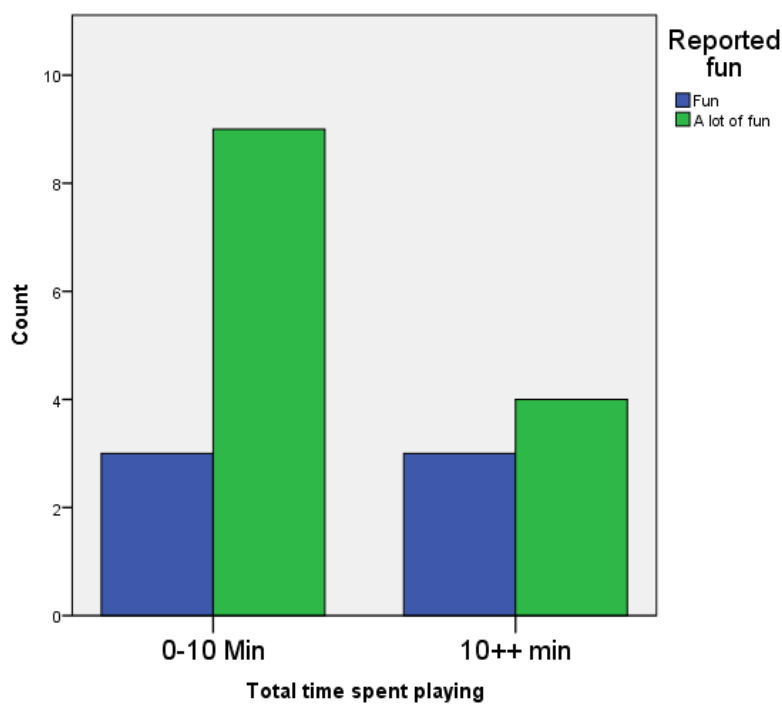


Figure 24: Clustered bar chart showing reported level of fun grouped by time spent playing.

	Chi-Square Value	df	Sig. (P-value)	Sig. (Fisher’s p-value)
TimeXFun	.65	1	.41	.61

Table 4: Statistical tests: 2x2 Chi-Square test and Fisher’s exact test. Time spent playing and reported level of fun. Df=Degree of freedom, Sig= Significance. Level of significance at .05.

This test could not be calculated on each minigame as the question of fun only referred to the entire game experience.

Hypothesis 5: Reported level of difficulty can predict time spent playing.

- H_{5_0} : There is no difference between time spent playing and reported levels of difficulty.
- H_{5_1} : There is a difference between time spent playing and reported levels of difficulty.

To test this hypothesis, the data needed some manipulation. None of the participants had reported the game to be difficult or too difficult. In order to avoid much null data in the calculations, the self reported difficulty variable was split into two new categories, "Easy" and "Just right". The "Time spent playing" variable had previously been split into two groups, "0-10 minutes" and "Above 10 minutes". The time variable was used for this calculation as well. A 2x2 Chi-Square test was run and there was no relationship between time spent and reported difficulty ($\chi^2(1) = 0.10, p(.91) > .05$) (See Figure 5).

As with the previous hypothesis, the expected Chi-Square expected count turned out to be less than five and the results might not be as strong. The Fisher's exact test was calculated and also displayed no correlation between time spent playing and reported difficulty (Fisher p-value: 1.0). The alternative hypothesis therefore had to be rejected.

	Chi-Square Value	df	Sig. (P-value)	Sig. (Fisher's p-value)
TimeXdifficulty	.10	1	.91	1.0

Table 5: Statistical tests: 2x2 Chi-Square test and Fisher's exact test. Time spent playing and reported level of difficulty. Df=Degree of freedom, Sig= Significance. Level of significance at .05.

Hypothesis 6: Concept familiarity will affect experience positively

100% of participants reported to having played Minecraft. 90% of participants reported agreeing or strongly agreeing to the Minecraft inspired graphics. This might indicate that the hypothesis has merit for future work. Due to lack of control possibilities, no statistical calculations were made.

7 Discussion

7.1 Hypothesis Discussion

7.1.1 Research Question 1: Can we measure and observe improvement in gameplay?

Hypothesis 1: Time spent playing can predict progression.

The time variable had a statistically significant correlation with progression through the difficulty levels. This is not a surprise, as it is expected to progress as one keeps playing a game. What this result reveals is that the balance of the game appears to be well adjusted. As some participants played the game for a shorter period of time, they naturally stop at earlier levels. The results from the tests suggest that if these were to play the game longer, they would also progress. This remains to be tested. The results from this test can only be interpreted as indicative, as the majority of participants did not progress to the very high levels. Game 1-4 appear to have players spend more time to progress as they reach the very high levels. There were only 4-5 participants who reached very high levels, making it hard to tell if this trend was significant or not.

The progression lines do not give info about the perceived difficulty of the game. Constant progression could also suggest that the game is too easy. The more difficult levels of the game do suggest that the difficulty was starting to catch up to the participant's skill. Data from the questionnaire tells us that 70% of participants rated the difficulty as "Just right", whilst 30% rated it "Easy" or "Too easy". This tells us that the majority of participants were happy with the difficulty. The observations made during the tests support this as the majority of the participants were observed as concentrated and interested during play. The clustered bar chart in Figure 25 shows how one is not likely to find any correlation between perceived difficulty and time played as the distributions are relatively similar.

Overall, the difficulty of the minigames appear to be well balanced, participants were improving as they played and were mostly pleased by the progression.

Hypothesis 2: There is transfer in learning across the minigames.

It was expected to see some transfer between the minigames as most of them have similar tasks and characteristics. The data gathered from the current gameplay is positive to the transfer of learning between the minigames.

The graph in Figure 22 displays how there is a peak around the 3rd played game of the session. This graph implies that participants gradually made mistakes until the skill required was mastered. This graph could be interpreted as a learning curve, showing how one makes a lot of mistakes while learning. Once a required skill has been obtained, there are fewer mistakes made. As this graph displays the mean of all 5 minigames, we can make assumptions of the game experience as a whole. The graph keeps the peaked shape after all minigame results are averaged. This suggests that the learning curve is independent of the order the minigames are played, meaning that transfer of learning happens between all games. This was clearly displayed in Figure 23 where the mean mistakes were zero or near zero for all minigames when played

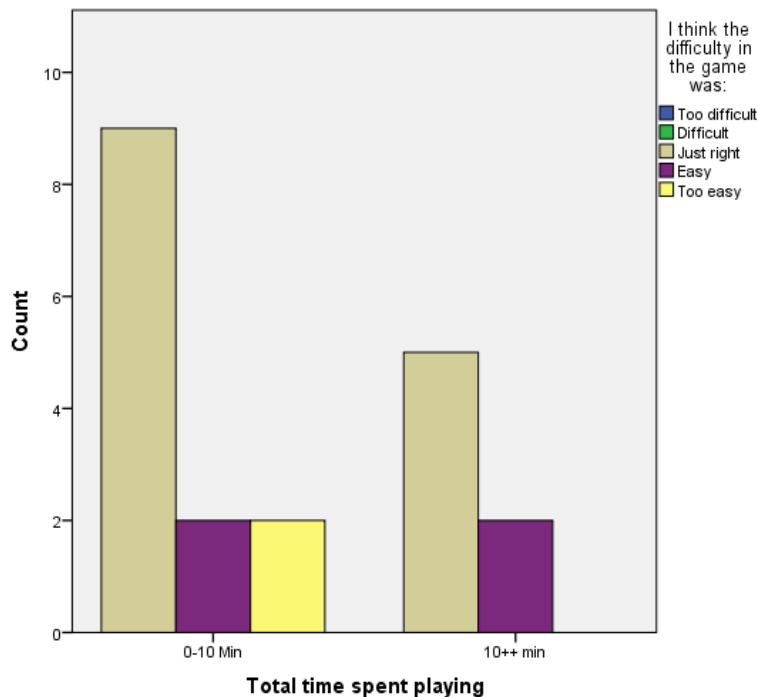


Figure 25: Self reported difficulty and time played.

last. This could mean that participants improve their overall skill as they play through the various minigames. Data from hypothesis 1 suggests that the game is well balanced. In light of this, one could hypothesize that the order of play does not matter as much in well balanced games.

These findings alone do not tell if the game is effective for improving actual working memory, but it lays the foundation for further exploration as there is a demonstrable learning effect.

7.1.2 Research Question 2: What makes children want to play a Working Memory Game?

Hypothesis 3: Reported immersion and flow can predict player intent to use.

The results from testing hypothesis 3 shows how there was a significant relationship between reported immersion and intent to use. Reported flow had no significant relationship to intent to use, but did show a slight relationship to immersion. This suggests that flow could be a predictor to immersion, which in effect is a predictor to intent to use. There are many similarities to how flow and early stages of immersion are described. According to Brown and Cairns, there is a strong link between immersion and flow [24]. The findings supports this sentiment. The findings from the statistical tests related to this hypothesis must be interpreted with caution as the sample size was not high enough to give a strong result.

Hypothesis 4: Participants who play the game longer find it more fun.

The relationship tests showed that there was no correlation between self reported fun and time spent playing. This was also apparent in the scatterplot which also showed no linear tendencies.

This indicates that participants who played the game for shorter periods of time found the game as fun as those who played longer.

It should be noted that the overall results of this question was positive. Only one participant rated the game as boring and no one chose "very boring". This might have affected the results. The results of these tests are also highly grouped. For simplicity, the variables were split into two subgroups. Time spent playing was reduced to "Under 10 Minutes" and "Over 10 Minutes" and "Fun"/"A lot of fun". This simplification could, in certain situations, have skewed the results. In this case, 65% rated the game as "a lot of fun" and 15% as "fun", 15% as "A little bit of fun". The original intent of this measure was to find out differences in experience, ultimately Fun/Not Fun. As the results were overall positive, it becomes harder to tell what would constitute a bad experience through time and reported fun measures. Therefore it was decided to group the feedback into "Fun" and "A lot of fun" to test for subtleties in the data. It is possible, however, that the children had different interpretation of what "Fun" and "A lot of fun" means. Due to these factors, we cannot claim that playing the game longer will make the experience more fun. At best, the results can be seen as indicative of how a good game might be experienced as fun even if played for short or long periods of time.

It is also possible that the game offered a fun experience that children like. The qualitative observations support this as the majority of the children appeared enthusiastic and concentrated when playing the game.

Hypothesis 5: Reported level of difficulty can predict time spent playing.

The results from the statistical test show that there was no relationship between time spent playing and reported level of difficulty. It is possible that the results from this test would have been different with a larger sample or a test situation where participants had to play for an extended amount of time. Participants could start and exit the minigames as they pleased and most did not play for a very long period of time. Many participants never reached the harder levels of the minigames. If one is to judge by the results from the participants who did well, it still didn't affect how they marked the game difficulty. This could mean that the overall game was too easy. When asked after playing, there were a few children that expressed how the game should have been more difficult or should progress faster.

Hypothesis 6: Concept familiarity will affect experience positively

In some sense, one could assume this hypothesis to be true as all children were positive and all had played Minecraft. However, due to lack of control, any conclusion would only be assumptions and wishful thinking. Due to its immense popularity, it is difficult to find children who have not played Minecraft. It would have been possible to test this hypothesis by using the same game with different graphics as control. This could be done in a similar manner as Nacke and Lindley [23] suggests in their work on creating tweaked versions of the same game as control conditions. While these results aren't proof, it is still safe to say that the familiarity was vastly positive and never negative. Many participants were interested in the game because of the familiar graphics and expressed curiosity and excitement when looking at the graphics. This is an uplifting result which is positive to further exploration.

7.2 Participant Bias

There were more male than female participants (8 girls and 12 boys). There was no apparent gender difference in the response, but there is always a chance of bias due to this. The majority of the participants were attracted to the Minecraft poster outside the test lab and expressed that they wanted to try it because it looked like Minecraft. All participants reported that they had played Minecraft before. There could be nuances to this as the questionnaire did not ask how much the participants had played Minecraft. Some participants could have been seasoned Minecraft players while others could only have tried the game a few times. The familiarity with Minecraft could also have affected how the participants reacted to and rated the game. This might also be a possible explanation to the positive bias in the answers. The circumstances make it possible to assume that many participants were above average in interest in technology and computer games. As there were no participants who had not played Minecraft, this bias could not be tested in the user tests. Minecraft is an incredibly common game among Norwegian children, making it harder to find this group without adding many other biases.

7.2.1 Questionnaire

The initial questionnaire results were very positive. Children have been reported to sometimes choosing the extremes of scales [52] and there is reason to believe this happened. This behaviour could be observed in the survey answers as the majority of answers were positive. One participant answered no to trying any game again, yet rated the game experience highly. All participants were volunteers recruited from the local science center. Not much is known about the children who visit science centers as a group versus those who do not visit science centers. It can be hypothesized that children who visit science centers are more likely to be enthusiastic about trying new things. Since parents or guardians often are the ones bringing children to science centers, it can also be hypothesized that these parents are more interested in science and technology, therefore influencing their children. Children who often visit science centers know that there will be new things to try and can therefore be primed to want this. If this is the case, this could offer an explanation to the positive results from the science center sessions.

7.3 Game Design

The Again Again table revealed how the participants favoured the Gem and Labyrinth minigames and were less inclined to want to replay the Arrow, Zombie and Animal minigames. Through observation it was noted how many participants failed to understand how to play the Arrow minigame. The only action needed in the Arrow minigame is tapping the correct order of numbers previously displayed. The game also displayed a bow that shot arrows to the sign the participant clicked.

The Gem and Labyrinth game was favoured by the children in rating and mean playtime. This could be caused by many factors such as game design. The Labyrinth will automatically take up more time as the player is required to navigate the toon through the labyrinth instead of simply tapping where the object were placed. This is probably the reason why the labyrinth got a higher mean playtime than all the other minigames. It is important to note that the test environment could also have had affect on the average playtime. This might have affected the playtime as the

participants could be eager to continue their visit to the science center.

Most of the minigames used in this study are based of the Simple Span tests, which are used to test short term memory. As working memory is connected to complex cognitive tasks, these tasks should be given priority. The Labyrinth and animal minigames are the minigames with the more complex tasks. There is a lack of complex tasks in the minigames. The N-back and running memory span tests are examples of complex span tests also used in current cognitive training games. Future development of the game could include more tasks related to or inspired by these tests.

Observational data revealed how all participants needed an explanation to understand the animal game. Even though an explanation allowed the participants to continue without problems, it revealed that the current state of this minigame is dependant on an instructor. It might be necessary to add a tutorial mode to explain this game in greater detail. Some complex tasks require more explanation than others. Most of the games had relatively simple tasks. If future versions of the game are to include more complex tasks, it will be important to include tutorial modes or other ways of explanation.

The start and continue buttons should also be improved as some participants spent some time waiting for the game to start, not understanding that they needed to click the button.

It was also observed that many participants reacted quicker than the game would allow. In some cases, the participants had started tapping the sequence of numbers before the sequence had disappeared.

7.4 Validity of Questionnaire

When discussing validity we have to ask ourselves if we actually measure what we think we measure.

All questionnaire questions after the Again Again Table were relating to the overall game experience instead of each minigame. Participants showed a varied preference of the minigames, clearly preferring some minigames over the others. The only questionnaire data that directly targets each minigame is the Again Again Table. This is a limited set of data which does not reveal differences in minigames other than wish to try again. There was an observed difference in how the participants understood and experienced the different minigames. There is a possibility that one minigame could affect the entire game experience positively or negatively. This possible bias is not available through the questionnaire and therefore weakens its validity. Because of this, one should be careful about drawing conclusions about the game experience based of the questionnaire alone.

7.5 Weaknesses of this study

The study included only one play session per participant. Participants were completely voluntary. This can both be a weakness and a strength. The participants had no pressure and could control the situation, how long they wanted to play each minigame and which order. They could stop at any time. It is possible that some participants tried the game for a shorter time because they felt no obligation to play a long time. The participants were recruited from the local science center. Therefore it is possible that the participants wanted to continue their visit and rushed

through the minigames and the questionnaire. This might affect the reliability of the game and questionnaire data.

Only 7 of the final 20 participants played the game over 10 minutes. This might be the reason the clusters of participants at the lower levels of the game.

The test environment wasn't completely shielded from noise or intrusion. The room was separate from the science center, but the door was left open. This was chosen because the children could give consent themselves and the guardian wasn't always next to them. Therefore the door was left open to make it possible for an accompanying guardian to see where the child was located. In most cases this caused no issues, but there were a few cases where the guardian or other visitors started talking through the door. For the most part, it did not seem to weaken the participants' concentration, but it is possible for this to affect the final results.

As with all studies with a great amount of raw data, there is always the risk of error during the data cleaning. Keeping data clean is the most important part of data mining as any mistake will corrupt all later calculations. Manual cleanup was kept to a minimum before moving data over to SPSS for calculations. Any name changes were automated in order to avoid manual errors. There were few inconsistencies in the raw data and the majority could be imported to SPSS as it was recorded.

Recruiting atypical participants

When conducting this study, it was considered appropriate to use neuro atypical participants. Since the game used was primarily intended to train those with cognitive impairment, such as adhd, typical participants were also wanted as a possible control condition. Recruiting young atypical participants proved to be a challenge. When recruiting children, there are many steps of consent that need to be in place before testing can be conducted. As a consequence of this, the opportunity to test with atypical participants did not present itself until very late in the process. It was decided that there was not enough time to include these participants and only include participants with no known disabilities.

8 Conclusion

This study has shown how there may be a transfer effect happening between working memory minigames and that it is possible to measure improvement in gameplay by using objective game measures. The results also show how children are positive to a game for working memory training and how immersion might be thought of as a predictor to the continued use of an application.

The possible transfer of skill is an interesting finding, as it suggests improvement across minigames. This could be of interest to WM game research. The promise of WM training is that training a specific ability can transfer into other abilities. While this finding doesn't provide proof of improved WM skills, all the minigames are based on WM tests that are used in well known training programs. There is still uncertainty of how far the transfer of skill can reach and if there is validity to this style of training (See Section 3.5). If skills from the WM game can be correlated to improved WM abilities, there is reason to believe that this game could be used instead of other commercial titles.

The positive feedback from the questionnaire and observations heavily suggests that children are happy to try out a new game for WM training. Being familiar with the game concept may be a catalyst for attention, which can be hard to maintain, especially among children with attention deficiency.

Overall, the findings from this study provides an opportunity to be optimistic about the use of games as a supplementary tool in diagnostics and training of WM training.

9 Future Work

The most important aspect to keep investigating is how a group of atypical participants would react to the same conditions. Even though a typical participant could benefit from playing a WM game, the potential benefit is much higher for someone with weakened cognitive abilities. In addition, future studies should explore if repeated use of the game affects the working memory ability as a whole or if the transfer stays within the tasks from the game.

A future study should also seek to improve the game logging system and the questionnaire. The logging system should be able to describe each play session in much greater detail in order to make sure the data is as valid as possible. The questionnaire should be extended to include follow up questions after each minigame.

In this study, participants were allowed to freely chose the order and time spent playing the game. Future work could test if results would be similar if the children had to play each game in a set order and amount of time. It would be interesting to test if the play order of minigames affects the results. Future studies should ask the questions related to game experience after each minigame. Participants could be asked to systematically play each minigame and fill out survey questions straight after each game. After playing through all games, the participant would answer survey questions relating to the overall experience of the game. Participants could be split into groups the followed different orders of play. Different conditions should be explored in order to test if the learning curve and transfer effect observed in Section 6.5.1 still stands. A possible condition to test would be a version of the game with neutral graphics compared to a graphics heavy version. Another possibility would include removing all feedback from the current game and testing if this changes the enjoyment and performance in a control group.

There is also an interest in how the metrics from this game could be used to display progress and recorded changes. Future development could add a more detailed logging system that displayed the game metrics as as easily understandable reports available for parents and health professionals. The possibility of using working memory games as assessment rather than treatment could be explored. A similar game system could be developed by translating more of the standard working memory tests into minigames. The system could be made to report positive, negative and no progress to health professionals as a way of assessing progress of current treatment. A system like this might be a positive way of making tedious assessment tasks more fun for children and other patients.

Bibliography

- [1] Folkehelseinstituttet. Folkehelse rapporten 2014. Technical report, Helse- og omsorgsdepartementet, 2014. URL: <http://www.fhi.no/publikasjoner-og-haandboker/folkehelse rapporten> (Visited 02.10.2014).
- [2] Prins, P. J., DAVIS, S., Ponsioen, A., Ten Brink, E., & Van der Oord, S. 2011. Does computerized working memory training with game elements enhance motivation and training efficacy in children with adhd? *Cyberpsychology, behavior, and social networking*, 14(3), 115–122.
- [3] Shipstead, Z., Hicks, K. L., & Engle, R. W. September 2012. Cogmed working memory training: Does the evidence support the claims? *Journal of Applied Research in Memory and Cognition*, 1(3), 185–193. URL: <http://www.sciencedirect.com/science/article/pii/S2211368112000629>, doi:10.1016/j.jarmac.2012.06.003.
- [4] Medietilsynet. Barn og medier 2014. Technical report, Kulturdepartementet, 2014. URL: http://www.medietilsynet.no/Documents/Barn%20og%20medier-unders%C3%B8kelsene/Rapport_Barnogmedier_2014.pdf (Visited 02.10.2014).
- [5] Association, E. E. S. 2014 sales, demographics, and usage data. Technical report, 2014. URL: http://www.theesa.com/facts/pdfs/ESA_EF_2014.pdf (Visited 02.10.2014).
- [6] Research, N. Z. G. M. The global games market report 2014. Technical report, 2014. URL: <http://www.newzoo.com/insights/global-games-market-will-reach-102-9-billion-2017-2/> (Visited 02.10.2014).
- [7] Medietilsynet. Barn og medier 2012. Technical report, Kulturdepartementet, 2014. URL: http://www.medietilsynet.no/PageFiles/11282/barnogmedier_2012_oppdateret_ISBN.pdf (Visited 02.10.2014).
- [8] Minecraft (game), 2009, Mojang. Mojang, Microsoft Studios (dev.). URL: <https://minecraft.net/> (Visited Nov.2014).
- [9] Clark, D. B., Tanner-smith, E. E., & May, S. K. 2014. Digital games for learning: A systematic review and meta-analysis.
- [10] Melby-lervåg, M. & Hulme, C. 2012. *Is Working Memory Training Effective? A Meta-Analytic Review*. URL: <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.281.7759&rank=1>.
- [11] Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J. S., Weems, S. A., Smith, V., Bobb, S., Bunting, M. F., & Dougherty, M. R. September

2013. Training working memory: Limits of transfer. *Intelligence*, 41(5), 638–663. URL: <http://www.sciencedirect.com/science/article/pii/S0160289613001062>, doi:10.1016/j.intell.2013.07.013.
- [12] DAVIS, S., VAN DER OORD, S., WIERS, R., & PRINS, P. 2012. Can motivation normalize working memory and task persistence in children with attention-deficit/hyperactivity disorder? the effects of money and computer-gaming. *Journal of Abnormal Child Psychology*, 40(5), 669–681. URL: <http://dx.doi.org/10.1007/s10802-011-9601-8>, doi:10.1007/s10802-011-9601-8.
- [13] SHAW, R., GRAYSON, A., & LEWIS, V. 2005. Inhibition, ADHD, and computer games: The inhibitory performance of children with ADHD on computerized tasks and games. *Journal of Attention Disorders*, 8(4), 160–168.
- [14] NACKE, L. 2009. Affective ludology: Scientific measurement of user experience in interactive entertainment.
- [15] SEIF EL-NASR, M., DRACHEN, A., & CANOSSA, A. 2013. Introduction. In *Game Analytics*, Seif El-Nasr, M., Drachen, A., & Canossa, A., eds, 3–12. Springer London. URL: http://dx.doi.org/10.1007/978-1-4471-4769-5_1, doi:10.1007/978-1-4471-4769-5_1.
- [16] HAZAN, E. 2013. Contextualizing data. In *Game Analytics*, Seif El-Nasr, M., Drachen, A., & Canossa, A., eds, 477–496. Springer London. URL: http://dx.doi.org/10.1007/978-1-4471-4769-5_21, doi:10.1007/978-1-4471-4769-5_21.
- [17] HUIZINGA, J. 1949. *Homo Ludens*. Routledge.
- [18] SMITH, R. February 2010. The long history of gaming in military training. *Simulation & Gaming*, 41(1), 6–19. URL: <http://sag.sagepub.com/content/41/1/6>, doi:10.1177/1046878109334330.
- [19] VBS3 (game), 2014, Bohemia Interactive Studios. Bohemia Interactive Simulations (dev.). URL: <https://bisimulations.com/> (Visited 02.10.2014).
- [20] DETERDING, S., DIXON, D., KHALED, R., & NACKE, L. 2011. From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek '11, 9–15, New York, NY, USA. ACM. URL: <http://doi.acm.org/10.1145/2181037.2181040>, doi:10.1145/2181037.2181040.
- [21] INITIATIVE, T. S. G. The serious games initiative (online). 2014. URL: <http://www.seriousgames.org/> (Visited Nov. 2014).
- [22] MCGONIGAL, J. 2011. *Reality is broken: Why games make us better and how they can change the world*. Penguin.

- [23] Nacke, L. E. & Lindley, C. A. 2010. Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. *arXiv preprint arXiv:1004.0248*. URL: <http://arxiv.org/abs/1004.0248>.
- [24] Brown, E. & Cairns, P. 2004. A grounded investigation of game immersion. In *CHI'04 extended abstracts on Human factors in computing systems*, 1297–1300. ACM. URL: <http://dl.acm.org/citation.cfm?id=986048>.
- [25] Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. September 2008. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9), 641–661. URL: <http://www.sciencedirect.com/science/article/pii/S1071581908000499>, doi:10.1016/j.ijhcs.2008.04.004.
- [26] Csikszentmihalyi, M. 2000. *Beyond boredom and anxiety*. Jossey-Bass.
- [27] Nakamura, J. & Csikszentmihalyi, M. 2002. The concept of flow. *Handbook of positive psychology*, 89–105.
- [28] Nacke, L. E., Drachen, A., Kuikkaniemi, K., Niesenhaus, J., Korhonen, H. J., Poels, K., & IJsselstein, W. 2009. Playability and player experience research.
- [29] Al-Aulamie, A., Mansour, A., Daly, H., & Adjei, O. June 2012. The effect of interinsic motivation on learners' behavioural intention to use e-learning systems. In *2012 International Conference on Information Technology Based Higher Education and Training (ITHET)*, 1–4. doi:10.1109/ITHET.2012.6246057.
- [30] Prasolova-Førland, E., Fominykh, M., & Darisiro, R. January 2013. Virtual afghan village as a low-cost environment for training cultural awareness in a military context. In *Scaling up Learning for Sustained Impact*, Hernández-Leo, D., Ley, T., Klamma, R., & Harrer, A., eds, number 8095 in Lecture Notes in Computer Science, 569–572. Springer Berlin Heidelberg. URL: http://link.springer.com/chapter/10.1007/978-3-642-40814-4_60.
- [31] Authority, N. M. Dataspill i skolen (computer games in school (online)). 2014. URL: <http://dataspilliskolen.no> (Visited Nov. 2014).
- [32] Global conflicts (game), 2007. Serious Games Interactive (dev.). URL: <http://globalconflicts.eu/> (Visited Nov.2014).
- [33] Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. 2013. A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105(2), 249.
- [34] Ibrahim, R., Yusoff, R. C. M., Khalil, K., & Jaafar, A. January 2011. Factors affecting undergraduates' acceptance of educational game: An application of technology acceptance model (TAM). In *Visual Informatics: Sustaining Research and Innovations*, Zaman, H. B., Robinson, P., Petrou, M., Olivier, P., Shih, T. K., Velastin, S., & Nyström, I., eds, number 7067 in Lecture Notes in Computer Science, 135–146. Springer Berlin Heidelberg. URL: http://link.springer.com/chapter/10.1007/978-3-642-25200-6_14.

- [35] Davis, F. D. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. In *MIS Quarterly*, volume 13, 318–339.
- [36] Tang, D. & Chen, L. 2011. A review of the evolution of research on information technology acceptance model. In *2011 International Conference on Business Management and Electronic Information (BMEI)*, volume 2, 588–591. doi:10.1109/ICBMEI.2011.5917980.
- [37] Yusoff, A., Crowder, R., & Gilbert, L. March 2010. Validation of serious games attributes using the technology acceptance model. In *2010 Second International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*, 45–51. doi:10.1109/VS-GAMES.2010.7.
- [38] Murah{"id":201929, Mohd Zamri and Zamri", "First_name":"mohd and "last_name":"murah" and "page_name":"mohdzamrimurah" and "domain_name":"ukm" and Murah", "Display_name":"mohd Zamri and "url":"http://ukm.academia.edu/mohdzamrimurah" and "is_analytics_public":false and "photo":"/images/s65_no_pic_borderless.gif" and Malaysia", "university":{"name":"National University of and "url":"http://ukm.academia.edu/" and Maklumat", "department":{"name":"Faculti Teknologi Sains and "url":"http://ukm.academia.edu/Departments/Faculti_Teknologi_Sains_Maklumat"} and Member"}, p. Towards educational games acceptance model (EGAM): A revised unified theory of acceptance and use of technology (UTAUT). URL: http://www.academia.edu/2744902/Towards_Educational_Games_Acceptance_Model_EGAM_A_Revised_Unified_Theory_of_Acceptance_and_Use_of_Technology_UTAUT_.
- [39] Shipstead, Z., Redick, T. S., & Engle, R. W. 2012. Is working memory training effective? *Psychological Bulletin*, 138(4), 628–654. URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0027473>, doi:10.1037/a0027473.
- [40] Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. 2004. Neuroplasticity: Changes in grey matter induced by training. *Nature*, 427(6972), 311 – 312. URL: <http://search.ebscohost.com/login.aspx?direct=true&db=afh&AN=12046926&site=ehost-live>.
- [41] Cogmed (game), 2014, Pearson Education. URL: <http://www.cogmed.com/> (Visited 11.11.2014).
- [42] Lumosity (game), 2014, Lumos Labs. URL: <http://www.lumosity.com/> (Visited 11.11.2014).
- [43] Cognifit (game), 2014, Cognifit. URL: <https://www.cognifit.com/> (Visited 11.11.2014).
- [44] Jungle memory (game), 2014, Memosyne Ltd. URL: <http://junglememory.com/> (Visited 11.11.2014).
- [45] Mindsparke (game), 2014, mind evolve, llc. URL: <http://www.mindsparke.com/> (Visited 11.11.2014).

- [46] Roche, J. D. & Johnson, B. D. May 2014. Cogmed working memory training product review. *Journal of Attention Disorders*, 18(4). URL: <http://jad.sagepub.com/content/18/4/379>, doi:10.1177/1087054714524275.
- [47] Egeland, J., Aarli, A. K., & Saunes, B.-K. 10 2013. Few effects of far transfer of working memory training in adhd: A randomized controlled trial. *PLoS ONE*, 8(10). URL: <http://dx.doi.org/10.1371/journal.pone.0075660>, doi:10.1371/journal.pone.0075660.
- [48] Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. 2008. Transfer of learning after updating training mediated by the striatum. *Science*, 320(5882), 1510–1512.
- [49] Rapport, M. D., Orban, S. A., Kofler, M. J., & Friedman, L. M. 2013. Do programs designed to train working memory, other executive functions, and attention benefit children with adhd? a meta-analytic review of cognitive, academic, and behavioral outcomes. *Clinical Psychology Review*, 33(8), 1237 – 1252. URL: <http://www.sciencedirect.com/science/article/pii/S0272735813001219>, doi:<http://dx.doi.org/10.1016/j.cpr.2013.08.005>.
- [50] J. Lazar, J. & Hochheiser, H. 2010. *Research Methods in Human-Computer Interaction*. Wiley.
- [51] Drachen, A., Seif El-Nasr, M., & Canossa, A. 2013. Game analytics – the basics. In *Game Analytics*, Seif El-Nasr, M., Drachen, A., & Canossa, A., eds, 13–40. Springer London. URL: http://dx.doi.org/10.1007/978-1-4471-4769-5_2, doi:10.1007/978-1-4471-4769-5_2.
- [52] Sim, G. & Horton, M. 2012. Investigating children’s opinions of games: Fun toolkit vs. this or that. In *Proceedings of the 11th International Conference on Interaction Design and Children*, 70–77. ACM. URL: <http://dl.acm.org/citation.cfm?id=2307105>.
- [53] Sim, G., Cassidy, B., & Read, J. C. 2013. Understanding the fidelity effect when evaluating games with children. In *Proceedings of the 12th International Conference on Interaction Design and Children*, IDC ’13, 193–200, New York, NY, USA. ACM. URL: <http://doi.acm.org/10.1145/2485760.2485769>, doi:10.1145/2485760.2485769.
- [54] Read, J. C. 2008. Validating the fun toolkit: an instrument for measuring children’s opinions of technology. *Cognition, Technology & Work*, 10(2), 119–128. URL: <http://link.springer.com/article/10.1007/s10111-007-0069-9>.
- [55] Statistics, L. Laerd statistics (online). 2014. URL: <https://statistics.laerd.com> (Visited Nov. 2014).
- [56] Camilli, G. & Hopkins, K. D. 1978. Applicability of chi-square to 2× 2 contingency tables with small expected cell frequencies. *Psychological Bulletin*, 85(1), 163.

A Appendix 1 - Informed Consent



Samtykkeskjema

Brukeropplevelse og evaluering av spill for kognitiv trening

Kontakt:

Hanne Fagerjord Karlsen: Master Applied Computer Science - Høgskolen i Gjøvik

hanne.karlsen2@hig.no Tlf: 99 26 90 45

Dr. Simon McCallum: Førsteamanuensis Høgskolen i Gjøvik simon.mccallum@hig.no

Prosjektbeskrivelse

Hanne Fagerjord Karlsen, masterstudent i Applied Computer Science v/Høgskolen i Gjøvik arbeider med masteroppgave om spillteknologi for helse. Dette prosjektet skjer i samarbeid med Universitetet i Oslo ved doktorgradstipendiat Charlotte Lunde, som også er utdannet lege. Gjennom bruk av spillteknologi er det mulig å trene opp flere kognitive egenskaper som arbeidsminne, hukommelse og impuls kontroll. Dette kan være nyttig for *alle* barn i alle aldre, kanskje spesielt dem med ADHD.

Dette prosjektet ønsker å bidra til forskningen i forståelsen av spill til nyttige formål.

Praktisk Informasjon

Tid: Undersøkelsen vil ta maksimum 10-15 minutter og barnet kan avslutte når som helst.

Deltakere i prosjektet vil spille et nettbrett-spill for arbeidsminnetrening. Spillet består av flere små spill som utfordrer visuell hukommelse og impuls kontroll. Følgende informasjon fra spillet vil bli lagret i en database: tidsbruk i spillet, riktige/feil klikk og opptjente poeng. Dette er anonym data som ikke kan kobles til en spesifikk deltaker. Denne informasjonen vil bli brukt til analyse i forskningen. Etterpå vil deltakeren bli bedt om å fylle ut en spørreundersøkelse om hvordan de opplever spillet. Undersøkelsen vil også spørre om alder, kjønn og hvor ofte barnet spiller dataspill. Det vil også bli utført et kort intervju om spillopplevelsen(anonymt).

Deltakelse i prosjektet er frivillig.

All data er anonymisert og vil ikke under noen omstendighet kunne bli koblet direkte til deltakeren. Vi har ingen opplysning om deltakerens medisinske bakgrunn.

- Jeg har lest og forstått samtykkeskjemaet og samtykker at mitt barn deltar i prosjektet.
- Jeg samtykker at uttalelser fra mitt barn kan benyttes i rapporten(anonymisert).

A Appendix 2 - Questionnaire

Alder:

7-8-9 ÅR	10-11-12 ÅR	12+ ÅR

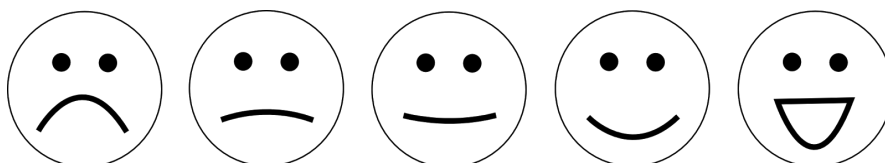
Kjønn:

Jente	Gutt

1: Hvor ofte spiller du dataspill(mobil, pc, mac, xbox, playstation)?:


Aldri	Noen ganger i året	Noen ganger i måneden	Noen ganger i uken	Hver dag

2:Jeg er flink med data:

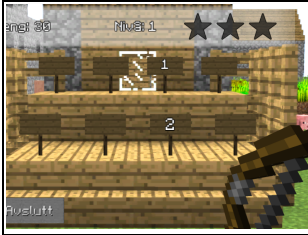
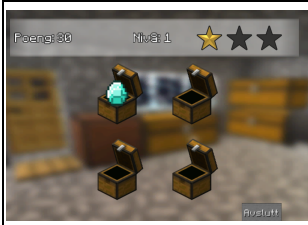
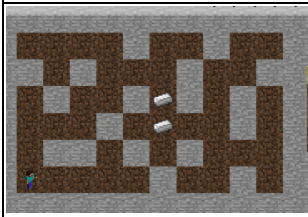




Uenig	Litt uenig	Litt	Enig	Veldig enig

3: Har du spilt Minecraft?

	JA	NEI
		

4: Har du lyst til å prøve det igjen?

	Ja	Kanskje	Nei
			
			
			
			
			

5: Det var spennende:



Uenig	Litt uenig	Litt	Enig	Veldig enig

6: Det så fint ut:



Uenig	Litt uenig	Litt	Enig	Veldig enig

7: Jeg likte at det så ut som Minecraft:



Uenig	Litt uenig	Litt	Enig	Veldig enig

8: Jeg følte at jeg var et annet sted



Uenig	Litt uenig	Litt	Enig	Veldig enig

9: Jeg glemte alt rundt meg



Uenig	Litt uenig	Litt	Enig	Veldig enig

10: Jeg følte meg flink:



Uenig	Litt uenig	Litt	Enig	Veldig enig

11: Jeg synes vanskeligheten i spillet var..:



For vanskelig	Vanskelig	Akkurat passe	Lett	For lett

12: Spillet var..:



Veldig kjedelig	Kjedelig	Litt gøy	Gøy	Veldig gøy

A Appendix 3 - Game Event Logging

The logging used in this thesis extended an already existing logging system in the game. Below follows an example of how a wrong or correct click was saved:

```

if(hadGem && !(errors == 0 && gemsFound == gemCount)){
    stats.RegisterClick(miniGameID,2,0,"right "+Input.mousePosition);
    AwardPoints(PlayerPrefs.GetInt ("Minigame2Difficulty", 1));
}
else if(!hadGem){
    stats.RegisterClick(miniGameID,2,1,"wrong"+Input.mousePosition);
    AwardPoints(-20);
}

```

Figure 26: Example code from Gem game. Information about wrong and correct clicks is saved. If the player found a gem, the stats saved it as a correct click. If the player did not find a gem in the chest, it is logged as a wrong click. Yellow marks the added code.

```

public void RegisterClick(int miniGameID, int new_eventID, int new_eventsSubID, string new_EventData){
    gameEvents.Add(new GameEvent(PlayerPrefs.GetInt ("Minigame1Difficulty", 1),
    PlayerPrefs.GetString ("UserCode", ""),miniGameID, new_eventID, new_eventsSubID, new_EventData));
    SaveEvents();
}

private void SaveEvents(){
    string savepath = GetSavePath();

    XmlSerializer serializer = new XmlSerializer (typeof(System.Collections.Generic.List<GameEvent>));
    FileStream stream = new FileStream (Application.persistentDataPath + "/" + eventfilename, FileMode.Create);

    serializer.Serialize (stream, gameEvents);
    stream.Close ();
}

private string eventfilename = "EventData-" +
    System.DateTime.Now.Month+"-"+
    System.DateTime.Now.Day+"-"+
    System.DateTime.Now.Hour+
    System.DateTime.Now.Minute+
    ".xml";

```

```
<GameEvent>
  <time>2014-10-18T14:10:09.4013395+02:00</time>
  <difficulty>1</difficulty>
  <userID>participant17</userID>
  <miniGameID>2</miniGameID>
  <EventID>2</EventID>
  <EventSubID>1</EventSubID>
  <EventData>wrong(389.0, 470.0, 0.0)</EventData>
</GameEvent>
```

Figure 27: Example of saved XML structure on click event.

A Appendix 4 - Pearson Correlations - Time and Level

Correlations

		Time1Arrow	MaxLv1Arrow
Time1Arrow	Pearson Correlation	1	,682**
	Sig. (2-tailed)		,001
	N	20	20
MaxLv1Arrow	Pearson Correlation	,682**	1
	Sig. (2-tailed)	,001	
	N	20	20

** . Correlation is significant at the 0.01 level (2-tailed).

Correlations

		Time2Gems	MaxLv1Gems
Time2Gems	Pearson Correlation	1	,882**
	Sig. (2-tailed)		,000
	N	20	20
MaxLv1Gems	Pearson Correlation	,882**	1
	Sig. (2-tailed)	,000	
	N	20	20

** . Correlation is significant at the 0.01 level (2-tailed).

Correlations

		Time3Labyrinth	MaxLv1Labyrinth
Time3Labyrinth	Pearson Correlation	1	,915**
	Sig. (2-tailed)		,000
	N	20	20
MaxLv1Labyrinth	Pearson Correlation	,915**	1
	Sig. (2-tailed)	,000	
	N	20	20

** . Correlation is significant at the 0.01 level (2-tailed).

Correlations

		Time4Zombies	MaxLvlZombie
Time4Zombies	Pearson Correlation	1	,927**
	Sig. (2-tailed)		,000
	N	20	20
MaxLvlZombie	Pearson Correlation	,927**	1
	Sig. (2-tailed)	,000	
	N	20	20

** . Correlation is significant at the 0.01 level (2-tailed).

Correlations

		Time5Animals	MaxLvlAnimal
Time5Animals	Pearson Correlation	1	,986**
	Sig. (2-tailed)		,000
	N	20	20
MaxLvlAnimal	Pearson Correlation	,986**	1
	Sig. (2-tailed)	,000	
	N	20	20

** . Correlation is significant at the 0.01 level (2-tailed).