Hildegunn Elin Olaussen Nedregård
Jane Wiik Larsen

Screening pre-reading children for dyslexia using a gamification approach

Master's thesis in Interaction design, game and learning technology
Supervisor: John Krogstie
June 2023

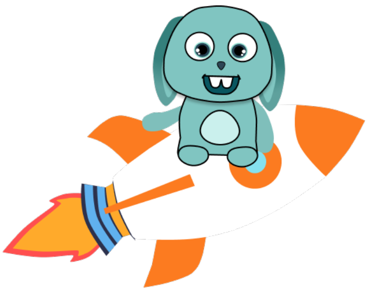
Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Hildegunn Elin Olaussen Nedregård
Jane Wiik Larsen

Screening pre-reading children for dyslexia using a gamification approach



Master's thesis in Interaction design, game and learning technology
Supervisor: John Krogstie
June 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Abstract

This thesis aims to create a version of the early dyslexia screening application, Magno, that is suitable for pre-reading children aged 5-7 years. This new version adapts the application's usability and adds gamification elements to create a motivating application suitable for the target age group. The need for effective screening tools for dyslexia in children arises from research highlighting the potential consequences untreated dyslexia has on a child's educational advancement and mental well-being. Magno is based on the magnocellular theory, which states that dyslexia stems from deficits in the magnocellular system, leading to visual processing difficulties. In this thesis, the navigation, understanding of the tasks and the motivation for the task are being examined. Through an iterative process of the design and creation research method, the application was designed, implemented, and user-tested with 18 children. The final application effectively facilitated task comprehension for all participating children and ensures smooth navigation throughout the user interface. Six out of seven children perceived the final application as enjoyable, however, two of these children struggled with concentrating throughout the task. Consequently, further testing of gamification elements should be made to enhance the sense of flow within the application and address potential attention-related challenges. Specifically, the implementation of feedback and goals holds a significant interest for this application, thus demanding further examination.

Sammendrag

Denne masteroppgaven har som mål å utvikle en tilpasset versjon av dysleksiscreeningsappen Magno egnet for barn i alderen 5-7 år. Denne oppdaterte versjonen tar hensyn til brukervennlighet og inkluderer spillkomponenter for å skape en motiverende applikasjon som passer for målgruppen. Behovet for effektive dysleksiscreeningsverktøy for barn er basert på forskning som tydeliggjør de potensielle konsekvensene ubehandlet dysleksi kan ha på barns utdanning og mentale helse. Magno er basert på den magnocellulære teorien, som antyder at dysleksi oppstår på grunn av mangler i det magnocellulære systemet og resulterer i utfordringer knyttet til visuell bearbeiding. Denne avhandlingen undersøker navigasjon, fortåelse for oppgaven og motivasjon i applikasjonen. Gjennom en iterativ prosess med design og brukertesting ble applikasjonen designet, implementert og testet med 18 barn. Den endelige applikasjonen bidro til at alle barna fikk en forståelse av testene og den sikret en enkel navigering gjennom brukergrensesnittet som alle barna klarte. Seks av syv barn oppfattet den endelige applikasjonen som underholdende, men to av disse barna opplevde vanskeligheter med å opprettholde konsentrasjonen gjennom oppgaven. Resultatene fra oppgaven viser at det er et behov for ytterligere testing av spillkomponentene for å forbedre flytopplevelsen i applikasjonen. Spesifikt er implementeringen av tilbakemeldinger og mål av betydelig interesse for denne applikasjonen og krever derfor ytterligere undersøkelse.

Preface

The work presented in this thesis is part of our master's degree in informatics at the Norwegian University of Science and Technology (NTNU). The Department of Computer Science oversees this project, which is being conducted under the supervision of Professor John Krogstie.

We want to thank Professor John Krogstie for giving us this interesting topic and for his guidance throughout this project. A special thanks to all the kindergartens who showed interest in our project and wanted to participate as well as to the children who took part. In addition, we want to thank Marcus Haaland for being the voice actor for the application and Erling Andreas Aabrekk for his helpful feedback. Thanks to both for being there as support throughout this project.

Table of Contents

List of Figures

List of Tables

1 Introduction

Dyslexia is a learning disability that can affect a person's reading and writing abilities. Characteristic features of dyslexia are difficulties in phonological awareness, verbal memory, and verbal processing speed [1]. Dyslexia is a common disorder, estimated to affect 5-10% of the population, it is the most prevalent of all learning disabilities [50]. With 60 000 Norwegian children starting elementary school each year, thousands of children will likely have to work their way through significant challenges.

Despite the widespread recognition of dyslexia, there is no national screening of dyslexia on pupils in Norway today. Only pupils with reading and writing difficulties discovered by parents or teachers will be screened for dyslexia. This results in many pupils not getting diagnosed after many years in school. In a news article from 2022 [24], a high school reported that approximately 80 of their 850 pupils have dyslexia. Out of these, 50 were diagnosed with dyslexia only after starting high school. Many of these pupils expressed feelings of inadequacy in elementary school. This finding aligns with a report conducted by Dyslexia Norway (Dysleksi Norge), which reveals that approximately 20% of Norwegians with dyslexia do not receive a diagnosis until they are between 16 and 19 years old [22]. Research conducted on this topic reports that these children often experience feelings of failure which may result in them struggling with low self-esteem [2]. Moreover, additional studies reveal more severe consequences, with evidence of dyslectic children having an increased risk of mental health issues like anxiety and behavioral issues like acting out and aggression [18].

Discovering dyslexia at an early age could help remedy these consequences. Dyslexia can be detected from 2nd grade as today's tests require a minimum knowledge of phonological awareness, grammar and vocabulary to predict the risk of dyslexia [50]. However, recent research has shown that dyslexia can also be identified through linguistic and cognitive markers, even in preschool age [19]. This means that pupils that have a high risk of developing dyslexia can be discovered even before they have begun their formal reading education. This is important as research shows a significant decrease in the effectiveness of interventions over time [42]. From the 1st to the 5th grade, the impact of interventions can diminish by up to 70 percent [27], emphasizing the need for early detection and appropriate educational measures to prevent individuals from falling behind in their education.

This thesis aims to continue the development of a dyslexia screening tool for early detection of dyslexia named Magno. Up until now, Magno has been specifically designed and tested for adult users. The objective of this thesis is to develop a version of Magno that is suitable for pre-reading children aged 5 to 7. The overarching goal is to contribute to the early detection of potential dyslexia in children before they have started their formal reading education.

1.1 Project structure

This thesis is structured into 9 chapters as follows. In Chapter 2, an overview of previous work related to the Magno application is presented, including the preparatory project for this thesis and the derived requirements from that project. Chapter 3 provides a detailed discussion of relevant theories and literature for the project, including theories of dyslexia, gamification, and usability. Chapter 4 outlines the research approach, research objectives, and research questions addressed in this thesis. Chapter 5 provides a concise overview of the technological tools utilized in the project, along with a rationale for their selection. Chapter 6 presents the iterative design process employed in this thesis. This chapter is divided into three parts, each describing the design, implementation, and results of a specific iteration. Chapter 7 is a comprehensive discussion of the project and its outcomes in relation to the presented theories. Finally, Chapter 9 concludes with the main findings, presents limitations of the project, and provides recommendations for future work on adapting Magno for pre-reading children.

2 Previous work on Magno

The work on the Magno application started in 2015 when Wold [53] developed an Android application based on a program from the early 2000s. This program, called "Form", showed great promise as a screening tool for dyslexia [8]. The goal in creating the Magno application was to modernize Form to be compatible with today's technology. Since then, the application has been worked on during a total of four master theses [11, 25, 10, 48] and two preparational projects [47, 14]. This section will present the current state of the Magno application as well as earlier works related to Magno and children. Parts of the text in this section are based on text from this thesis' preparatory project [15].

2.1 Tests

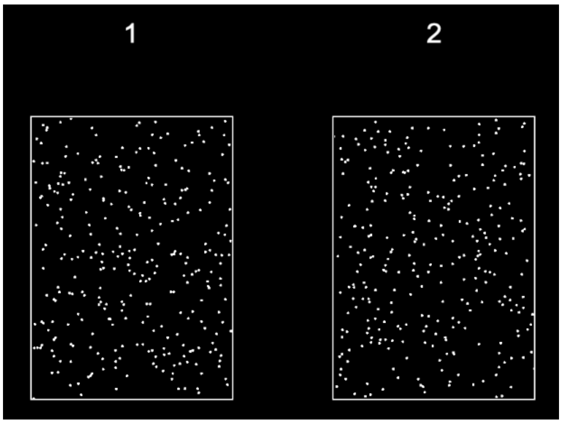
The application consists of three tests: a random-dot kinematogram test (referred to as motion test) and two static and random pattern tests (referred to as form tests). These tests evaluate visual processing abilities. There exists evidence that dyslectics have a deficit in the detection of motion stimuli, called the magnocellular theory (see section 3.1.1). This is tested by the motion test.

In a study by Hansen et al. [8] used these three tests to examine the connection between visual deficits in dyslexics and the dorsal stream function. Their results showed that dyslexics were less sensitive to the coherent motion test compared to a control group. For the form random test, there was a smaller yet significant correlation. However, when dyslexic participants were tested on the form fixed test, they did not show a significant difference in performance compared to the control group. The tests are described in greater detail below.

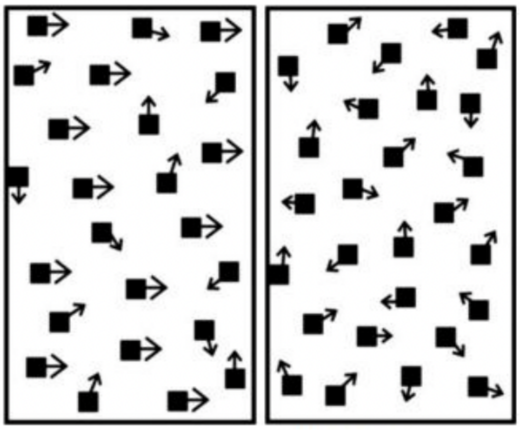
Motion test

The motion test (Figure 1a) consists of two boxes, each containing 300 dots with a size of one pixel. The user is given a time limit of five seconds to identify the box in which the dots move in a coherent manner from left to right, with their direction reversing every 0.572 seconds. In the other box, the dots change direction either after 0.572 seconds or when they collide with another dot. By default, all dots move at a speed of 50 pixels per second.

Once the five-second time limit is up, the dots disappear, and the user is prompted to select the box that contained the coherent moving dots. If the user selects the correct box, the test becomes more challenging in the subsequent round as the level of coherence decreases. However, if the answer is incorrect, the coherency increases, making the test easier.



(a) Implemented motion test

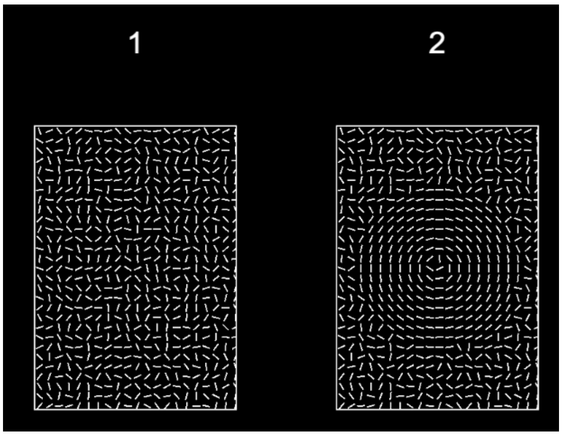


(b) Movement pattern of the dots

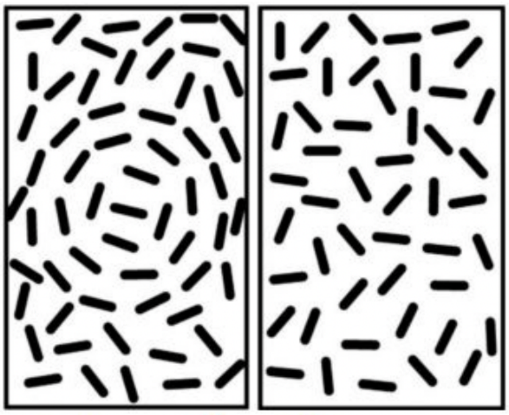
Figure 1: Motion test

Form tests

The form tests (Figure 2) consist of a total of 600 random lines evenly distributed throughout the box, each with a thickness of 1 pixel and a length of 0.4 degrees. The objective is for the user to identify the box that contains the line segments forming circles. If the user selects the correct box, the percentage of line segments forming circles will decrease in the subsequent test, making it more challenging. If the user chooses incorrectly, the percentage will increase, resulting in an easier test. There are two types of form tests: static global and random target. In the static global test, the placement where the line segments form a circle remains constant. In the random target test, the circle is randomly placed within the box.



(a) Implemented form test

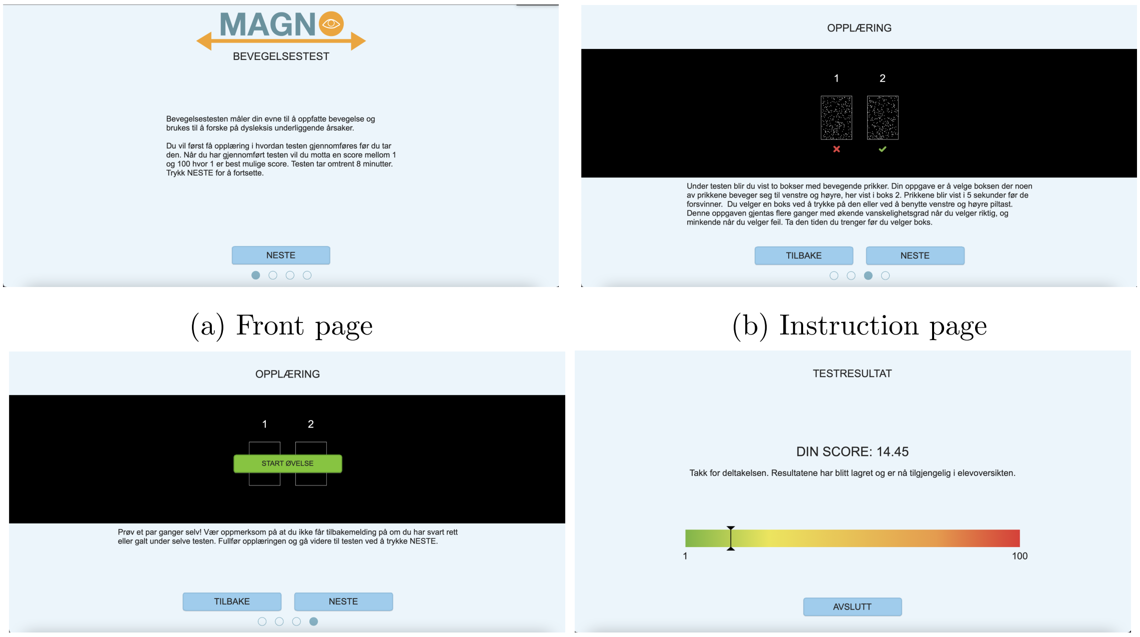


(b) Left: circle. Right: random pattern

Figure 2: Form test

2.2 User interface

Johansen and Kirkerød [11] worked on evaluating and improving the user experience of the interface among adult participants. The System Usability Scale (SUS) score yielded a value of 92.7, indicating a high level of usability. Furthermore, the application incorporates specific modifications intended to cater to individuals with reading disorders. Parts of the interface of the user tests are shown in Figure 3.



(a) Front page

(b) Instruction page

(c) Trial page

(d) Results

Figure 3: Magno user interface

Additionally, Syvertsen [48] created a database with a UI that researchers can use to analyze the results of the tests. The result of this master thesis was a platform with a fully functional graphical user interface, a server that handles API calls, and a database storage system. The platform also integrates the existing tests and allows teachers to add pupils to the platform.

2.3 Magno tested with children

A few research projects have utilized these tests for research related to children, one of these being a project by Sigmundsson et al. [36]. The goal of this project was to explore the connection between math skills and deficits in visual processing. In this project, the parameters of the tests themselves were adjusted for the user group of children aged 10 years old. One of the parameters was the coherency percentage which was defined as ”the ratio of stimulus elements in the target signal to the remaining noise elements" [36]. While this coherency percentage when previously tested on adults has been set to have a starting point of 66.7% [8] and 50% [10], Sigmundsson here used a setting with 75%. The reason for this high percentage was to have a starting value where the pattern to be identified was deemed to be easily detected for all participants. In the motion test, 75% is the maximum obtainable coherency percentage when accounting for a dot lifetime of 4 seconds. Sigmundsson has also used this threshold in another article with children in the same age range [37]. In this study, the increase and decrease in stimuli were set to 1 dB from its previous value for each incorrect and 3 dB for each correct value.

Besides technical specifics, the children were instructed by one of the test leaders about how to complete the tests and had an instructor present throughout the test. Furthermore, they were allowed to complete practice trials where all the tests were fixed to 75% coherence for all tests. The final results of the project where visual processing was compared to mathematics skills showed that there was a definite connection between the children with visual deficits and their math skills.

2.4 Recommendations for adapting Magno to children

Klevstuen [14] explored how the Magno application could be better suited for younger children. This investigation involved interviewing experts in the fields of obstetrics and special education. One of the key findings was how the experts meant that children under the age of 10-12 could have difficulties maintaining focus throughout the test. In addition, younger children may not have developed the skills to recognize patterns and systems effectively, making it harder for them to complete tests within the given time limit of 5 seconds. Additionally, as the difficulty of the tests increases, there is a risk that children might resort to selecting answers at random or become discouraged and give up entirely.

In a separate study by Syvertesen [47], one teacher and one special educator were interviewed to gain insights into adapting the application for testing on children. The findings of this project align with Klevstuen's research, emphasizing the importance of close supervision by a supervisor during the test. This allows the supervisor to make necessary adjustments, such as providing short breaks if needed. Evaluating the tests requires awareness of factors that could impact the results, such as the child's understanding of the test and instructions, as well as the effort they make.

2.5 Preparatory project

The preparatory project for this thesis [15] investigated how cognitive assessments could be adapted to children aged 5 to 7 by looking at gamification elements as well as usability requirements for this age group. The project drew on earlier research on cognitive assessments for children and insights from previous research conducted by Klevstuen and Syvertsen.

Findings show that when designing Magno for the target age group, several considerations must be made in terms of usability. Younger children tend to have shorter attention spans. The application should incorporate vivid colors and interaction into the application to grab and retain their attention. Additionally, since the target age group lacks reading skills, the tests should be presented using illustrations and audio explanations and potentially incorporate a storyline to aid comprehension. To remove test anxiety, findings showed that it is recommended to create a less formal testing environment. Caution should be taken to show test results or negative feedback to children, as it could stress them and harm their self-esteem. Possibly causing them to abandon the test. Adapting the tests to match the child's cognitive skill level and allowing more time for processing choices is also suggested. Children are more accustomed to touch input so the Magno application should be adapted for tablet use, with appropriately scaled buttons and tests to accommodate children's limited motor and cognitive skills.

In addition to usability considerations, the project explored the integration of gamification elements into the Magno application to enhance engagement. Ensuring individuals' focus and optimal performance is crucial in cognitive assessments [52]. However, young children may struggle to grasp the significance of this and find it insufficiently motivating to give their best effort [51]. Research indicates that incorporating gamification into cognitive assessments has the potential to enhance motivation and engagement among participants [13].

While the impact of gamification on cognitive assessments for children is still debated [52, 17], potential elements that could positively influence motivation using the self-determination theory (see section 3.3.2) were identified: customization, feedback, story and theme. Customization allows children to have a sense of influence over the game (autonomy), feedback should be positive to boost self-esteem (competence), and the inclusion of a story or theme can create a more immersive experience (relatedness).

2.5.1 System requirements

Based on the insights gained from relevant research as well as earlier work on Magno, the preparatory project resulted in a list of requirements which are presented in Table 1. Functional requirements (FR) describe the features and functions of the product and non-functional requirements (NFR) describe the general function of the system. NFR1 was included to ensure maintainability as well as easy and fast coding (see section 5). NFR2 focuses on optimizing the code.

ID	Description	priority
FR1	The test should give positive feedback on progress	high
FR2	The instructions should be given by the use of audio and illustrations	high
FR3	The instructions can be repeated	high
FR4	The application should have a continuous story	low
FR5	The tests time constraints should be increased to 10 seconds	high
FR6	The test should have sound feedback for all actions	high
FR7	The test results should not be visible to the child	high
FR8	The test should be possible to perform on a tablet	high
FR9	The test should be scaled	high
FR10	The test should have a vivid color palette	high
FR11	The instructions can be skipped at any time	medium
FR12	The test should include background music with positive associations	medium
FR13	The test should instruct the child to choose the correct box when the time limit is up without text	medium
FR14	The test should have engaging animations	medium
FR15	The test should include a personalizing option	medium
FR16	The test should include a personalizing option	medium
NFR1	The code should be fast and easy to change	high
NFR2	The application must have a maximum latency of 2 seconds for video, animation and image display	high

Table 1: Derived requirements for adapting Magno to children aged 5-7 years

3 Background

This section presents the theoretical framework that serves as the foundation for this thesis. It outlines the key concepts, models, and theories that are used in analyzing and interpreting the data. Section 3.1 are based on text from this thesis' preparatory project [15].

3.1 Theories of dyslexia

Researchers today agree that dyslexia is a lifelong neurological disorder with a genetic origin. However, even though it is one of the most researched psychological special disorders, the underlying biological and cognitive causes are still being debated [28].

Multiple studies have demonstrated that dyslexic individuals have impaired phonological awareness and that measures of phonological skill have predictive value for reading ability [9]. This is referred to as the phonological theory. One critique of this theory is its inability to provide a comprehensive explanation for why some children struggle to develop literacy skills [44]. As a result, it becomes challenging to differentiate dyslexic poor readers from other potential factors that contribute to children's difficulties in learning to read, such as ineffective teaching methods, inadequate family support, or irregular school attendance. To address this limitation, alternative theories aim to elucidate the underlying biological causes of reading difficulties. There are many theories on low-level abnormalities that could lead to impairments in phonological awareness such as basic auditory processing, magnocellular, cerebellar, attention-shifting, or general sensorimotor deficits. The theory of deficits in the magnocellular system is the most relevant to this study given the fact that the assessment tests in the Magno application are based upon this theory.

3.1.1 The Magnocellular Theory

The magnocellular theory was formed to link the many different theories about the causes of dyslexia on a biological level [43]. It suggests that deficits in the visual processing systems may be linked to dyslexia. More specifically within the dorsal pathway of the brain, where specialized magnocellular neurons detect visual motion.

Today, several studies have linked deficits in the visual magnocellular function with dyslexia, putting further weight behind the magnocellular theory [44]. Evidence has been found not only with psychophysical tests of magnocellular function, but also studies with electrophysiology, eye movement, attention, imaging, interventional, and genetic findings have confirmed a reduction in visual motion sensitivity at each level of the visual system [45].

Skottun [39] has criticized the magnocellular theory, higlighting the fact that most of the research supporting this theory comes from studies looking at the sensitivity to contrast. Here, he criticize its support, claiming that it is ambiguous. Only 4 of the 22 studies Skottun investigated are in line with the theory. And this, he argues, is evidence that the theory is too uncertain. This critique is answered by Stein in a recent paper [45]. He refers to Skottun and Skoyles [40] saying that visual magnocellular weakness cannot be detected in every dyslectic. Some children with mildly impaired magnocellular function may still learn to read. Thus this cannot be the sole cause of dyslexia [41]. In another article, Stein also explains that dyslexia may also involve difficulties in auditory processing and other neural temporal processing [44]. Thus, he emphasizes the relevance of considering these additional factors in order to gain a comprehensive understanding of the dyslexia disorder.

3.2 Usability

Usability is defined as the extent to which a product can be used by specified users to achieve established goals, with effectiveness, efficiency, and satisfaction in a defined context of use [21].

Children have unique cognitive, perceptual, and motor abilities that should be considered when designing cognitive assessment applications. By considering the cognitive capacity of different age groups, designers can make applications usable and appealing to younger audiences.

3.2.1 Piaget's Theory of Cognitive Development

When designing cognitive assessment tools, it is essential to consider the cognitive capacities of the target age group [16]. Tasks should be carefully crafted to allow children to integrate new information based on their existing knowledge and adapt their understanding when faced with novel or challenging tasks. By aligning assessments with children's cognitive abilities, the tools become more meaningful, engaging, and developmentally appropriate.

This section introduces Piaget's theory of cognitive development [33], which provides valuable insights into the cognitive growth and understanding of children. Piaget's theory emphasizes the active role of children in constructing knowledge and understanding through interactions with their environment. The theory consists of the following key components:

Stages of development
Piaget has identified four main stages of cognitive development that children progress through. These are the sensorimotor stage (0-2 years), preoperational stage (2 to 7 years), concrete operational stage (7 to 11 years) and formal operational stage (11 years and older). The preoperational stage is the most relevant to this thesis. According to Piaget, the preoperational stage is characterized by the development of symbolic thinking and the emergence of language skills in children. However, their thinking is primarily egocentric, meaning they aren't yet able to effectively take other people's perspectives.

Assimilation and Accommodation
Piaget proposed that cognitive development involves two complementary processes: assimilation and accommodation. Assimilation refers to the incorporation of new experiences or information into existing mental frameworks (schemas). Accommodation involves modifying existing schemas or creating new ones to accommodate new information or experiences that do not fit existing mental structures. Piaget emphasized the importance of achieving a balance, or equilibrium, between assimilation and accommodation. Equilibration occurs when children are able to resolve cognitive conflicts and achieve a harmonious understanding of the world. This process promotes intellectual growth and development.

3.3 Motivation in cognitive assessments

Motivation is the driving force behind all human behavior. Motivation in cognitive assessments refers to the internal or external factors that influence an individual's level of engagement, effort, and persistence during the assessment process, ultimately impacting their performance and results and thereby the validity of the assessment [4]. It includes the motivational factors that underlie an individual's behaviors, choices, and attitudes toward the assessment tasks. Understanding the multifaceted nature of motivation is crucial for comprehending its impact on cognitive assessment outcomes [32].

3.3.1 Intrinsic and extrinsic motivation

According to Ryan and Deci [31], intrinsic motivation involves engaging in an activity for the inherent satisfaction it brings, rather than for external rewards or consequences. When intrinsically motivated, individuals are driven by the enjoyment and challenge of the activity itself, rather than external factors like rewards or pressures. In contrast, extrinsic motivation arises when an activity is pursued to achieve a separate outcome. It is often employed in situations that require short-term engagement, where the focus is on obtaining external rewards rather than finding inherent interest in the activity.

In short, intrinsic motivation stems from internal factors, while extrinsic motivation arises from external sources. However, these two types of motivation can coexist and influence our behavior. An example is someone working on a project may be extrinsically motivated to meet a deadline, while also being intrinsically motivated by their enjoyment of the project and desire to produce excellent results. Consequently, our motivations often involve a combination of intrinsic and extrinsic elements [34].

3.3.2 Self-determination theory

Self-Determination Theory (SDT) [32] is a theory closely related to motivation that differentiates between intrinsic and extrinsic motivation. The term self-determination refers to a person's own ability to manage themselves, to make confident choices, and to think on their own. Self-determination theory focuses primarily on intrinsic motivation.

Self-determination theory posits that individuals are driven by three fundamental and universal psychological needs: competence, relatedness, and autonomy. According to this theory, personal well-being is directly influenced by the extent to which these basic psychological needs are satisfied.

Competence: is when someone has sufficient intellect, judgment, skill and/ or strength to perform a given task. A person that is competent feels a sense of mastery over their environment. Feelings of competence can increase when a person's skills match the demands of a task or positive feedback is received. If the demands are too high or a person receives negative feedback, feelings of competence can decrease.

Relatedness: involves feelings of closeness and belonging to a social group. Connections are important to gain access to help and support. Feelings of relatedness can increase when a person is part of an inclusive environment where they are respected and cared for. Competition with others, cliques and criticism from others can decrease feelings of relatedness.

Autonomy: is being able to make your own decisions. Feelings of autonomy are increased when a person is given the choice to control their own behavior and when other people acknowledge their feelings. If a person is controlled by others or has to operate according to deadlines, feelings of autonomy can decrease.

3.4 Gamification

A popular definition of gamification was created by Detering who defined it as "the use of game design elements in non-game contexts" [5]. It involves incorporating features such as challenges, rewards, leaderboards, interactive elements, and narratives to transform a task into an interactive and enjoyable experiences [16].

3.4.1 The Gameflow model

The Gameflow model [46] provides a holistic perspective on the factors that contribute to an engaging and enjoyable gaming experience. Though the Gamflow model is initially meant for evaluating and enhancing games, its insights are also valuable in the design and evaluation of gamified cognitive assessments. The model's emphasis on engagement and motivation aligns with the goals in gamified cognitive assessments, making it an effective framework for enhancing the assessment experience.

By considering the key components of the Gameflow model, researchers and practitioners can create assessments that are enjoyable, and engaging, and promote optimal performance outcomes. This section will describe the key components of the Gameflow model: concentration, challenge-skill balance, clear goals, feedback, control and immersion.

Concentration: The capacity to fully focus and immerse oneself in a game is referred to as concentration. The more focus required by a task in terms of attention and workload, the more immersing it will be. To achieve a state of concentration, a lot of relevant stimuli should be offered to attract and hold the user's attention. It is important that the workload is appropriate to the user's perceptual, cognitive, and memory limits. Players should not be diverted from tasks on which they need to concentrate.

Challenge and player skills: According to Gameflow, challenge is the most important aspect of good game design. A challenge that exceeds the user's skill-level causes feelings of anxiety and can result in the user giving up. Challenges that are less than the user's skill-level can result in apathy or boredom. Challenges therefore must match the users skill-level, provide different challenges for different users and provide new challenges at an appropriate pace. The game should be learned through interesting and absorbing tutorials to get involved quickly and easily. The skills needed for the game should be learned as a part of the game. When the level of challenge is appropriately matched to the player's abilities, it promotes a sense of flow and optimal engagement.

Control: To experience flow, the player must experience some degree of control over their actions. It involves providing meaningful choices, allowing players to make decisions, and influencing the game's outcome. The game should avoid providing decisions that have no impact or no importance. Offering a sense of control enhances player engagement and satisfaction.

Clear goals: Clear goals provide players with a sense of purpose and direction. The game should have an overriding goal presented early in the game, in addition to several intermediate goals in each level. Well-defined objectives and clear instructions guide the player's actions and progress within the game, enhancing motivation and engagement.

Feedback: Feedback plays a crucial role in the Gameflow model. It involves always letting players get feedback on progress towards their goals, get immediate feedback on their actions and know their status or score. Effective feedback helps players understand their strengths and weaknesses, supports their learning, and fosters a sense of competence and mastery.

Immersion: Immersion describes the extent to which the player feels deeply engaged and absorbed in the game world. Elements such as realistic graphics, captivating narratives, and immersive audio contribute to the sense of immersion and enhance the overall gaming experience.

4 Research methodology

4.1 Research goal

Based on the previous work on the Magno application [10, 11, 14, 25, 47, 48, 53] as well as our preparatory project [15], this research project aimed to develop and test the usability of a dyslexia screening application that could be used to indicate the risk of dyslexia in children at the beginning of their formal reading and writing education. For this reason, the target group for this project was Norwegian children aged 5-7 years old.

It is important to emphasize that this thesis did not aim to test the validity of the screening tests. The purpose of this project was to adapt the Magno application in such a way that the target group would be able to complete the tests, with a focus on usability and intrinsic motivation. The final results from this new version of Magno would then lay the foundation for future work on the application.

This thesis was based on the premise that the screening would be performed in a situation where the child did not have one-to-one supervision with an adult. Rather, the screening would be conducted at a group level, with each child taking the test independently while being supervised by one adult. This approach was chosen to ensure a simple and cost-effective facilitating process.

4.2 Research questions

The following research questions was addressed in this thesis:

RQ1: *How can Magno be adapted to be more user-friendly for Norwegian children aged 5-7 years old?*

> **RQ1.1:***Do the children understand how to complete the tasks based on the instructions given to them?*
>
> **RQ1.2:***How do children navigate the Magno application?*

RQ2: *How motivating is the Magno application for children aged 5-7 years old?*

> **RQ2.1:** *How does the inclusion of gamification elements affect the motivation for the tasks?*

4.3 Research methodology

4.3.1 Research design

The research in this paper was based on the design and creation research method proposed by Oates [23]. An overview of Oates' model was shown in Figure 4. Design and creation consisted of three phases: problem identification, solution generation, and evaluation. An iterative approach was used for the two final phases.

The first phase, problem identification, was based on the narrative literature review done in this thesis preparatory project [15] as well as the findings from previous work on Magno [14]. From this, a list of requirements was deduced that served as the starting point for this project's new prototype of the Magno application. The second phase, solution generation, involved the design and development of a new version of the dyslexia screening application. The final phase, evaluation, entailed testing the usability and effectiveness of the dyslexia screening application through usability testing with children aged 5 and 6 years. This is further described in section 4.3.3. The findings gained from this usability testing were used in a new iteration with solution generation and evaluation. In this way, the design and creation research method provided a systematic and iterative approach to developing and testing the dyslexia screening application.

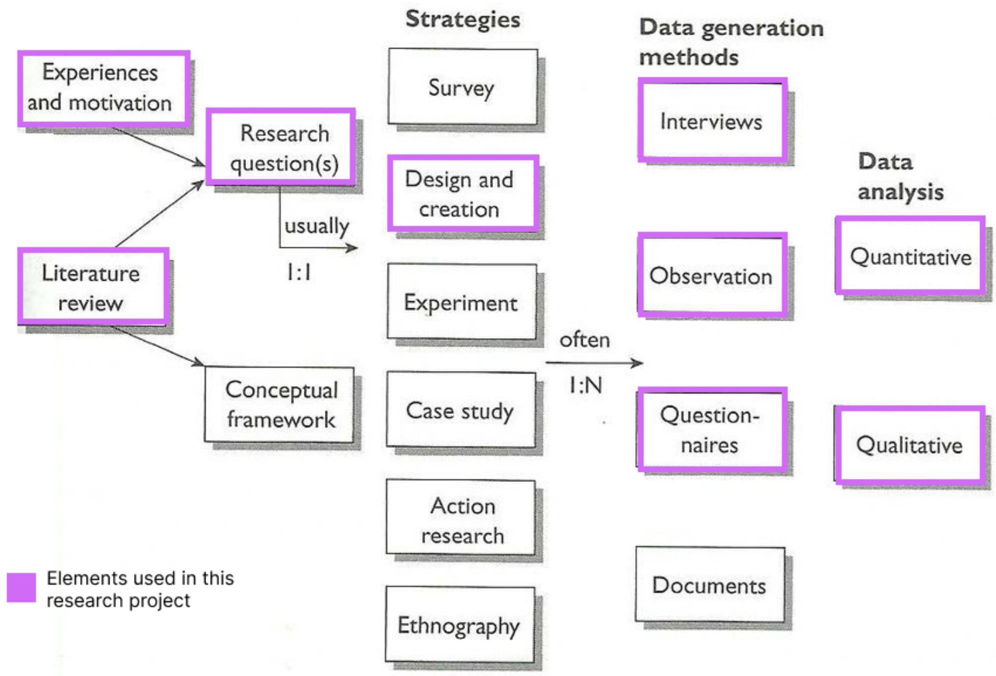


Figure 4: Oates model on research processes

4.3.2 Sampling

The sampling strategy employed for this study utilized the self-selection sampling technique [23]. The population of this study was all Norwegian children aged 5-7 years old. From this, the sampling frame consisted of preschool children of all genders, aged 5-6 years. Participants were recruited from all kindergartens in Trondheim. This was done to ensure that our sampling frame encompassed preschool children of a diverse set of backgrounds.

For each iteration, a minimum of 5 children was required for the sample size. Usability testing research suggests that 5 participants are sufficient to identify the main issues with an application [20]. To ensure adequate data for valid test results in case of illness or unwillingness to participate, efforts were made to recruit more than five children. The final sample size was set at a minimum of 18 children to ensure unfamiliarity with the application across the three iterations.

From the initial pool, 12 kindergartens expressed interest in participating. The selection process prioritized the four kindergartens with the highest number of children in the target age group. This approach aimed to maximize the likelihood of recruiting a sufficient number of participants. The selected kindergartens were provided with an approval scheme to be forwarded to the parents, as explained in subsection 4.3.6. The application was tested on those children who obtained parental approval and expressed their willingness to participate. To mitigate nonresponse bias from the parents, the kindergarten teachers were responsible for reminding them about the approval scheme.

It is important to note that the sample can have some self-selection bias as only kindergartens who volunteered were participating in the study. In addition, the application was not tested on children aged 7 years. The reasoning behind this was that if 5 and 6-year-olds managed to use the application, it would be likely that it would be usable for 7-year-olds as well, though this needs to be validated. This may have affected the generalizability of the results for the whole population.

4.3.3 Data collection methods

This study aimed to assess the comprehension, ease of use, and engagement of pre-reading children when they utilized the Magno application. The data collection methods employed for this purpose involved conducting user tests, during which data was gathered through observations and user feedback. Throughout the tests, the presence of a facilitator ensured oversight of the child's interactions, while an observer took notes on the child's behavior. Verbal feedback was also obtained and recorded using voice recordings.

Observation

During the participant's interaction with the application, careful observation was conducted to identify any instances of confusion or hesitation. Emphasis was placed on the research questions related to how participants engaged with and navigated the application. The children's behavior, including their enthusiasm, persistence, and interest, was observed to measure their level of motivation, engagement, or disengagement throughout the tests.

Additionally, particular attention was given to assessing the participants' success in completing the assigned tests. Difficulties encountered by users during specific tests might have suggested a potential lack of comprehension regarding the application's usage.

User feedback

After each test, participants were requested to share their feedback regarding their understanding, difficulties, and enjoyment throughout their usage of the application. To accommodate the cognitive development of children, simple questions were posed to elicit their feelings during the test, and the children answered by using a smileyometer to express their emotions.

Smileyometers were a way of measuring satisfaction by showing a scale of smiley faces ranging from negative to neutral to positive. This form of measuring user satisfaction had been shown to work well on small children in the context of UX design [30]. However, research showed that it was important that the smileyometer represented the child's feelings and what they wanted to communicate [7]. For this reason, overly sad or angry smiley faces were not used. The smileys used in the tests were shown in Figure 5. When analyzing the data a 4 rating indicated the maximum satisfaction score, while a 1 rating represented the minimum satisfaction score.

After collecting quantitative data through the smileyometer, the children were asked some follow-up questions to try to understand why they chose a particular smiley. These questions aimed to gain further insights into the children's thoughts on the application and learn more about what they liked or disliked about it.

Figure 5: Smileyometer used in user tests

4.3.4 Special considerations for user testing on children

Because the testing was with young children, special considerations were taken. This section will explain the measures that were taken to account for children's special needs.

To encourage children to think aloud, an approach was taken to tell them that they were experts and that the goal was to learn their thoughts and perspectives on the application. Positive affirmations such as *"That is some really good feedback"* or *"You did a great job testing this application"* were used to instill confidence and create a safe environment for them to share their thoughts.

Traditionally, children tend to feel apprehensive about undergoing various types of assessments. As highlighted in the literature review by Lumsden et al. [16], research suggested that test-related anxiety could negatively impact a child's performance in cognitive assessments. To mitigate performance anxiety and create optimal testing conditions, warm-up questions were incorporated to engage the child in casual conversation and make them feel more comfortable. If they wanted to, they could also bring an adult they knew to feel safer. One of the goals of the application was to enable children to independently navigate the tests. Minimal interference was desired, and for this reason, the assessments took place in a quiet room separate from other children and adults in the kindergarten. This environment minimized distractions and helped the child focus on the task at hand.

4.3.5 Data analysis techniques

Thematic analysis

The data collected from observations and follow-up questions underwent thematic analysis, which combined qualitative insights with quantitative data collection methods. Thematic analysis is a qualitative research technique that involves identifying patterns or themes within the data [3]. The analysis process consisted of several steps. Initially, the interview notes and transcripts were thoroughly read to gain a comprehensive understanding. Next, individual observations and quotations were tagged with appropriate codes, and an affinity-diagramming technique was utilized to group related data and reveal significant themes. In this study, a deductive approach was employed, where codes were developed as the data was reviewed. This method aligned well with the investigation of new ideas or concepts. The analysis commenced with initial coding to gain a broad overview of the data, followed by line-by-line coding. The selection of this approach, instead of software and journaling, was based on its suitability for collaborative work and its simplicity.

Frequency distributions
The responses to the smileyometer were analyzed by calculating the frequency distribution for each available response option. This analysis provided insight into the number of participants who chose each response option. By examining these results, patterns or trends regarding the perceived usability of the application could be identified.

4.3.6 Ethical considerations

This study was conducted in an ethical and responsible manner, ensuring that the rights and well-being of research participants were protected. This included obtaining informed consent, maintaining confidentiality, and ensuring participant anonymity. Before recruiting participants, the project was approved by NSD [38].

The participants in the tests were minors. For this reason, written consent from the child's legal guardian was obtained before conducting the testing. The guardians received information about the purpose, procedure, and data that would be collected in a consent scheme, which they returned to the kindergarten if they wanted their child to participate.

The privacy and confidentiality of the research participants were protected. Any personal information collected during the study was only accessed by the researchers. All data was stored securely and anonymously. Identifying information was removed from the data before analysis.

5 Technological Tools

This section contains an overview and description of frameworks, tools and technologies that have been used in this project in addition to changes from previous versions of Magno and why these changes were made.

5.1 Tests with PixiJS

PixiJS is a rendering engine designed to render complex 2D graphical content designed for web applications. It is high-performance and supports a wide range of features for creating graphic elements and interaction. The framework is built on top of WebGL, providing a solid foundation for graphics rendering, user input management, and animation creation [26].

In this project, PixiJS was used to display the tests and the test trials. PixiJS leverages WebGL and can display thousands of moving sprites efficiently. This was relevant as the tests in the Magno application consisted of hundreds of moving elements. The logic is the same as the one implemented in earlier versions of the Magno application.

5.2 UI with React and Typescript

React is a front-end library that is built on top of JavaScript and is designed to simplify the process of building dynamic and responsive user interfaces [29]. React uses a declarative programming model, which simplifies the process of building complex interfaces and makes it both more efficient and flexible. In addition, React is also highly modular and extensible and allows for breaking down the interfaces into smaller, reusable components.

Typescript is an extension of Javascript, compatible with React, that enhances the developer experience by letting developers add type safety to projects [49]. Typescript supports a tighter integration with the code editor that catches errors earlier. This makes the project easier to maintain and collaborate with others as it provides reference validation and other language features.

As mentioned above, the original interface was written entirely in PixiJS. PixiJS is powerful, which is needed to render the animations in the tests. However, in terms of development flexibility, PixiJS may have limitations, particularly when it comes to making rapid and effortless changes to the UI as it is a low-level API. For this reason, it was decided to change the UI to React as it is a popular and well-documented framework. Using both PixiJS and React with Typescript, it was possible to leverage the strength of both frameworks to build a powerful and flexible application.

5.3 Design with Figma

Figma is an all-in-one design platform where you can easily create digital designs. It is cloud-based and supports real-time collaborations with teammates. With Figma, users can create wireframes, prototypes, and high-fidelity designs for websites, apps, and other digital products [6].

With a strong emphasis on design and multiple iterations, Figma's collaborative capabilities provided an ideal environment for efficient collaboration and decision-making. Figma also is also capable of creating vector graphics and downloading them as SVGs. As one of the key elements in a child-friendly application is illustrations, this was a valuable and easy way of creating custom designs and illustrations for the project.

5.4 Instructions videos with Keynote

Keynote is a presentation software developed by Apple Inc. With keynote, users can include text, images and charts and multimedia elements like audio and video. Keynote also offers animation and transition effects that can be applied to individual elements to create a dynamic and engaging experience [12].

In this project, effectively explaining the tests to the users was an important issue, as highlighted in section 6.1.1. To achieve this, creating videos with illustrations and voiceovers was considered to be essential. Keynote, with its comprehensive set of tools, offers easy ways of animating elements, incorporating voiceovers, and exporting them as videos that could be seamlessly integrated into the application.

6 Results

6.1 First iteration

In this iteration, the primary objective was to create a new version of the Magno application that where usable for pre-reading children. The focus was to implement the necessary changes to ensure easy navigation and understanding of the application. By focusing on only the necessary changes to make the application easy to navigate and understand, the goal was to create a base point on which later iterations could be measured. In this section, the design changes, implementation details, user testing, and evaluation of the first iteration are described.

6.1.1 Design

The changes of this iteration were based upon the findings in the preparatory project [15] related to usability in web applications for children. This section outlines the requirements for the first iteration and introduces the new solutions that were implemented based on these requirements. Screenshots of the new interface are presented in appendix A.

Requirements
The requirements for the first iteration were based on selected relevant requirements from the preparatory project (see Table 1). The full list of changes in this iteration is presented in Table 2 shown below.

ID	Description
FR1	The test should give positive feedback
FR2	The instructions should be given by the use of audio and illustrations
FR3	The instructions can be repeated
FR5	The test's time constraints should be increased to 10 seconds
FR6	The application should have sound feedback for all actions
FR7	The test results should not be visible to the child
FR8	The test should be possible to perform on a tablet
FR9	The test should be scaled to a larger size
FR10	The application should have a vivid color palette
FR11	The test instructions can be skipped at any time
FR13	The test should instruct the child to choose the correct box when the time limit is up without the use of text
NFR1	The code should be fast and easy to change

Table 2: Functional requirements for iteration 1

Pre-reading children don't understand written instructions

In the previous version, Magno relied mainly on text instructions. However, since the application was designed for pre-reading children, this was identified as a crucial issue during the first iteration. To address this concern, it was decided that instructions should be presented through videos featuring illustrations and voiceover (FR2). These videos effectively communicated the test concepts using simplified animations that highlighted movements or patterns. Concrete examples were also provided to demonstrate both correct and incorrect answers. After watching the video, children were given the opportunity to independently interact with the trial screen, allowing them to practice and become familiar with the tests before moving on to the actual test.

Children have shorter attention span

As mentioned in section 2.5, it was recognized that engaging elements are necessary to effectively capture children's attention. To enhance their focus, several modifications were made to the application. Firstly, all colors were adjusted to more vibrant hues (FR10) to enhance visual appeal. Additionally, feedback sounds were incorporated for all interactions (FR6). Specifically, a distinct "ping" sound played when an interactive element was touched. Moreover, when the time limit for the tests expired, additional instructions were implemented (FR13), including a hand gesture pointing towards the boxes and reminding the child to select their preferred choice.

Children need more time to process information than adults
Section 2.4 emphasized the notion that younger children might not have fully developed the cognitive abilities necessary to effectively discern patterns and systems. Taking this aspect into consideration, they were allowed to attempt the test multiple times on the trial page (FR3) until they felt sufficiently familiar with the test. Moreover, they were given the opportunity to watch the instructional video multiple times. Additionally, the time limit for choosing a box was extended to 10 seconds (FR5).

Children are more familiar with touch
The preferences of young children indicate a greater comfort level when interacting with touch interfaces as opposed to trackpads or mice (FR8). For this reason, the interface was specifically designed for tablet devices by incorporating touch listeners and larger buttons.

Children might experience test anxiety
As indicated in section 2.5, it was important to implement measures aimed at preventing children from experiencing test anxiety. The removal of test results (FR7) and its substitution with a confetti animation served to provide positive reinforcement to the child (FR1). Moreover, the goal was to foster the child's motivation through the exclusive use of positive feedback sounds and encouraging voiceovers.

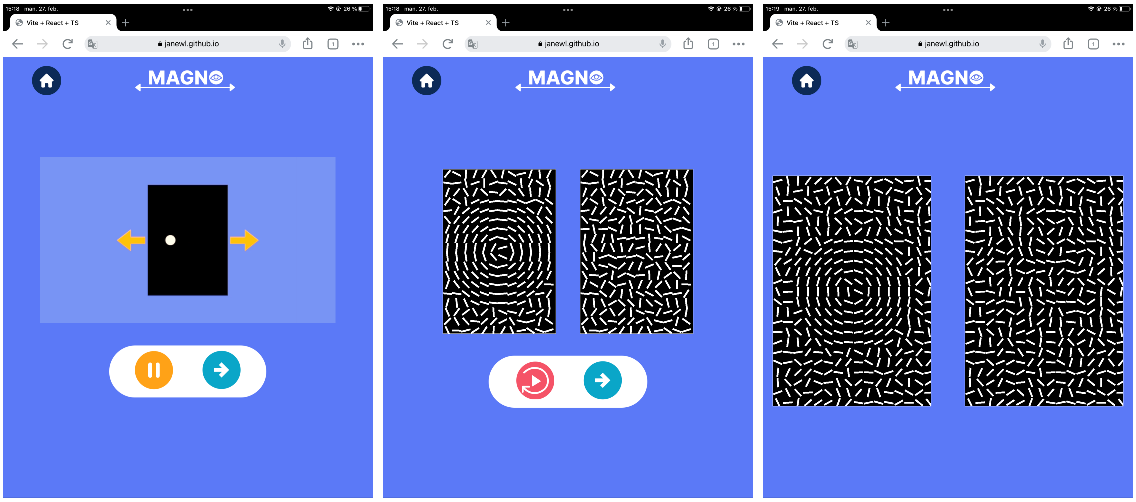
6.1.2 Implementation

Building upon the foundation by Syvertsen [48], it was possible to tailor the system's functionality to the target age group and increase the system's maintainability.

Rewriting parts of the application
A transition was made from only utilizing PixiJS scripts to incorporating React components in accordance with NFR1. The rationale behind this decision was presented in section 5.2. The result of this change was that the new user interface now only uses React with Typescript. However, the tests themselves, along with their underlying functionality, remained mostly unchanged from previous versions of the application and continued to be implemented using PixiJS and JavaScript. Nonetheless, specific alterations were made to the tests.

Change parameters on tests
To create a more child-friendly testing experience, several adjustments were made

to the parameters of the tests. In accordance with FR9, the size of the tests was increased to enhance visibility and ensure ease of interaction. Furthermore, the length of the test, measured in terms of reversals, was reduced from 8 to 6 reversals, thereby shortening the overall test duration. This was one in the effort to maintain focus the child's focus for the duration of the test [14]. A comprehensive overview of the complete test settings can be found in Appendix B. To assess the validity of the new version, some initial tests, explained in the next section, were conducted.



(a) Tutorial video (b) Trial page (c) Test page

Figure 6: Interface version 1

6.1.3 Validating changes to visual processing tests

In line with the objectives stated in section 6.1.2, modifications were made to the size and number of reversals of the tests in order to better align them with the capabilities of the target age group. To verify that these changes did not result in a significant gap in test results compared to the original test, a validation process with five test subjects was conducted. These individuals were previously unacquainted with the Magno application. The results obtained from their performance on both versions of the test have been carefully documented and are presented in Figure 7.

The tests give each user a score from 1 to 100, where scores closer to 1 indicate fewer chances of having dyslexia and higher scores indicate higher chances. The analysis of the test results indicated that the modifications made to the test parameters resulted in a maximum deviation of 4.8 points from the original test. On average, the variations amounted to 2.9 points for the motion test, 4.2 points for the fixed form test, and 3.1 points for the random form test. Given these findings, it was deemed acceptable to implement the aforementioned changes in the new version, as they did not significantly affect the overall performance and remained within an acceptable range of deviation.

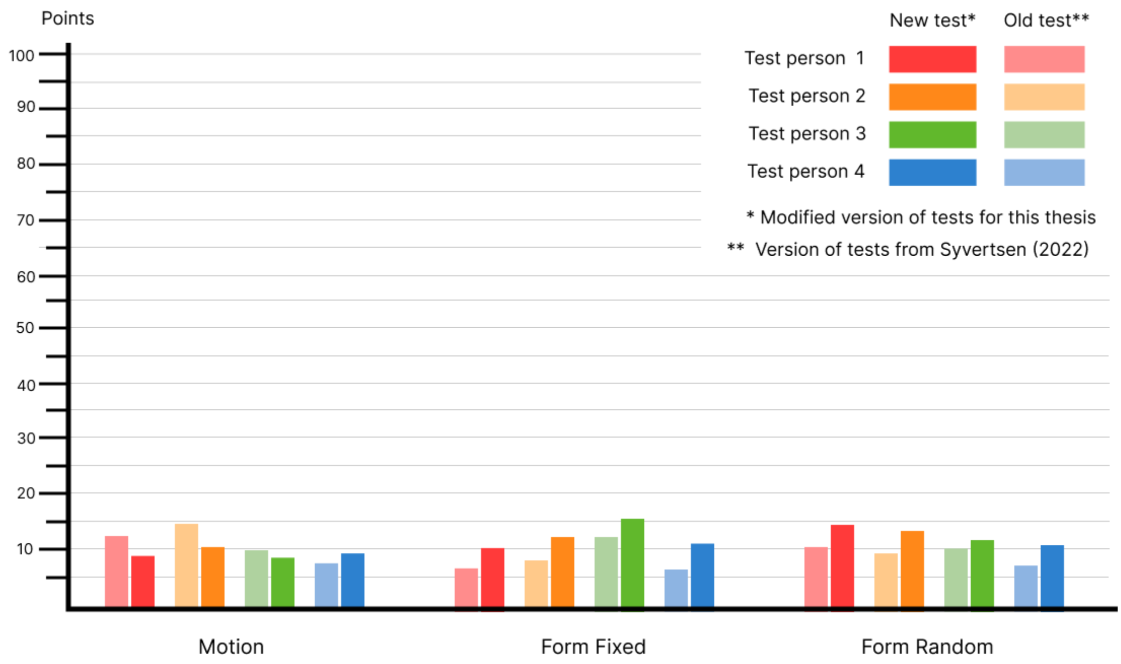


Figure 7: Results from validation of test changes

6.1.4 User testing

The new version of the Magno application was user tested with five children within the target age group. This section will present the results from these tests which were based on observations, quantitative data, and answers to follow-up questions. This feedback will be used as a basis for enhancing the application for future iterations.

Understanding the test

During the user testing sessions, it was observed that several children attempted to interact with the screen while the instructions video was still playing. This behavior occurred when the voiceover in the video prompted them with statements such as *"In one of these images there is a circle/horizontal moving dots, can you see what image it is?"*. Furthermore, one child even responded to the video by stating, *"No, I do not"*.

Following these interactions, the children proceeded to skip the trial test. Only one child demonstrated a partial understanding of the test in the motion test, but expressed confusion and wanted confirmation that they understood the test correctly. Despite skipping the trial test in the form tests, most of the children seemed to understand the concept of identifying the circle, promptly initiating their search for the circle's location. Still, one child struggled with the form test and required an explanation before starting. This child had also skipped the tutorial.

In addition, two children also expressed confusion about the purpose of the application as a whole, one asked *"What am I really supposed to do in this?"* and another asked *"How do I exit this?"* in the middle of the test.

Navigating the application

In this version of the application, the children were granted a significant degree of control over the navigation in the user interface. However, user testing revealed that the children found several aspects of the navigation confusing.

All children exhibited a tendency to bypass the trial page and proceed directly to the actual test without attempting the trial test. It was observed that the audio instruction, stating *"Click on the arrow button when you are ready to begin,"* prompted the children to locate and click the button immediately.

Additionally, two children encountered difficulties in navigating through the test due to utilizing the navigation buttons in a different way than intended. They switched back and forth between the video and the trial page, failing to advance to the actual test. Consequently, these children required assistance to proceed with the test.

Motivation for the test

Throughout the majority of the tests, the children demonstrated a high level of concentration and engagement. However, a notable exception arose when the tests became repetitive. One child explicitly expressed their boredom, stating, *"It's boring to just, click, click"*. This suggests that the repetitive nature of the tests may have led to decreased interest and motivation among the children.

Nevertheless, in most cases, the children made a genuine effort to perform well, at least at the beginning of the tests. They either exhibited confidence in their selection of the correct box or took a few seconds to carefully consider both boxes before making a choice. However, after reaching lower levels of coherency most children started clicking at random or looked at us for validation before clicking. This mostly attentive approach was evident across the tests, with the exception of the motion test. In the case of the motion test, one child expressed reluctance to participate altogether, while three children immediately resorted to clicking randomly without following a specific strategy or pattern. One of these children seemed unmotivated throughout the entire test, also in the form tests. This child sighed loudly and several times stopped to look at other objects in the room seeming to lose interest in the application.

Self perceived satisfaction

At the conclusion of each test, the children were asked some questions to provide feedback regarding their experience during the testing process. The feedback collection process was facilitated by the utilization of the smileyometer, as described in section 4.3.3. Subsequently, based on the child's initial responses, follow-up questions were posed to gain further insights into their perceptions and feelings. The feedback received from the children exhibited a mixed response overall. However, it is important to note that in the case of the motion test, one child was unable to complete the test, resulting in a response of N/A for this specific test.

Positive: Two children consistently rated both tests in the application with a rating of 4. These children expressed that the tests were somewhat challenging, but despite this, they still found them enjoyable. When asked about what aspects they found fun, the reasons mentioned were *"It was fun to find the circle"* and *"It was fun to click"*. One additional child did not manage to complete the motion test but reported enjoying the form test, mentioned finding the "ping" button confirmation sound funny, and gave this test a 4 rating.

Negative: Two other children expressed dissatisfaction with their experience following the completion of the tests. One child assigned ratings of 2 (motion) and 3 (form) to the respective tests and said *"It made me tired"* as the reason for their dissatisfaction. Another child gave a rating of 1 (motion) and 2 (form) and described their experience as *"Very boring"*.

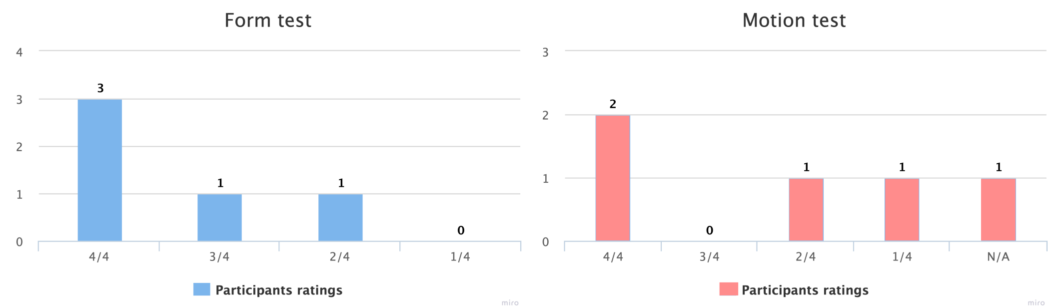


Figure 8: Self perceived satisfaction from version 1

6.1.5 Key findings

- KF-1.1: Too much freedom in navigation causes children to either skip or get stuck on the trial page.
- KF-1.2: Instructions must be specific.
- KF-1.3: The concept of the motion test is hard to understand.
- KF-1.4: The tests are perceived as repetitive and boring.
- KF-1.5: Audio feedback on interactions is perceived as positive.
- KF-1.6: Three children reported enjoying the tests, while two children reported they disliked the tests.
- KF-1.7: Two children expressed confusion about the purpose of the application.

6.2 Second iteration

The second iteration of the application aimed to enhance its overall quality based on the feedback obtained from the previous round of user testing. The primary objectives of this iteration were to improve the overall usability and navigation of the application and to introduce gamification elements to enhance user engagement. A new round of user tests was then carried out to evaluate this new version's effect on motivation and usability. The key findings from this user test are then presented at the end of this section.

6.2.1 Design

This new version was based on the version from the first iteration and the findings from its user testing. The new design was also based on the same requirements as used in the last iteration (Table 2). All the illustrations and designs used in this iteration were exclusively created for this project using Figma. In this section, an examination will be made of the additional requirements that were identified following the previous round of user testing and the corresponding design choices made to address these requirements. Furthermore, any requirements deemed unnecessary or dropped based on insights gained from user feedback will also be highlighted.

Requirements

Following the user testing conducted after the first iteration of the application, several new requirements were identified and included in the revised list of requirements (see Table 3). These newly added requirements are FR17 to FR22, with FR15 and NFR2 being taken from the original set of requirements (Table 1).

Furthermore, based on the insights gained from the user testing, it was decided to drop two requirements from the initial list. FR3, which focused on the ability to repeat the instructions, and FR11, which aimed to provide the option to skip the instructions at any time, were deemed to only cause confusion in the navigation and were for this reason removed from the included requirements.

FR15	The test should include a personalizing option
FR17	The application should guide the child to navigate correctly through the application
FR18	The application should clearly communicate how it is intended that the child should interact with it
FR19	The application should include elements that decrease the feeling of repetitiveness
FR20	The application should include elements that grab the child's attention
FR21	The application should have stated a clear goal to the child
FR22	The tests should be explained in a short and precise manner
NFR2	the application must have a maximum latency of 2 seconds for video, animation and image display

Table 3: New requirements implemented in iteration 2

The navigation was confusing for the child

In order to fulfill the requirements stated in FR17 and facilitate the intended navigation of the application, modifications were made to the trial page (Figure 9c). The skip button now appears only after the child has correctly chosen two boxes. This adjustment ensures that the children do not skip the trial page before proceeding to the actual test (KF-1.1).

To address the challenges encountered in meeting the intended navigation goals, the skip and replay buttons associated with the instruction video were removed. These buttons led to confusion as the child was either taken back to the beginning of the video or redirected to the trial page without having seen the tutorial.

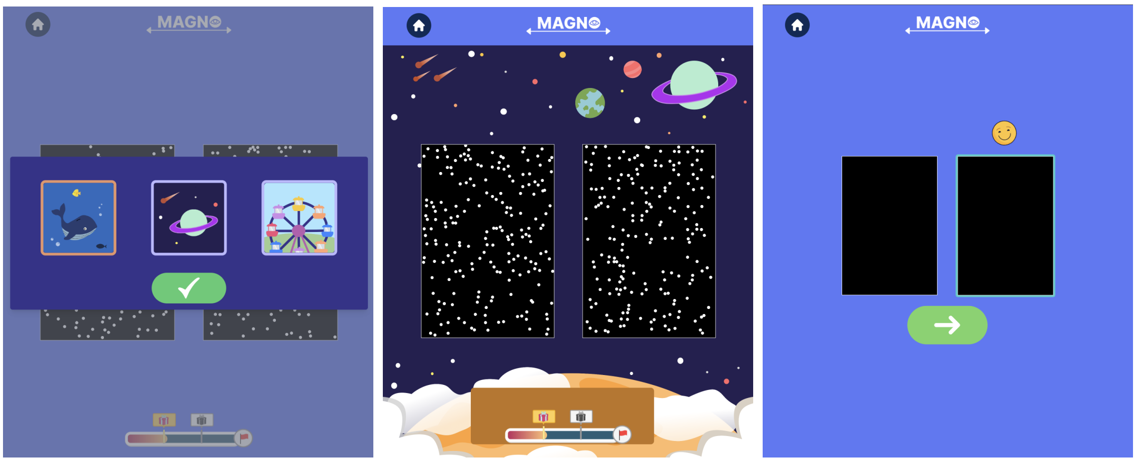
The tutorial was confusing for the child
In order to fulfill the objectives outlined in FR22, which aimed to present the tests in the application in a more accessible and comprehensible manner, several adjustments were made. The instruction video was shortened to convey the information in a more direct manner. By reducing unnecessary content and focusing on the key concepts, the revised video aimed to provide a clearer explanation of the tests to the children. Furthermore, examples that could be misunderstood as the trial page were removed to avoid any confusion or ambiguity (FR18, KF-1.2).

Additionally, upon entering the trial page, the voiceover was modified to deliver a concise summary of the test's objective. This alteration aimed to provide a brief reminder to the children about the goal of the test, allowing them to approach the test with a clear understanding of what they needed to accomplish.

The test was repetitive and boring for the child
Test results indicated that some of the children experienced the test as repetitive and boring (KF-1.4). As mentioned in the preparatory project, this could lead to reduced concentration, thereby affecting the validity of the test results. To address this issue, gamification was identified as a potential solution. Hence, it was decided to test the effect on motivation by incorporating gamification elements.

Feedback was recognized as one of the key gamification elements that could enhance engagement (see section 2.5). The findings also revealed that the children enjoyed the audio feedback provided in the first iteration (KF-1.5). This suggested that additional feedback sounds could increase motivation. Therefore, in this iteration, more feedback sounds were implemented as part of FR15, which required the test to include a personalization option. The inclusion of this requirement aimed to reduce the feeling of repetitiveness throughout the test (FR19). To make the test page more interesting and fun for the child (FR20), customization options were introduced, such as the ability to choose a background theme (space, ocean, or carnival) (Figure 9a) and select a preferred feedback sound.

Furthermore, a progress bar was added to the application. This progress bar served as a visible goal and aimed to alleviate the feeling of not receiving feedback on progress (FR21). Additionally, it included two small gift icons to indicate how far the child had progressed in the test, providing an incentive to reach the next milestone and receive the next prize (Figure 9b).



(a) Customization menu (b) Test with background (c) Trial page

Figure 9: Interface version 2

6.2.2 Implementation

Performance optimizations

As the application was meant to run on lightweight iPads and tablets with varying performance capabilities, optimization was an important consideration (NFR2). To address these issues, all images and animations were compressed and fetched dynamically. Preloading on audio files was also implemented to reduce delays.

6.2.3 User testing

In the current iteration, user testing was conducted in a kindergarten with a total of six preschool children aged 5 and 6 years old. In this section, outcomes derived from the thematic analysis conducted on the collected data will be presented. Furthermore, the feedback obtained from the users and key findings derived from this iteration will be presented.

Navigation

In this iteration, it was observed that the navigation was easier for the children compared to the last user testing. The absence of a skip and replay button for the tutorial video ensured that they did not get stuck in repetitive video loops or completely skip the video and fail to understand what to do.

However, some issues with the navigation in the application were still identified. Four of the children did not understand that they had to click on the green button to confirm their choice and exit the customization modal.

Additionally, five of the children remained on the trial page. The children who did not navigate from the trial page either did not realize that it was not the actual test or found the smileys indicating the correctness of their answers amusing. In these instances, the children needed encouragement to continue with the test.

Understanding the test

During the form tests, all the children demonstrated an understanding of the test thanks to the new tutorial video and the repetition of the goal during the presentation of the trials. However, when it came to the motion test, all of the children encountered difficulties in identifying the correct box, with two of them resorting to randomly clicking on boxes on the test page. When asked for their thoughts on the motion test, two children mentioned that it was difficult. Furthermore, one child remarked that the motion test was harder compared to the form random test.

Observations also revealed that the children who initially attempted the form tests appeared to be less confused by the motion test. Conversely, the two children who started with the motion test were observed to struggle considerably more than those who began with the form tests. One child who began with the motion test quickly became frustrated and refused to continue when they were unable to comprehend the instructions. However, this child successfully completed both the video and trial of the form test without encountering any difficulties.

Motivation for the test

During the user testing, it was observed that the children displayed a higher level of focus on the tests compared to the previous iteration. Throughout the form tests, all of the children remained seated and did not exhibit signs of becoming unfocused or bored. In the motion test, one child immediately started clicking at random, and two children resorted to random clicking when the coherence percentage dropped. Notably, when engaging with the sound customization feature, all children, except one, demonstrated signs of enjoyment by laughing or smiling.

Furthermore, it was observed that three of the children hesitated to choose a box when they were uncertain about their answer. Despite the presence of a hand animation pointing at the boxes, it was not sufficient to motivate these children to make a selection within the time limit. In such cases, additional encouragement like "If you were to guess, then?" was required.

Another interesting observation was that two of the children seemed to expect that their chosen customizations would have some impact on the actual test. One child expressed the expectation that the icon for the sound customization they selected (a rabbit) should be placed inside the boat displayed in the ocean theme. Similarly, another child who chose the carnival background inquired about what they were supposed to do within the carnival setting.

Self perceived satisfaction

The responses gathered through the smileyometer indicated an overall improvement in the children's perception of the tests compared to the previous round of user testing. However, it is important to note that in the motion test, two responses were marked as N/A. One child expressed a reluctance to complete the test, while another child encountered technical difficulties toward the end of the motion test, leading to invalid test results for that particular test. The detailed results from the smileyometer can be seen in Figure 10.

Positive: Two children expressed their enjoyment, with one stating enthusiastically, *"It was fun! I liked clicking."* Both children specifically mentioned their liking for a particular customization sound option. These children gave the tests in the application a score of 4, indicating their high level of satisfaction.

Neutral: Two additional children enjoyed choosing the customizations, mentioning that *"The carnival and sheep were fun"* and *"Choosing carnival was fun."* However, despite their positive reactions to the customization feature, both children also mentioned finding the application boring, indicating mixed feelings about their overall experience. These children assigned ratings of 2, 3, or 4 to the tests, reflecting varying levels of satisfaction and engagement.

Negative: Furthermore, two other children described the application as "boring," with one child elaborating that *"Choosing the background and sound was boring"*. One of these children assigned lower ratings of 1 and 2 to the tests, indicating a lack of enjoyment. The other child provided a rating of 4 and when asked why, the child simply said, *"Because it is happy."*

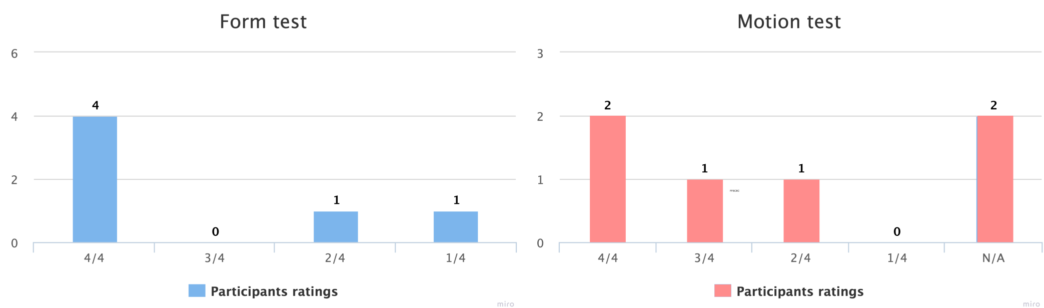


Figure 10: Self perceived satisfaction from version 2

6.2.4 Key findings

- KF-2.1: Fewer options for navigating lead to less confusion for the children and more independence.
- KF-2.2: The children did not always understand the meaning of the "continue" and "confirm choice" buttons in the application.
- KF-2.3: The children did not want to choose a box when they were not confident about their answer.
- KF-2.4: The customization options were enjoyed by the children.
- KF-2.5: Two children reported enjoying the application, two reported having mixed reactions, two reported having negative feelings.
- KF-2.6: Focus was higher in this iteration compared to the last.
- KF-2.7: The motion test was hard to understand for the children, especially when it was the first test they tried.
- KF-2.8: The children understood what to do in the Form Fixed and Form Random tests.
- KF-2.9: The children expected the customization to have a bigger impact on the test.
- KF-2.10: The children stayed on the trial page.

6.3 Third iteration

In the third iteration, the goal was to improve the application based on the feedback from the previous iterations. Creating an enjoyable application that would capture the children's attention and motivate them to put in their best effort throughout the test was one of the main focuses. As this was the last iteration presenting the final results from user tests, the design and findings will be described in more detail than the previous iterations.

6.3.1 Design

The design in the last phase was based on versions from the previous iteration and the findings from its user testing. The new design was also based on the same requirements used in the previous two iterations (Table 2, Table 3). As in the last iterations, the illustrations and designs used in this iteration were custom-made for this project. This section presents requirements that were added or dropped after the last user testing and their corresponding design choices.

Requirements
Based on the findings from the user testing of the tests after the second iteration, three new requirements were added to the system (Table 4). FR14 and FR16 are from the original list of requirements (Table **??**). Additionally, the requirement FR5, "The test's time constraints should be increased to 10 seconds," was removed. The reason for this change is discussed below.

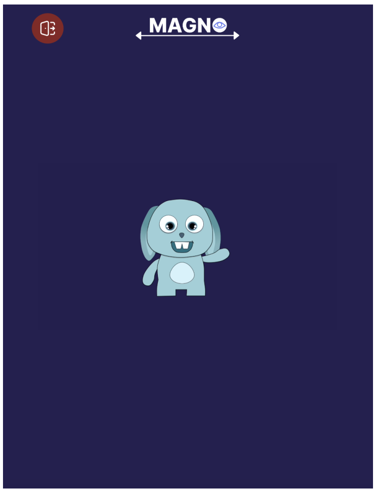
FR4	The application should have a continuous story
FR14	The test should have engaging animations
FR16	The test should include a theme
FR23	The trial page should have a maximum number of trials

Table 4: New requirements implemented in iteration 3

By incorporating these requirements into the system, all of the functional requirements gathered during the preparatory project were included (see section 2.5). The only exception to this was FR12: the test should include background music with positive associations. This requirement was excluded due to concerns that the additional music would be distracting, considering the application already incorporated a significant amount of audio.

The application was not motivating enough for all the children
In the previous iteration of testing, an equal number of children reported the application as enjoyable and uninteresting (KF-2.5). During the preparatory project for this thesis (see section 2.5), themes and storytelling were identified as having positive effects on motivation. Therefore, in this new version (FR16, FR4), the application was designed to revolve around a story of traveling. The customization options from previous iterations, which included different backgrounds, were used to display the various places the children would visit on this journey.

The journey start with a short animated introduction video (Figure 11a). In the video, an avatar named Magno, who is a blue bunny, introduces himself. Magno explains his fondness for traveling and solving tasks. After the video, the child was directed to the front page (Figure 11b), which now featured a progress path illustrating the journey they would undertake. The progress path ensured that the tests were taken in a specific order (KF-2.1, KF-2.7) and will be described in more detail below. The journey started in the ocean, proceeded to the carnival, and ended in space. Each time the child advanced to the next step, they viewed the front page and traveled using the vehicle they had chosen during customization.

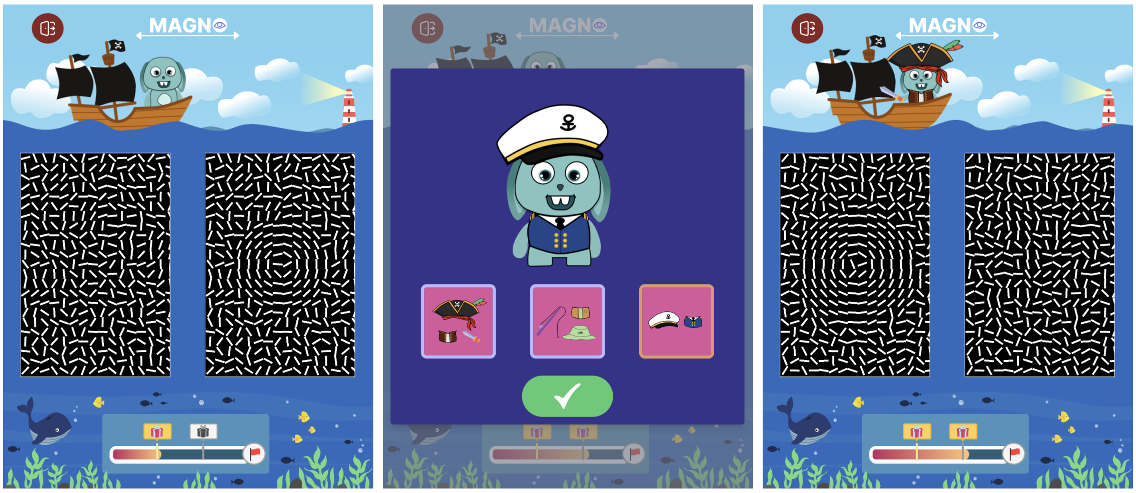


(a) Magno introduction video

(b) Journey overview page

Figure 11: Interface version 3 - Introduction and overview page

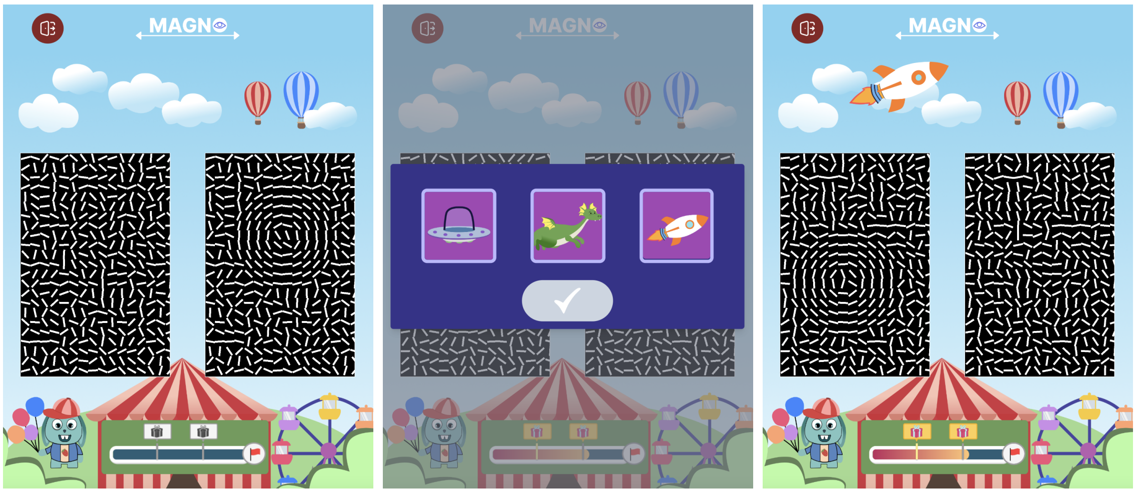
Previously, the children were both observed as well as self-reported to enjoy the customization option (KF-2.4). Also, enjoyment was higher in the second iteration compared to the first iteration (KF-1.6, KF-2.5), suggesting that the element of choice could have had a positive impact on motivation (KF-2.6). To enhance this gamification element, additional customization options that aligned with the themes were implemented. In the ocean section of the test (form fixed test), the child had the opportunity to choose a boat and later a naval clothing option. The chosen boat and avatar with the selected clothing would then appear at the top of the screen (Figure 12). Additionally, FR14 was implemented to increase interactivity within the application. The waves were animated to move, and a calm dipping animation was added to the boat.



(a) The avatar in chosen boat (b) Choose outfit menu (c) Boats and outfit

Figure 12: Interface version 3 - Form fixed test

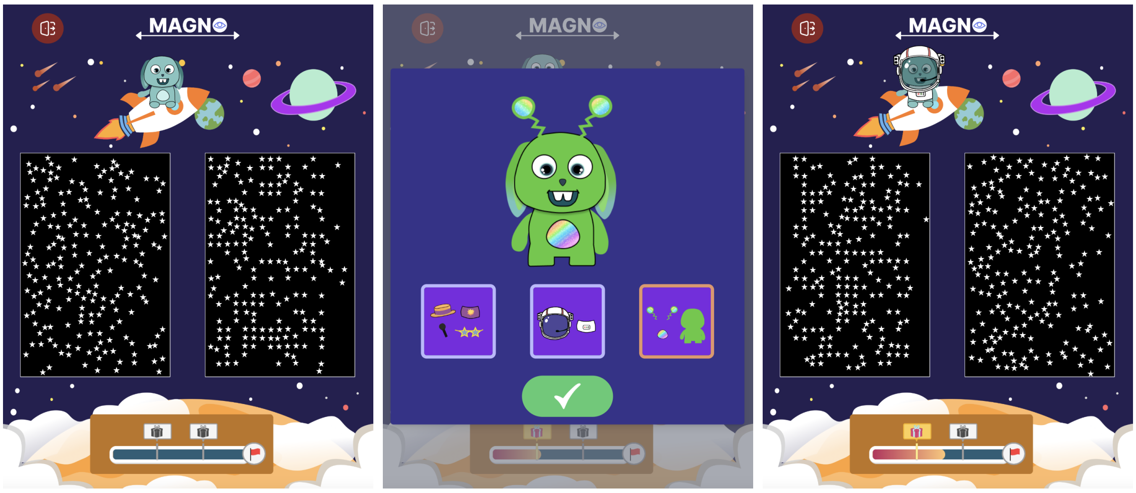
The second step of the journey (form random test) featured the carnival theme (Figure 13). Here, the child first customized the feedback sound. The sound options were the same as used in iteration two. The second customization involved choosing a vehicle for traveling to space, which would then be displayed in the sky at the carnival. Similar to the boat, the spaceship had a subtle vertical movement animation.



(a) Carnival theme (b) Choose spaceship menu (c) After spaceship option

Figure 13: Interface version 3 - Form random test

In the third and final step of the journey (motion test), the space theme was featured (Figure 14). Magno appeared at the top of the screen, piloting the vehicle chosen by the child in the previous test. The first customization option involved selecting a space outfit for Magno, while the second customization option allowed for choosing a new feedback sound. The chosen sounds remained the same as those selected during the carnival customization. To align with the theme, the dot sprites were replaced with star sprites, and the spaceship drifted calmly in space with a subtle movement. Throughout all the tests, great care was taken to ensure that the animations and background images did not disturb or distract the child's attention. The animations were deliberately designed to be slow, small, and repetitive. The backgrounds have intentionally no disturbing elements behind the boxes.



(a) Space theme (b) Outfit menu (c) Avatar with outfit

Figure 14: Interface version 3 - Form random test

The navigation was not clear for the children

The feedback received from the previous iteration indicated that the order in which the children took the tests had an impact on their understanding and, consequently, their results (KF-2.7). It was also observed that limiting the number of navigation options made it easier for the children to independently navigate through the test without adult guidance (KF-2.1, KF-2.2). Based on these findings, it was decided that the tests should follow a predetermined order. The test now always started with the form fixed test, followed by the form random test, with the motion test being the final one.

In addition to the fixed order of the tests, more audio cues were incorporated into the test. In the previous iteration, the children consistently responded as intended when provided with simple prompts by the facilitators, such as *"Click the green button when you are ready"* (KF-2.2). In this version, audio cues were added in the application to remind the children to click the button at the bottom when they were ready, both in the modal and during the trial (KF-2.2).

To encourage children who remained on the trial page (KF-10) to proceed to the actual test, a maximum of 5 attempts was allowed on the trial page before the child was redirected to the test page (FR23). Additionally, the child was required to select two correct answers in a row before advancing to the next stage. This measure was implemented to ensure that the children comprehended the tasks before progressing further.

The tests were difficult for the children

To encourage the children to make a choice when they were uncertain about the correct answer (KF-2.3), a new audio cue was introduced. Similarly to the button navigation cues this audio was implemented based on the encouragement the children needed in the last user tests. If the children hadn't provided an answer within 15 seconds, an encouraging voice now kindly suggested, *"You're allowed to take a guess if you're uncertain"*. Furthermore, the hand animation and the 15-second time limit were removed. This decision was made based on observations that indicated these elements did not significantly motivate the children to make a choice, but rather increased their uncertainty (KF-2.3).

Efforts were made to address the issues regarding the children's understanding of the motion test, with some even giving up entirely (KF-2.7, KF-1.3). Firstly, the video was modified. It now included a more accurate example from the test, and the language was adjusted to provide an easy-to-relate-to explanation of the test. The wrong box was described as chaos, while the right box was described as orderly and nice. Furthermore, the coherence percentage was increased, as discussed in the implementation section below, in order to improve the test experience.

6.3.2 Implementation

Change to coherency percentage

Earlier projects on this application have used coherency percentages of 50 [14]. In this iteration, the coherency percentage was changed to 75 percent. This was done in accordance with earlier studies with children and the tests in the Magno application [36, 37].

6.3.3 User testing

In the final iteration of this project, the last version of the application was subjected to user testing with seven preschool children in our target age group. This section presents the user feedback collected on the application, which will serve as the basis for future improvements.

Understanding the test

During this iteration, the form trials proceeded smoothly, and all participating children demonstrated a solid understanding of the tests. Notably, every child consistently selected the correct answers in their initial two attempts, further reinforcing the findings from the second iteration that the children comprehend these tests.

As described in the design section of this iteration, the tutorial video for the motion test was modified compared to the previous version. This change, along with the revised order of the test, resulted in six out of seven children successfully completing the motion trial without difficulty. These children correctly chose two answers in a row before proceeding to the actual test. The effectiveness of the video was evident as some children nodded or replied *"yes"* to the question, *"Do you see the difference between these two pictures?"* while the video provided an example of a correct and incorrect box. The one child who initially struggled with the trial eventually succeeded after a few attempts. This child even vocalized, *"That one! In that one, the dots do not go everywhere,"* demonstrating their understanding.

When transitioning to the actual test, two children randomly selected boxes. However, the remaining five children showed that they managed to complete the tests, choosing mostly correct boxes. In the motion test, they easily identified the correct boxes, although most of the children encountered challenges as the coherency percentage significantly decreased, leading to more frequent guesses. One child specifically commented at the end of the test that the motion test was the most challenging. Another child expressed frustration, exclaiming, *"What should I do?"* when the test became too difficult for them.

Navigating the application

The newly introduced audio prompts proved effective, as observations indicated that six out of seven children consistently listened to and followed the instructions whenever a prompt was given. Additionally, setting a maximum number of attempts on the trial page successfully guided those who wished to remain on the trial page into proceeding to the test.

The only area in the application where navigation posed challenges for the children was the progress path on the front page. The start test button lacked an audio prompt and was somewhat difficult to locate, being positioned at the bottom of the page. Only one child managed to navigate this section without assistance from the facilitator. However, apart from this aspect, five out of seven children successfully navigated the application without requiring help from the facilitator.

One child needed a prompt to understand that they had to click to choose a customization option. This child stated that they thought *"This was hard to understand"* when presented with the customization modal. One additional child struggled with almost all the navigation in the application. However, this may be a result of test anxiety which is presented in more detail in the section below.

Motivation

When the test started, three of the children immediately began clicking at random, showing little effort in attempting to choose the correct box. This behavior persisted across all the tests, despite them answering correctly during the trial, indicating that understanding was not the issue. On the other hand, the remaining four children started the tests with great concentration, carefully examining the boxes and doing their best. However, as the tests became more challenging, one child remained focused until they reached a point where they could no longer identify the correct box. From that moment onwards, they resorted to clicking randomly, even when the coherency percentage returned to its original value, where they had previously had no difficulty identifying the correct box. This child exhibited signs of boredom throughout the test, sighing and looking around the room.

In contrast, the last three children managed to maintain their concentration and continued to put forth their best efforts in every test, even as the difficulty level increased. Additionally, five of the children visibly enjoyed the application, expressing their enjoyment through smiles and laughter during the test. Some children also shared their positive feelings, expressing their liking for the different vehicle and clothing options.

An interesting observation was that one child displayed hesitation in initiating any action independently, even when prompted by the application. Instead, they consistently sought confirmation from the test leader or the kindergarten teacher present during their test. When identifying the box they believed to be correct, the child would ask, *"Is it this one?"*, seeking validation for their choice. Taking time to explain to the child that this was their choice and no answer was right or wrong did not make the child more comfortable. Another child showed reluctance to engage in casual conversations before the test began. Sensing their hesitation, the child was given the option to leave if they wished, but they chose to stay while also declining to answer questions.

Self perceived satisfaction

To avoid taking away the focus from the child under the test the children only answered the smileyometer and the follow-up questions once. This way, their answer to the smileyometer was not with a specific test in mind but more of their experience with the application as a whole. The results are shown in Figure 15, and the children's self-reported satisfaction is described below:

Positive: Five children expressed only satisfaction with the application. They all reported finding the application fun. When asked about their favorite aspects of the application, one child mentioned *"It was fun to choose the rocket and the boat"*. Two other children expressed similar enjoyment of the customization options, with one child exclaiming *"Everything was fun"*. These children all gave the application a 4 or 3 rating. One additional child did not want to answer any questions but rated the application 4.

Neutral: One child showed a more hesitant reaction to the application. When asked if they thought the application to be fun, boring, or in between the child answered *"Not sure"*. This child gave the application a rating of 3.

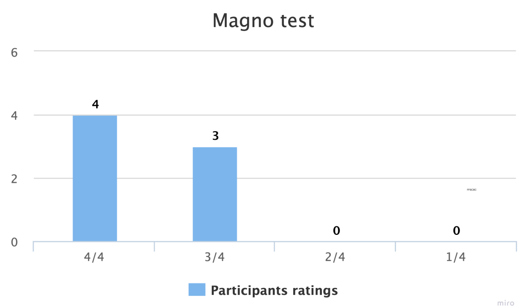


Figure 15: Self perceived satisfaction from version 3

6.3.4 Key findings

- KF-3.1: The navigation on the front page to the tests is hard to understand for the child initially.
- KF-3.2: Audio prompts help the children learn to navigate the application.
- KF-3.3: The application manages to motivate all the children to try their best in the trial.
- KF-3.4: The application motivated five out of seven children to stay focused through the test.
- KF-3.5: The customization options were enjoyed by five out of seven children.
- KF-3.6: The children enjoyed the story and theme of the application.

- KF-3.7: 75% coherency percentage, a more descriptive instruction video, and placing the motion test last eased understanding for the child.
- KF-3.8: Choice of words is important for the child's understanding.

7 Discussion

This section aims to discuss and evaluate if the application successfully achieved the objectives of promoting high usability and motivation among children aged 5 to 7 years. The evaluation will be conducted by looking at the results from the user testings and employing the theories of Piaget, Self-Determination Theory (SDT) and Gameflow.

7.1 Understanding the test

The findings suggest that the choice of words was very important when it came to preschool children's understanding (KF-3.2, KF-3.8). This is in line with Piaget's theory of cognitive development which states that children in the preoperational stage tend to think in a more concrete and literal manner. For this reason, using age-appropriate and concrete language is crucial for adjusting the application to their comprehension and cognitive growth. An example of this literal thinking was apparent when almost all of the children in iteration two immediately pressed the button when the voiceover said they could click it when they were ready. They did not comprehend that *"When you are ready"* indicated that they should complete the trial page first. Another example of this was shown related to the motion tutorial video. At one point the voiceover in the video asked if the child saw the difference between the two boxes. This resulted in several children starting to explain the goal of the test out loud. This could indicate that the video asking them questions made them reflect and actively think about how to identify the correct window, helping them process the goal of the test. An interesting point was when a kindergarten teacher asked if the child know what the word *"guess"* meant. The child did not remember, indicating that the language used was too complex. This demonstrates how kindergarten teachers have a lot of insights into what preschool children do as understand.

On the trial page, the child got to try the test with feedback on whether the answer was correct or incorrect. An important point was forcing the child to try the test until a minimum of two correct answers were provided. This seemed to be important to ensure that the children who did not understand the test by watching the video, understood through learning-by-doing.

To cope with children's cognitive abilities, changes were done to the coherency percentage. Initially, the low level of coherency presented a challenge that was too difficult for the child, resulting in the children feeling frustrated and overwhelmed.

However, by starting at a more suitable level of coherency aligned with the child's skill level, the test provided a smoother and more manageable progression. This enabled the child to better engage with and comprehend the test, fostering an enhanced understanding of the visual patterns (KF-3.7). In addition, time constraints on the tests were removed in accordance with findings from Klevstuen [14]. Earlier work on Magno has argued that the test's time constraints are not the primary limiting factor for coherent motion detection [37]. However other studies show that dyslexics have higher response time as a result of visual processing deficits [35]. This implies that removing this upper time limit could harm the validity of the test results, though this needs to be verified.

After several changes were made related to the video, the test parameters, and the order of the tests, the understanding seemed to increase. This was especially true for the motion test, though the understanding of the form tests seemed to be high already in the second iteration (KF-2.8). Because so many factors were included at the same time it is difficult to isolate which of these changes had the biggest impact, though it is likely that the combination could be important.

7.2 Navigating the application

It did not seem like many preschool children were familiar with navigating simple interfaces. Problems with understanding the purpose of the continue, confirmation of choice, skip, and replay buttons were common in earlier iterations (KF-1.1, KF-2.2). Most of these issues were solved by adding guiding audio prompts and animations that played whenever the child did not interact with the test within a given time. These prompts often indicated what would happen if a specific button was clicked. That these explanations made the children understand what to do could be the reason why most of the children only needed to be prompted once to know what this button meant for the duration of the test. For this reason, lack of knowledge about navigation and buttons' meaning could be the reason for the children's hesitations; children's inexperience with using tablets required a learning process. In accordance with Piaget's theory of cognitive development, the absence of assimilation suggests that the children could not fit the information about the navigation and the buttons' meanings into their existing mental frameworks. Therefore, they had to actively adjust their understanding and accommodate these new concepts to navigate the application successfully. This process of accommodation reflects their cognitive effort to adapt to unfamiliar technology and integrate it into their cognitive schema. In addition, fewer navigating choices and limiting the attempts on the trial page also created limited choices for the child. Minimizing the options available reduced the learning curve associated with the buttons, resulting in reduced confusion regarding which button to click.

7.3 Motivation

In iteration one, some children were intrinsically motivated by the mere engagement with the application, enjoying the experience regardless of the feedback provided. This is seen when one child commented *"It was fun to click"*. However, for others this appeal was seen to be lower, one child commented *"It was boring to just click, click"*. Gaps in motivation can also be seen in later iterations, for instance among the customization options.

To explain these differing views, it is important to recognize that intrinsic and extrinsic motivation can vary among individuals. What may be intrinsically motivating for some children, may not be for others [31]. The child enjoying the clicking may be intrinsically motivated by this, while other children find this less interesting. For this reason, looking at specific elements and determining how this motivates the children can be difficult. However, the Gameflow model and SDT can be used to analyze in more detail how motivation can be increased by including different factors.

7.3.1 Gameflow

Some of the children seemed to choose the wrong box on purpose, not trying their best to choose correctly (KF-3.3). Interestingly, several children gave us feedback that the test was easy and that their satisfaction was high, even if they choose all the wrong boxes or clearly chose boxes at random. This could be the result of several factors, and this section will employ the Gameflow model to evaluate the reasons for this shortcoming.

Clear goals
An important element in Gameflow is the presence of clear goals. It was observed a significant improvement in the children's engagement and understanding when introduced to a more coherent story with the clear goal of traveling from the ocean to space. This positive change was shown as the children no longer expressed complaints or questioned the purpose of the tests, as they did during the initial iteration (KF-1.7).

However, it is important to note that Gameflow also emphasizes the importance of intermediate goals. The implementation of a progress bar at the bottom of the screen during the test phases was meant to fulfill this by indicating the child's progress in the test. The progress bar included gift icons that would light up as the child reached specific milestones, with the gifts being different customization options. However, it is uncertain whether the children paid much attention to this progress bar, especially considering its placement at the bottom of the screen, which might have made it difficult to notice.

This raises the question of whether it would have been possible to make it easier for the children to understand these intermediate goals provided throughout the application. For instance, it may have been beneficial to incorporate additional visual cues or prompts to highlight the significance of reaching certain points in the progress bar. By making the milestones more noticeable and understandable, it would have been possible to enhance the children's sense of progress and direction within the application, aligning with this principle of clear goals in Gameflow.

Feedback
In the Gameflow model, feedback is an element emphasizing the importance of providing immediate feedback to children, informing them about their actions and their progress toward the goal. Being informed about their actions also aligns with SDT, which highlights how feedback enhances the sense of competence. When positive feedback is received, the feeling of mastery and competence increases. In the application, children primarily receive feedback on the trial page, where they are informed whether their responses were correct or incorrect. All the children displayed a high level of concentration and were careful not to choose incorrectly during this phase (KF-3.3). However, during the actual test phase, it was evident that some children who had selected the correct response in the trial phase immediately began to select random answers when the test phase started (KF-3.4).

This demonstration of increased focus when children received feedback, suggests that incorporating feedback in the actual test phase of the application could be beneficial. However, a potential concern arises when considering the inclusion of feedback in this phase. The preparatory project findings suggest caution against this, as feedback on performance could potentially lead to low self-esteem among children who perform poorly. In addition, SDT also warns that receiving negative feedback can decrease feelings of competence and thus also negatively affect intrinsic motivation. This presents a dilemma between avoiding potential negative effects on self-esteem and the potential benefits of increased feelings of competence through feedback.

Striking a balance between these two opposites is a challenge that requires careful consideration and experimentation to identify the most effective feedback strategies. In addition, it could be beneficial to explore other alternative approaches to providing feedback that is informative and supportive without risking negative emotional consequences.

Challenge and player skills

Another crucial aspect highlighted by Gameflow is the balance between challenge and player skills. As the test progressed, the coherency percentages decreased which made it inevitable that the test at some point would become too difficult for the child to determine the correct answer. According to Gameflow, challenges that surpass the user's skills can lead to feelings of anxiety and ultimately result in the user giving up. This was evident during the motion test, which the children consistently found harder compared to the form tests. On the other hand, when tests were too simple and did not require much effort or skill, users would lose interest and become disengaged or even bored. This could have posed a risk in the form tests, as they may have been perceived as repetitive and overly easy over time due to their similarities. However, as there are limitations in how much can be done to modify the actual tests themselves without risking ruining their validity, it is possible that this issue can be addressed by improving other elements of Gameflow.

Control

To facilitate flow in the application, the Gameflow element of control needs to be incorporated. By providing children with the opportunity to choose customizations as prizes, they were given a small degree of influence over the application which Gameflow states are important. The customization options were added to enhance the sense of autonomy the child had in the application in accordance with SDT.

However, it is important to note that these customizations lacked a specific connection to the test itself. While they did impact the story and avatar, they did not directly affect the tests or their objectives. Moreover, the story and theme did not align well with the purpose of the tests. The only exception is the inclusion of star sprites used in the motion test. However, this is only a vague link. This is problematic because Gameflow highlights that elements unrelated to the game can divert attention from the actual test. Additionally, Gameflow emphasizes the importance of impactful elements that have a significant influence on the user's environment. This gap between the gamification elements and the tests can result in the sense of control or real autonomy being lost, only leaving some superficial or less meaningful influence. Due to the nature of the tests, having a more cohesive and meaningful relationship between the gamification elements and the tests is difficult to achieve.

Consequently, limitations in achieving optimal levels of flow and autonomy within the application may arise. Adding more focus to the other elements of Gameflow and SDT discussed in the other sections could be beneficial to increase overall motivation.

Immersion
Immersion is another concept in the Gameflow theory which is important in facilitating a fully engaged and absorbed state during the assessment. In line with Gameflow principles, audio and narrative elements were used to enhance immersion in the application.

The inclusion of audio features, such as funny and exciting sounds, may have contributed to a sense of presence and enhanced the overall feeling of immersion. In addition, the narrative elements, particularly the story of the avatar character named Magno, were incorporated to create a compelling setting for the children. Magno's love for travel and problem-solving became the central theme of the application's narrative. The locations ocean, carnival, and space within the story further contributed to the immersive experience by evoking a sense of adventure. Findings showed that the story and theme setting were enjoyed by the children (KF-3.6).

In addition, the friendly and approachable presentation of the avatar was meant to create a further connection with the child. This, in addition to the familiar setting of the locations, could have increased the child's relatedness, which is an important aspect of SDT. This positive relationship with the avatar Magno could have helped foster intrinsic motivation, as the children could feel a sense of connection and identification with this character.

Concentration
The results indicate that the Magno application effectively addresses the aspect of concentration in the Gameflow theory. In accordance with the concentration criteria in Gameflow the children were presented with elements that grab their attention. This includes animations, colors, sounds, story and theme. These elements enhanced the experience while not distracting the user from the tasks. For instance, the animations were subtle and the background did not have disturbing elements behind the test boxes. The results demonstrated that most of the children exhibited a consistently high level of concentration throughout the test (KF.3.4). This is opposed to the first iteration where the test was seen as boring (KF-1.4) leading to some children struggling to maintain concentration. For this reason, there are grounds to argue that these elements have helped increase concentration.

However, another factor to consider was the presence of potentially distracting elements within the application. Gameflow emphasizes the importance of eliminating any elements that could divert the child's concentration from the main test. In this regard, it is possible to consider that the modal with prizes in the form of customization options was a potentially distracting element. However, this should be further investigated.

8 Conclusion

In conclusion, the new version of the Magno application has come a far way in becoming a usable and motivating application for children aged 5 to 7 years. Based on theory in addition to observations and user feedback from 18 children, significant changes were made to the application. Gamification elements and changes to increase usability were implemented in the second and third iterations as a result of findings from previous iterations. Some notable findings were the usefulness of audio prompts and forced navigation to help the child with the navigation, as well as feedback and customizations to increase motivation. The last version of the application was tested on seven children and received an average rating of 3.6 out of 4 when looking at self-perceived enjoyment of the application. Most of the children described the application as "fun". Still, there is work that needs to be done in managing to keep the child's concentration and attention throughout the test.

RQ1.1: Do the children understand how to complete the tests based on the instructions given to them?

In this new version of the Magno application, the children all seemed to understand the goal of the different tests. This was shown in the way the children all managed to complete the trials in a correct and confident way. To achieve this level of understanding, a clear instructions video showing concrete examples in a simple, straightforward manner, and in a way that represents the actual tests was important. In addition, it was important that the choice of words and phrases throughout the application was in line with the child's mental model as well as their level of understanding. To ensure those who did not understand the test after watching the video could grasp it better, it was essential to encourage children to experiment on the trial page, where they could receive feedback on what was right and wrong. Together, these elements ensured that all children seemed to understand all the tests in the Magno application.

RQ1.2: How do children navigate the Magno application?

The children navigated the application in a primarily independent way without the need for help or guidance from an adult. According to the results, most children would likely be able to navigate this test in a group setting as was the goal of this application. Though, this has to be tested. Through the process of making this application, it has been clear that few preschool children are familiar with interfaces and do not understand the concept of buttons with simple icons on them indicating continue, confirm, replay, and skip actions. However, the inclusion of audio prompts was identified as an effective measure for explaining the purpose of these buttons by facilitating a learning process on their meaning. This was shown when most of the children understood the buttons after receiving the audio prompt for the first time. In addition, removing several of the navigation options and creating a more linear and set path for the child through the application resulted in less confusion. However, the issue with the initial navigation on the front page still needs to be resolved.

RQ2: How motivating is the Magno application for children aged 5-7?

Throughout this thesis, the Mango application has come a long way in making the application more enjoyable and thus also more motivating for pre-reading children. The last iteration showed that 6 out of 7 children perceived the application as enjoyable and one last child had a more neutral reaction. This was in contrast to the results from the self-perceived enjoyment of the tests in the first iteration. Here, 2 out of 5 children disliked the application, and one additional child not managing to complete the full test. Still, there is a need to figure out how to further adapt the application to manage to engage all children. This is especially true as even though most children self-reported enjoying the application, observations showed that 2 out of 7 children were not motivated to try their best in the tests. This shows the distinction between self-reported enjoyment and actual motivation to try their best.

RQ2.1: How does the inclusion of gamification elements affect the motivation for the tests?

In this thesis, the gamification elements of feedback, story, theme, and customization were used to intrinsically and extrinsically motivate children to concentrate on their tests. It is not possible to measure the specific effect gamification had on motivation for this thesis. This is especially true since the changes were made to the usability at the same time. Most likely the combination of these elements is important.

While the application successfully captured children's attention, it fell short of eliciting optimal performance from all participants. The Gameflow model reveals several factors contributing to this outcome. Firstly, the application lacked explicit intermediate goals, which could be addressed by incorporating more noticeable and understandable milestones to enhance children's sense of progress within the application. Secondly, incorporating feedback on performance during the actual test phase could potentially have a positive impact on children's focus and performance. Additionally, potential distractions, such as the modal with customization options, warrant further investigation to optimize the overall experience. The Gameflow elements of control and challenge were also important to achieve flow according to this model, but these may be harder to incorporate due to the nature of the tests. For this reason, focusing on the other aspects mentioned may be more beneficial.

8.1 Limitations

In this section, potential limitations that may have affected the validity of the testing results obtained during the user tests of the application will be discussed.

Prior to each test, the children were reassured that it was acceptable if they did not like the application, and that their feedback would not be taken personally. As seen in the result section in each iteration, several participants stated that they enjoyed the application and that the tests were fun. However, the observed body language of the participants during the tests implied that the participants were not always being honest when giving these answers. Several of the children who said they enjoyed the tests or the application displayed dismissive body language with their heads in their hands and eyes fixed on the ceiling. Some children also sighed and seemed annoyed during the test. When asked why they said the application was fun, the children often said they were unsure. For this reason, there are grounds to believe that some of the children, motivated by a desire to please, may have refrained from giving their honest opinions. If this is the case, the validity of the user feedback could be questioned.

In addition, questions could be raised about the ecological validity during user testing. The testing environment, with two facilitators observing the child, deviates from the intended real-world group setting in which the application is designed to be used. These testing conditions may have contributed to test anxiety among certain children, potentially impacting the reliability of the test results. Furthermore, external factors may have had an influence. Many of the children were engaged in play before the user tests. Even though the children participated willingly, there is a possibility that some of the children were eager to complete the test hastily in order to resume their previous activities.

Finally, there were several inconsistencies in the introduction and facilitation during the user testing. One of them was how some of the children had the application presented to them as a game, while some as a task or a website. This distinction could lead to the participants having different mental models before doing the tests which again can lead to differing results. For this reason, a more structured and predetermined manuscript on how to present the difficulties to the participants should have been made.

8.2 Future work

This last section will present suggestions for future work on the project. Concrete action points needed for the application to advance further will also be suggested.

One of the main issues with this new version of Magno is that the tests are not thoroughly validated. The application should be tested and the results compared to the validated version of the Magno application from Wold [53]. Observations made during the tests showed that the children manage to advance to quite low coherency percentages, however, this needs to be measured more precisely. In addition, the results need to be compared to the child's own reading competence when advancing through school. By comparing the application's screening results with the child's own reading competence over time, insights can be gained into the application's ability to identify dyslexia at an early stage.

As mentioned in the discussion, comments from the kindergarten teacher were helpful to realize how some of the wordings in the application were too hard for the child to understand. For this reason, showing the application to experienced kindergarten teachers or teachers working with first and second graders and asking for their insights would give valuable information for making the application suited in the best possible way for the target age group.

In addition, an interesting approach could be to test the application on children who are seven years old. Children who are in their first year of school may demonstrate good adaptability in terms of understanding, sustained focus and test-solving abilities due to being accustomed to these activities in a school setting. Consequently, maintaining concentration during the assessment could potentially be easier for these children compared to the children the application was user tested on in this thesis. In addition, using iPads is common at early ages at school.

The following list outlines the final key areas that require attention and further development in order to enhance the current application's usability as an effective assessment tool:

- Measures according to the discussion on Gameflow should be taken to increase intrinsic motivation in the application.

- The application's responsiveness needs to be enhanced to ensure compatibility across various tablet sizes. Currently, the application is designed for the Apple iPad Air 10.9", with only a few simple measures made towards responsiveness for other tablet sizes.
- Lagging has been an issue during the implementation of this application. Measures have been taken to eradicate such issues, but further testing and optimization are needed.
- The application should know what child is taking the test and send the test results to the database made by Syvertsen [48].
- This test is limited to only testing for dyslexia related to visual deficits in the magnocellular system. As dyslexia can be the cause of several different reasons more tests covering other theories should be added to the system. This way, the system can become a more comprehensive tool for assessing the risk of dyslexia.
- The application should be tested in a group setting with one facilitator.

Bibliography

[1] The British Dyslexia Association. *phoneme.* URL: https://www.bdadyslexia.org.uk/dyslexia/about-dyslexia/what-is-dyslexia (visited on 7th Apr. 2023).

[2] M. E. Boyes et al. 'Relationships between reading ability and child mental health: Moderating effects of self-esteem'. In: *Australian Psychologist* 53.2 (2018). DOI: 10.1111/ap.12281.

[3] V. Braun and V Clarke. 'Using thematic analysis in psychology'. In: *Qualitative Research In Psychology* 3.2 (2006). DOI: 10.1191/1478088706qp063oa.

[4] J. DeRight and Randall S. Jorgensen. 'I just want my research credit: Frequency of suboptimal effort in a non-clinical healthy undergraduate sample'. In: *The Clinical Neuropsychologist* 29.1 (2014). DOI: 10.1080/13854046.2014.989267.

[5] S. Deterding et al. 'From game design elements to gamefulness'. In: *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek 11* (2011). DOI: 10.1145/2181037.2181040.

[6] Figma. *It's time to design on your terms.* URL: https://Figma.com/design (visited on 15th May 2023).

[7] L. Hall, C. Hume and S. Tazzyman. 'Five degrees of happiness'. In: *Proceedings of the The 15th International Conference on Interaction Design and Children* (2016). DOI: 10.1145/2930674.2930719.

[8] P. C. Hansen et al. 'Are dyslexics visual deficits limited to measures of dorsal stream function?' In: *Neuroreport* 12 (2001). DOI: 10.1097/00001756-200105250-00045.

[9] T. Helland. 'Trends in Dyslexia Research during the period 1950 to 2020-Theories, Definitions, and Publications'. In: *Brain Sci.* (2022). DOI: 10.3390/brainsci12101323.

[10] F. Jenssen. 'Magno: An Application for Detection of Dyslexia - From Native to Web App'. Master thesis, Department of Computer Science, NTNU, Trondheim. 2021.

[11] T. H. Johansen and M. Kirkerød. 'Magno: An Application for Detection of Dyslexia - Dyslexia and Interface Design'. Master thesis, Department of Computer Science, NTNU, Trondheim. 2017.

[12] Keynotes. *Presentations that stand out. Beautifully.* URL: https://www.apple.com/keynote/ (visited on 15th May 2023).

[13] A. Khaleghi, Z. Aghaei and M. A. Mahdavi. 'A Gamification Framework for Cognitive Assessment and cognitive training: Qualitative study'. In: *JMIR Serious Games* 9.2 (2021). DOI: 10.2196/21900.

[14] P. A. Klevstuen. 'Magno: Screening Young Children for Dyslexia'. Preparatory project for master thesis, Department of Computer Science, NTNU, Trondheim. 2021.

[15] J. W. Larsen and H. O. Nedregård. 'Magno: A Gamification Approach for Screening Dyslexia on Pre-Reading Children'. Preparatory project for master thesis, Department of Computer Science, NTNU, Trondheim. 2022.

[16] J. Lumsden et al. 'Gamification of Cognitive Assessment and Cognitive Training: A systematic review of applications and efficacy'. In: *JMIR Serious Games* 4.2 (2016). DOI: 10.2196/games.5888.

[17] A. T. Miranda and E. M. Palmer. 'Intrinsic motivation and attentional capture from gamelike features in a visual search task'. In: *Behavior Research Methods* 46.1 (2013). DOI: 10.3758/s13428-013-0357-7.

[18] P. L. Morgan, G. Farkas and Q. Wu. 'Do poor readers feel angry, sad, and unpopular?' In: *Scientific Studies of Reading* 16.4 (2012). DOI: 10.1080/10888438.2011.570397.

[19] T. Nergård-Nilssen and O. Friborg. 'The dyslexia marker test for children: Development and validation of a new test'. In: *Assessment for Effective Intervention* 48.1 (2021), pp. 23–33. DOI: 10.1177/15345084211063533.

[20] Jakob Nielsen. *Why You Only Need to Test with 5 Users*. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ (visited on 24th Apr. 2023).

[21] NIST. *Usability*. URL: https://csrc.nist.gov/glossary/term/usability (visited on 12th Mar. 2023).

[22] Dysleksi Norge. *Praksis for utredning av spesifikke lese- og skrivevansker, matematikkvansker og språkvansker i Norge*. Tech. rep. 2021.

[23] B. J. Oates. *Researching Information Systems and Computing*. SAGE publications, 2006.

[24] E. Omland and V. Hjortland. *1 av 5 får påvist dysleksi på videregående: – Har vært mer stress*. URL: https://www.nrk.no/sorlandet/1-av-5-far-pavist-dysleksi-pa-videregaende-_-systemet-svikter-elevene-sier-dysleksi-norge-1.16103988 (visited on 26th Mar. 2023).

[25] T. A. Petersen. 'An Application for Detection of Dyslexia'. Master thesis, Department of Computer Science, NTNU, Trondheim. 2018.

[26] PixiJS. *PixiJS Guides*. URL: https://pixijs.io/guides/basics/what-pixijs-is.html (visited on 12th May 2023).

[27] Y. V. Post, B. R. Foorman and M. Hiscock. 'Speech perception and speech production as indicators of reading difficulty'. In: *Annals of Dyslexia* 47.1 (1997), pp. 1–27. DOI: 10.1007/s11881-997-0018-6.

[28] F. Ramus et al. 'Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults'. In: *Brain* 126.4 (2003). DOI: 10.1093/brain/awg076.

[29] React. *React*. URL: https://react.dev/ (visited on 15th May 2023).

[30] J. Read, S. MacFarlane and C. Casey. 'Endurability, Engagement and Expectations: Measuring Childrens Fun'. In: *Interaction Design and Children* (2009).

[31] R. M. Ryan and E. L. Deci. 'Intrinsic and extrinsic motivations: Classic definitions and New Directions'. In: *Contemporary Educational Psychology* 25.1 (2000). DOI: 10.1006/ceps.1999.1020.

[32] R. M. Ryan and E. L. Deci. 'Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.' In: *American Psychologist* 55.1 (2000). DOI: 10.1037/0003-066x.55.1.68.

[33] P. Sanghvi. 'Piaget's theory of cognitive development: A Review'. In: *Indian Journal of Mental Health* 7.2 (2020), p. 90. DOI: 10.30877/ijmh.7.2.2020.90-96.

[34] C. Sansone and J. M. Harackiewicz. 'Intrinsic and extrinsic motivation: The search for optimal motivation and performance'. In: *Academic Press* (2000).

[35] H. Sigmundsson. 'Do visual processing deficits cause problem on response time task for dyslexics?' In: *Brain and Cognition* 58.2 (2005), pp. 213–216. DOI: 10.1016/j.bandc.2004.11.007.

[36] H. Sigmundsson, S. K. Anholt and J. B. Talcott. 'Are poor mathematics skills associated with visual deficits in temporal processing?' In: *Neuroscience Letters* 469 (2010), pp. 248–250. DOI: 10.1016/j.neulet.2009.12.005.

[37] H. Sigmundsson, P. C. Hansen and J. B. Talcott. 'Do 'clumsy' children have visual deficits'. In: *Behavioural Brain Research* 139 (2003), pp. 123–129. DOI: 10.1016/s0166-4328(02)00110-9.

[38] Sikt. *Meldeskjema for personopplysninger i forskning*. URL: https://sikt.no/fylle-ut-meldeskjema-personopplysninger (visited on 3rd May 2023).

[39] B. C. Skottun. 'The magnocellular deficit theory of dyslexia: the evidence from contrast sensitivity'. In: *Vision Research* 40 (2000), pp. 111–127. DOI: 10.1016/S0042-6989(99)00170-4.

[40] B. C. Skottun and J. R. Skoyles. 'Some remarks on the use of motion VEPs to assess magnocellular sensitivity'. In: *Clinical neurophysiology* 115 (2005). DOI: 10.1016/j.clinph.2004.06.021.

[41] J. Skoyles and B. C. Skottun. 'On the prevalence of magnocellular deficits in the visual system of non-dyslexic individuals'. In: *Brain and Language* 88.1 (2004), pp. 79–82. DOI: 10.1016/s0093-934x(03)00162-7.

[42] K. E. Stanovich. 'Matthew Effects in reading: Some consequences of individual differences in the acquisition of Literacy'. In: *Reading Research Quarterly* 21.4 (1986). DOI: 10.1598/rrq.21.4.1.

[43] J Stein. 'To see but not to read; the magnocellular theory of dyslexia'. In: *Trends in Neurosciences* 20.4 (1997), pp. 147–152. DOI: 10.1016/s0166-2236(96)01005-3.

[44] J. Stein. 'The current status of the magnocellular theory of developmental dyslexia'. In: *Neuropsychologia* 130 (2019), pp. 66–77. DOI: 10.1016/j.neuropsychologia.2018.03.022.

[45] J. Stein. 'The visual basis of reading and reading difficulties'. In: *Frontiers in Neuroscience* 16 (2022). DOI: 10.3389/fnins.2022.1004027.

[46] P. Sweetser and P. Wyeth. 'GameFlow:A Model for Evaluating Player Enjoyment in Games'. In: *Computers in Entertainment* 3.3 (2005), pp. 3–3. DOI: 10.1145/1077246.1077253.

[47] J. M. H. Syvertsen. 'Magno: A Platform for Dyslexia Screening'. Preparatory project for master thesis, Department of Computer Science, NTNU, Trondheim. 2021.

[48] J. M. H. Syvertsen. 'Magno: A Platform for Dyslexia Screening'. Master thesis, Department of Computer Science, NTNU, Trondheim. 2022.

[49] Typescript. *What is TypeScript?* URL: https://www.typescriptlang.org/ (visited on 12th May 2023).

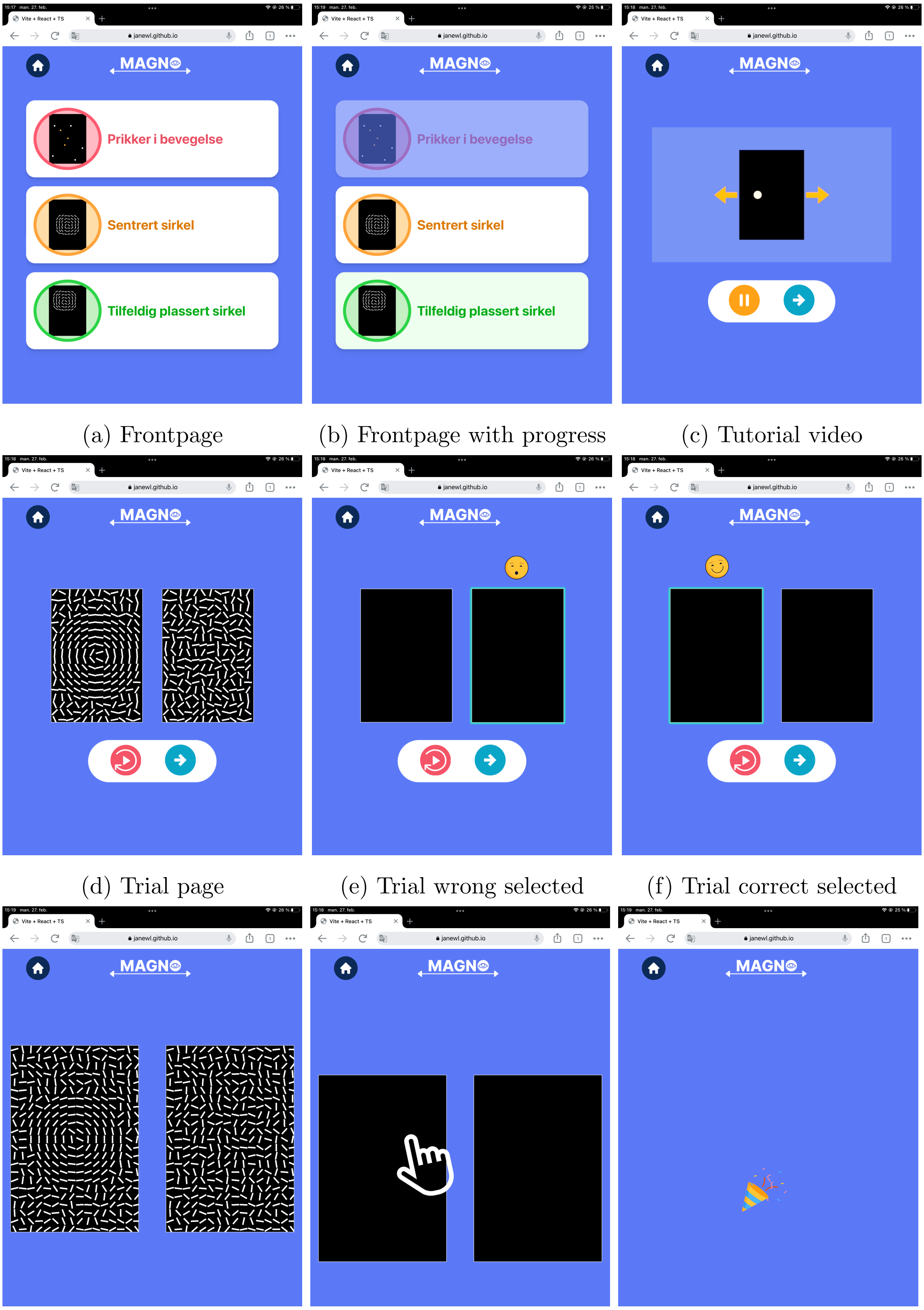
[50] Utdanningsdirektoratet. *Lese- og skrivevansker.* URL: https://www.udir.no/laring-og-trivsel/spesialpedagogikk/spesialpedagogiske-fagomrader/lese-og-skrivevansker (visited on 19th Apr. 2023).

[51] J. Verhaegh et al. 'In-game assessment and training of nonverbal cognitive skills using Tagtiles'. In: *Personal and Ubiquitous Computing* 17.8 (2012). DOI: 10.1007/s00779-012-0527-0.

[52] K. Wiley, S. Vedress and R. L. Mandryk. ‘How points and theme affect performance and experience in a gamified cognitive task’. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020). DOI: 10.1145/3313831.3376697.

[53] B. H. Wold. ‘App for Early Detection of Dyslexia’. Master thesis, Department of Computer Science, NTNU, Trondheim. 2016.

A Design iteration 1



(a) Frontpage (b) Frontpage with progress (c) Tutorial video

(d) Trial page (e) Trial wrong selected (f) Trial correct selected

(g) Test page (h) Prompt choose box (i) Finished test

Figure 16: Complete interface version 1

B Settings.ts file

```
// Patch settings
Settings.PATCH_GAP = 5;
Settings.PATCH_WIDTH = 21;
Settings.PATCH_HEIGHT = 31;

// Dot settings
Settings.DOT_HORIZONTAL_REVERSAL_TIME = 672;
Settings.DOT_RANDOM_DIRECTION_TIME = 672;
Settings.DOT_VELOCITY = 0.08;
Settings.DOT_RADIUS = 7;
Settings.DOT_SPACING = 4;
Settings.DOT_COHERENCE_PERCENTAGE = 50;
Settings.TRIAL_DOT_TOTAL_AMOUNT = 130;
Settings.DOT_TOTAL_AMOUNT = 200;
Settings.DOT_KILL_PERCENTAGE = 10;
Settings.DOT_MAX_ANIMATION_TIME = 15000;
Settings.DOT_MAX_ALIVE_TIME = 50000;

// Form settings
Settings.FORM_AUTO_MODE = true;
Settings.FORM_MAX_AMOUNT = 360;
Settings.FORM_DIAMETER_WB = 17;
Settings.FORM_CIRCLES = 6;
Settings.FORM_CIRCLES_GAP = 8.8;
Settings.FORM_LINE_LENGTH = 1.4;
Settings.FORM_LINE_HEIGHT = 4;
Settings.FORM_LINE_GAP = 0.4;
Settings.FORM_COHERENCY_PERCENTAGE = 100;
Settings.FORM_FIXED_DETECTION_TIME = 30000;
Settings.FORM_RANDOM_DETECTION_TIME = 30000;

// Staircase settings
Settings.STAIRCASE_CORRECT_ANSWER_DB = 1;
Settings.STAIRCASE_WRONG_ANSWER_DB = 3;
Settings.STAIRCASE_MAX_ATTEMPTS = 100;
Settings.STAIRCASE_REVERSAL_POINTS = 6;
Settings.STAIRCASE_REVERSALS_TO_CALCULATE_MEAN = 6;
```

Figure 17: Settings file for motion and form tests

NTNU

Norwegian University of
Science and Technology