DEVELOPMENT OF ADAPTIVE SERIOUS GAMES FOR THE TRAINING OF CHILDREN WITH SPECIFIC LEARNING DISABILITIES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

OĞUZCAN YILDIRIM

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF MULTIMEDIA INFORMATICS

SEPTEMBER 2020

Approval of the thesis:

DEVELOPMENT OF ADAPTIVE SERIOUS GAMES FOR THE TRAINING OF CHILDREN WITH SPECIFIC LEARNING DISABILITIES

Submitted by **OĞUZCAN YILDIRIM** in partial fulfillment of the requirements for the degree of **Master of Science in Modeling and Simulation Department, Middle East Technical University** by,

Date: **10/09/2020**

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : OĞUZCAN YILDIRIM

Signature : _____

ABSTRACT

DEVELOPMENT OF ADAPTIVE SERIOUS GAMES FOR THE TRAINING OF CHILDREN WITH SPECIFIC LEARNING DISABILITIES

Yıldırım, Oğuzcan

MSc., Department of Modelling and Simulation

Supervisor: Asst. Prof. Dr. Elif Sürer

September 2020, 71 pages

Recent developments in technology and the video game industry enabled video games to play essential roles, especially in education, health, and military services under the umbrella term serious games (SGs). Recent studies show that playing computer games is common among children and teenagers, and SGs ease the learning process by providing immersive and interactive platforms. Recently, SGs have been used as an alternative method in education and diagnosis of children with a specific learning difficulty (SpLD). However, these studies have focused on only one or two types of SpLD at the same time. To analyze a broader range of SpLD, in this thesis, five different SGs were designed and developed in order to be used in the training of children with any SpLD. Usability tests were applied to two different participant groups —10 educators and 25 students with SpLD— while they played the games during which think-aloud protocols were implemented. In the second part of this study, a rule-based adaptive difficulty level enhanced the two of the five games, and these modified games were played and tested by the students. The results show that both participant groups had a positive attitude on using SGs together with a complete training tool for students with any type of SpLD. Also, students performed better when they played the adaptive difficulty levels since the games were adjusted automatically due to their performance.

Keywords: Specific Learning Disability, Serious Games, Adaptive Games, Training.

ÖZ

ÖZGÜL ÖĞRENME GÜÇLÜĞÜ OLAN ÇOCUKLARIN EĞİTSEL ALIŞTIRMALARI İÇİN UYARLANABİLİR CİDDİ OYUNLAR GELİŞTİRİLMESİ

Yıldırım, Oğuzcan

Yüksek Lisans, Modelleme ve Simülasyon Anabilim Dalı

Tez Yöneticisi: Dr. Öğr. Üyesi Elif Sürer

Eylül 2020, 71 sayfa

Teknolojideki ve video oyunlarındaki yeni gelişmeler, video oyun sektörünün özellikle eğitim, sağlık ve askeri alanlarda "ciddi oyunlar" adı altında önemli roller almaya başlamasını sağlamıştır. Yapılan çalışmalar, çocuklar ve gençler arasında bilgisayar oyunu oynamanın ne kadar yaygın olduğunu ve ciddi oyunların öğrenme sürecini sürükleyici ve etkileşimli platform oluşturarak kolaylaştırdığını ortaya çıkarmaktadır. Son zamanlarda ciddi oyunlar özgül öğrenme güçlüğüne (ÖÖG) sahip çocukların teşhisinde ve eğitiminde kullanılmaya başlanmıştır. Ancak bu çalışmalar sadece bir veya iki tip ÖÖG'ye aynı anda odaklanmaktadır. ÖÖG'yi daha geniş açıdan ele alabilmek için bu tezde beş farklı ciddi oyun her tür ÖÖG'ye sahip çocuğun eğitiminde kullanabilmek için tasarlanmış ve geliştirilmiştir. İki ayrı katılımcı grubuna —10 eğitimci ve 25 öğrenci— kullanılabilirlik testleri ve oynama esnasında sesli düşünme metotları uygulanmıştır. Çalışmanın ikinci kısmında, beş oyun arasından seçilen iki oyuna kural tabanlı uyarlanabilir zorluk seviyesi eklenmiş ve öğrenciler tarafından test edilmiştir. Sonuçlar iki katılımcı grubunun da ciddi oyunların bütünsel bir eğitsel alıştırma materyali olarak herhangi bir tür ÖÖG'ye sahip çocukların eğitiminde kullanılmasına olumlu baktığını göstermiştir. Ayrıca, öğrenciler oyunların otomatik olarak kendi performanslarına göre seviye değiştirdiği uyarlanabilir zorluk seviyesinde oyunları oynarken daha iyi sonuçlar elde etmişlerdir.

Anahtar Sözcükler: Özgül Öğrenme Güçlüğü, Ciddi Oyunlar, Uyarlanabilir Oyunlar, Alıştırma.

To My Family

ACKNOWLEDGMENTS

First of all, I would like to thank my thesis advisor Asst. Prof. Dr. Elif Sürer for her tremendous help, support, and encouragement during these three years. She always welcomed me with her endless patience whenever I encounter a problem with any subject during this period. Her faith in me made everything easier.

I would like to thank DİSLEV Türkiye Disleksili Çocuklar Vakfı (Dislev Turkish Association of Children with Dyslexia) and especially Mr. Atıf Tokar for their immense help during data acquisition.

I would like to thank my friend Büşra Şenderin for our invaluable brainstorming sessions. Also, I would like to thank my friend Burak Güneş Özgüney for his contribution to the development of one of the games.

I would like to thank my thesis examining committee members for accepting to be committee members of my thesis and their invaluable feedback.

I would like to thank my family for their love and support throughout my life. None of this would be possible without their support and encouragement.

In addition to them, I would like to thank my friends, colleagues, and bosses for their understanding and support during this period.

I would like to appreciate TUBITAK (2210- National Scholarship Programme for MSc Students) for their financial support.

I finally would like to thank the special educators of Albatros Special Education Center (Albatros Özel Eğitim Merkezi, in Turkish) and students who participated in our study for playing the games and providing invaluable help and feedback.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
ADHD	Attention Deficit Hyperactivity Disorder
IDEA	Individuals with Disabilities Education Act
LDA	Learning Disabilities Association of America
SG	Serious Games
SpLD	Specific Learning Difficulty
SUS	System Usability Scale
TAM	Technology Acceptance Model

CHAPTER 1

INTRODUCTION

Special education is a personalized education type according to the needs of children with any kind of disability. Special education can be different for every child according to the type or degree of the disorder. It may include indoor or outdoor activities, one-to-one teaching, and a technological environment in the classroom or at home. In the United States of America, special education is provided in public schools, and this right is protected by laws, such as The Individuals with Disabilities Education Act (IDEA). Disabilities, which are included in IDEA, are categorized into 13 groups, while autism, deafness, attention deficit hyperactivity disorder (ADHD), and specific learning disability (SpLD) are some of the disabilities which require special education according to IDEA.

This thesis focuses on the usability and effects of serious games in the training of children with SpLD. Also, the design criteria for serious games in the literature were examined and discussed. In the first part of the thesis, five games were designed and developed for this purpose. In the second part of the thesis, two of the five games were enhanced with an adaptive difficulty system to improve the experience of children while playing the games.

All five games were tested by children with SpLD and their educators, and their answers for questionnaires – usability and technology acceptance model (TAM) – and comments on games were examined.

1.1. Purpose of the Study

The main objectives of this thesis are listed as:

- Design and develop a group of serious games for the training of children with any type of SpLD,

- Examine, implement, and evaluate the design criteria of educational serious games in the literature,

- Evaluate the effect of adaptive difficulty balancing by checking the performance of the players and collecting their feedback on the standard and balanced versions of the games,

- Test the games with both the educators and students with SpLD by evaluating their experiences in terms of usability and technology acceptance perspectives while also collecting their feedback via open-ended questionnaires.

1.2. The Goal of the Thesis

The main goal of this thesis is to develop a group of games that are specifically designed to be used in the special education of children with any type of SpLD. In the thesis, the potential of this collection of games has been investigated with usability tests. Another goal is showing the effect of the adaptive difficulty system on the playing experience of children with SpLD. An adaptive difficulty system has been implemented to measure its impact on the performance of the children with SpLD.

1.3. Developed Applications and Serious Games

For this study, five different serious games were designed and developed, which were aimed to be used for training children with any type of SpLD. The first game is named "Word Game," which is a simple word-forming game with given letters, according to a question asked by a child avatar. The second game is "Memory Game," where the player needs to memorize a given pattern in five seconds and recreate it after placing the pattern to the same grids on the screen. The third game is "Category Game," which includes two sections: At first, the player needs to mark words on the screen if they are suitable for a given category. In the second part, the player writes down other examples of given categories without any given clue. "Space Game" is the fourth game, which is a fast-paced action game. Players collect gems on the screen by flying through them while avoiding crashing enemy spaceships. The final game is "Math Game," which is based on doing simple calculations. The player taps on one of the four kites on the screen, which have possible answers written on them, before they fly out of the screen.

After the development process, the games were tested twice by two participant groups, including educators and students with SpLD between the ages of 7 and 11. During the first interview with the participant groups, the System Usability Scale (SUS) and Technology Acceptance Model (TAM) questionnaires were filled by the participants besides applying the think-aloud method during playing and answering open-ended questions after playing the games. After the first interviews, rule-based adaptive difficulty levels were added to two of the five games to reveal the effect of adaptiveness in the developed serious games. During the second interview with the children, only the open-ended questions and the think-aloud methods were used. For the educators, the same interview progress was used as before.

1.4. Contributions and Novelties

The main contributions of this thesis are as follows:

- Five different serious games were designed and developed to be used in the training of children with SpLD. These games allow educators to use a tablet while training children during the lessons.

- Two of the games were enhanced with an adaptive difficulty system, and its impact on the usability and performance was analyzed.

- Four interviews were held to evaluate the developed games in six months. In these interviews, target audiences of the games —students with SpLD and their educators— tested and commented on the games. Educators examined the first and the last versions of the games and answered usability, technology acceptance questionnaires, and open-ended questions about the games twice. Students played all the games and evaluated them with the same questionnaires. Also, they played an adaptive difficulty version of the games in another interview and answered the questions about how they feel for both versions of the games.

- Recently, serious games in the education of children with SpLD have become an appealing subject for researchers, and to the authors' best knowledge, this thesis is the first comprehensive study that deals with all types of SpLD entirely and evaluates the effect of adaptive difficulty system using both quantitative and qualitative methods.

1.5. The Outline of the Thesis

The structure of this thesis is as follows: in Chapter 2, background information about the SpLD, serious games, adaptiveness in serious games, and evaluation procedures are given. In Chapter 3, proposed games are examined, with their full design criteria and development process being explained in detail. In the following section, Chapter 4, details of the adaptive difficulty adjustment are presented. In Chapter 5, the results of the experiments are given, and these results are discussed comprehensively in Chapter 7. In Chapter 8, the final chapter, a summary of the study is presented, and possible future improvements are mentioned.

CHAPTER 2

RELATED WORK

2.1 Types of Specific Learning Disabilities

Specific learning disability (or specific learning disorder, SpLD) is a disorder that causes children to learn particular topics slower than expected from their age or educational level. SpLD is not synonymous with the Learning disability (or learning disorder, LD), but it is a medical term that is used for diagnosis and often referred to as LD. To prevent this confusion, the term of SpLD will be used for the rest of this thesis.

SpLD is not related to the intelligence level —i.e., children with SpLD can have average or superior intelligence levels, but their learning speed can be slower than their peers. Learning Disabilities Association of America (LDA) classifies SpLD under five categories as Dyslexia (reading disability), Dyscalculia (math disability), Dysgraphia (writing disability), Oral/Written Language Disorder and Specific Reading Comprehension Deficit, and finally, Non-Verbal Learning Disability.

Dyslexia, the most known SpLD type, causes spelling errors, slow and false reading in children. Children with Dyscalculia have difficulty with numbers and mathematical concepts, which include simple numerical operations and counting. Dysgraphia leads to writing problems on children, including misspelling, writing letters in incorrect shapes, and writing letters in different sizes in a word. Children with Oral/Written Language Disorder and Specific Reading Comprehension Deficit have problems with using the language for expressing themselves or understanding sentences. They usually cannot understand the concept of a sentence form, which causes forming false sentences. For example, they cannot understand the difference between the sentences “Soup is in a spoon” and “Spoon is in a soup.” Finally, Non-Verbal Learning Disability, which is not counted as a type of SpLD by IDEA, causes children to struggle with understanding facial expressions or body language. Dyspraxia is another disorder that is usually associated with SpLD types. Children with Dyspraxia could show different symptoms about motor skills, from not adjusting coordination between lips and tongue to handwriting. There is no consensus on types of SpLD and coverage of SpLD changes in every country. Therefore, Dyslexia, Dysgraphia, Dyscalculia, and Dyspraxia are accepted as SpLD types in this thesis since they are accepted as SpLD in Turkey and are suggested by educators of special education center that we have collaborated with. There are 41.600 SpLD diagnosed people

in Turkey, where 82% of dyslexic people drop out without even completing secondary education.

There is no known treatment for SpLD, except for special education. Children with SpLD take extra courses in groups or individually given by Special Education teachers, besides their regular education in school. This special education may include reading, writing, mathematics, improving motor coordination skills, and language education exercises. Moreover, attention and memory exercises can be a part of the special education if children have any other disabilities, such as ADHD, besides of SpLD. Some researches try to diversify these special education methods. Art therapy can result in a significant pre- and post-change in a positive way among children with SpLD between the ages of 7-15 (Freilich & Shechtman, 2010). In contrast, participants in the control group, who were treated with academic assistance, did not show any statistically significant changes.

2.2 Serious Games

Zyda (2005) defines serious games (SG) as tools that use entertainment for different purposes, such as training, education, and health. Nowadays, serious games have become an alternative way to traditional methods in training, military, education, and health. Since children are usually enthusiastic about games, using games as part of their education or rehabilitation is a motivating factor for them. Therefore, there are a couple of serious game researches for different purposes.

In recent years, the effect of games on the training of children with SpLD has been discovered, and the number of studies on this topic increased rapidly. Studies for SpLD started with the effect of regular video games on children. In a study (Franceschini, Gori, Ruffino, Viola, Molteni & Facoetti, 2013), researchers worked with 20 children with Dyslexia to see video games on reading skills. All the children were individually treated by playing Rayman Raving Rabbids, a commercial Wii video game, for a total of 12 hours with 80 minutes daily period. However, half of the group played mini-games, which were categorized as action video games, while the other half played non-action video games from Rayman Raving Rabbids. Although these two groups had similar results from phonological decoding and reading measurements before the video game training, the group who played action video games had significant improvement at the measurements after play sessions. Moreover, the increase in reading speed occurs without any additional misreading. It is also stated that this progress was equivalent to a one-year study with traditional methods.

An application was designed to detect Dyslexia among children between the ages of 6-7 (Bartolomé, Zorrilla & Zapirain, 2012). The application includes six different mini-games, each of which tests different skills, including word production, syllabic memory capacity, verbal work memory, auditory memory, syllable, and word reading. The application was found useful by teachers for both detection of possible Dyslexic children and correcting exercises. The application was also appreciated by parents since it can also

be used from home. Despite all these positive aspects, the application was language-dependent, and children should know reading and writing since verbal games require these skills.

To solve the first problem with language dependency, another application was developed to screening Dyslexia among children aged between 7 and 12 with language-independent games (Rauschenberger, Rello, Baeza-Yates & Bigham, 2018). The study includes a musical-based game and a visual-based game. Application (MusVis) aims to distinguish Dyslexic children from others according to their reactions to musical and visual cues in the games. They tested the games with 108 Spanish, 57 German students where 67 of them have Dyslexia. The results show that Spanish children with Dyslexia showed different performances in games, and it can be presented with eight significant indicators such as duration, average click time, or time to the first click. However, for German participants, the result was not the same, and there were only four significant indicators that are language independent. These indicators are listed as total clicks, hits, accuracy, and efficiency for the visual game. Indicators of the visual games were found to be language-dependent. Also, due to the design of games, pre-readers were not able to play these games, which prevented prediction among pre-readers. To overcome this problem, the application was modified for pre-readers (Rauschenberger, Rello & Baeza-Yates, 2018). Musical cues in the game completely changed, and the visual game simplified. Games are prepared to be tested for children with different mother-tongue, e.g., German, Spanish, and English.

In another study, researchers developed a system to diagnose developmental Dyslexia in preschoolers (Gaggi et al., 2017). Games include drawing paths, visual search, visual-spatial attention, completing the drawing of letters, and audial letters. Results show that children with familial risk for developmental Dyslexia had different scores than other children who are in no risk group.

Serious Games for dyslexic children are developed not only to diagnose disorders but also to help the education of the children. An application called Madrigale was developed to improve the reading skills of dyslexic children between the ages of 7-9 (Di Tore, Di Tore, Ludovico & Mangione, 2014). The game on the application is based on ordering letters only. Graphical and phonological signs were made clear for this task. Although there was only one game that was designed for this purpose, the difficulty level in the game can be adjusted with six different variables. After player levels up, the number of letters, number of similarly shaped letters, word length, the difficulty of the word, and disarray of letter distribution can be increased while more unknown words are chosen to make the game difficult. These variables increase repeatability since they increase variety.

Another study (Ludovico, Di Tore, Mangione, Di Tore & Corona, 2015), based on Madrigale, aims to measure the reading abilities of dyslexic children. They developed further Madrigale to a hidden object puzzle game. The goal of the game is simply revealing hidden objects in the background by reconstructing syllables in order. They tested the game with 12 students who were fifth grade at that time, and none of them were

diagnosed as Dyslexia. They aimed to find a correlation between in-game activities and reading skills while they also tested to determine possible Dyslexic players. The results show that proposed game activities correlate with the speed of reading while there is no correlation between accuracy and game activities. Since it is hard to evaluate the effect of only one game on different consequences of Dyslexia, researchers also focused on creating systems that include different mini-games.

Types of spelling errors of dyslexic children were determined for both in English and in Spanish in six groups, which are insertion an extra letter, the omission of a letter, substitution of a letter, transposition of letters, word boundary errors, and morphology errors (Rello, Bayarri, Otal & Pielot, 2014). After classification, they developed a mini-game for each spelling error type so that children with Dyslexia could overcome these errors by practicing on games. They tested games with 54 potentially dyslexic children who were selected by their teachers. The participant group was filtered to 48 children between the ages of 6-11 after applying a standard dyslexia diagnosis test in Spanish. The study continued for eight weeks, and 48 children had to play proposed games or a word search puzzle game that served the control group. Results show that the error rate of children with Dyslexia decreased by 20% for a word, which is a significant reduction amount when it is compared to the control group. However, even though students who played the proposed games showed a 9.4% decrease in reading errors, this result was not significant enough, which can be explained with the selection of games was according to spelling errors in writing, not to reading.

Since Dyslexia and Dysgraphia cause different difficulties in different languages and alphabets, SGs which developed for these disorders should be explicitly designed for that language. For instance, a group of games was created for Arabic children who suffer from either Dyslexia or Dysgraphia (El Kah & Lakhouaja, 2018). Since the shape of letters changes according to its position in word in the Arabic language, they divided games into three groups as letter games, word games, and sentence games. Results of a test study with 46 second grades, where 6 of them diagnosed with Dyslexia and 20 of them show the academic delay in the opinion of their teacher, show that games significantly improve performance of children in both reading and spelling according to paired sample T-test results based on pre-test and post-test results. It is also indicated that games increased students' motivation to learn according to their observations.

There are also studies for Dyscalculia. An augmented reality serious game was developed to be used in the education of dyscalculic children (Avila-Pesantez, Vaca-Cardenas, Avila, Padilla & Rivera, 2018). The game includes three different scenarios or gameplays, and each of them has three different difficulty levels. The participant group for this study consists of 40 students between the ages of 7-9. Children attended a total of 8 meetings in a month, where one of them lasted about 15 minutes. Randomly selected half of the children played the proposed game while the other half, the control group, was treated with a traditional method called Domino. Results showed that children who played the proposed game needed significantly less time than the control group while matching basic

mathematical operations to its answer. Also, it is shown that they performed better in given tasks.

2.3 Adaptive Difficulty Balancing

Serious games should be challenging enough to make children get fun out of the game, just like any other commercial game. For this purpose, adaptive difficulty balancing is suggested by researchers. The adaptive game difficulty was implemented balancing on two games where one of them is an existing SG called PEGASO (Personalized Guidance Services for Optimizing Lifestyle in Teenagers) and the other one developed for this study (Hendrix, Bellamy-Wood, McKay, Bloom & Dunwell, 2018). PEGASO is an action-based puzzle game, while the new one is a simple platformer game. According to survey results, for both games, participants were more likely to prefer the balanced version of games. However, participant groups were convenience samples. For PEGASO, 8 participants were selected among friends and families of developers, even though the game was an SG with a focus group. For the platformer game, thirty-seven participants were recruited online.

Moreover, due to design selections, games were adapting themselves only at the start of the game after checking the last performances. This disables the opportunity of level balancing during games. In a study (Cuschieri, Khaled, Farrugia, Martinez & Yannakakis, 2014), researchers designed an open-ended visual novel type game that can be played again for dyslexic kids. The game consists of 9 mini-games which target the linguistic skills of children, which are defined by dyslexia experts. Games are also designed to be played under teacher supervision and alone. They also developed an adaptation mechanism, which includes only a set of pre-defined difficulties decided by experts. The adaptation mechanism adjusts difficulty by checking the game performance of students and time elapsed since last practice. Because of that, it only changes difficulty at the start of the game.

2.4 Design Rules of Educational Serious Games

Studies are not only based on developing an SG for children with SpLD but also specifying designing rules of serious games. Six terms are defined to be considered while designing a serious educational game (Annetta, 2010). The first of the term is called “Identity,” defined as the behavior of players seeing themselves as protagonists in the game. “Immersion” is the second term and is used for being immersed by the content and mechanics of the game. The third term, ‘Interactivity, refers to the social interactions of players with other players or non-playable characters (NPC). Well-designed competence in games is called “Increasing Complexity.” The fifth and sixth terms are “Informed Teaching” and “Instructional,” which indicate a serious educational game should be informative, and this could be tested.

In a study (Bellotti et al., 2011), it is reported that researchers have determined six categories of design patterns for SGs. The first one is named "Integration," which is used for using game elements and adding learning objectives in pedagogically. "Cognition" is keeping the player focused on the game with feedback. The third term, "Presentation," is being sure that content passes to players. "Social Interaction" and "Teaching" are the fourth and fifth terms, which means teaching through social activities and offering solutions that facilitate learning. The final term, "Engagement," is called for experiencing a game as a whole experience. In another study (Shabbir & Bhatti, 2019), six user interface design rules were described to keep in mind while developing a serious game for dyslexic children. According to them, the user interface should not be overloaded with information, words in the layout should not include any jargon, the user interface should be consistent, instant alert messages should be avoided, graphics and images should be purposeful in layout, and also background should be plain and light-colored.

2.5 Evaluating the Performance Outcomes of Serious Games

Designing and developing games or tools is not enough to accept that they are useful and ready to use. They should be checked if they can serve their purposes. To evaluate that, various evaluation methodologies (see (Calderón & Ruiz, 2015; Mayer et al., 2014; Moreno-Ger, Torrente, Hsieh & Lester, 2012; Bernhaupt, 2015)) are studied by researchers. Calderon & Ruiz (2015) analyzed 102 serious game publications, and they indicated that 92 of them used questionnaires as the main evaluation method. The following subsections include an explanation of two commonly used questionnaires, which were also used as evaluation tools in this thesis.

2.5.1 Technology Acceptance Model Questionnaire

Information system research based on system use prediction at the beginning of the 1980s, since there was no commonly held study that investigates the user's acceptance of newly introduced technologies. Fred Davis (2016) suggested a model named Technology Acceptance Model (TAM), which is a questionnaire focused on perceived usefulness and perceived ease of use. Even today, TAM is a widely used model to evaluate the acceptance of users. Vanketash & Davis (1996) completed TAM to its known version today. Since then, Vanketash & Davis (2000) proposed and published TAM 2, and Vanketash & Bala (200) proposed another alternative, TAM 3.

2.5.2 System Usability Scale Questionnaire

Usability is one of the most important topics when it comes to the evaluation of educational serious games. Since games will be used in education or training of children, they should be accepted as usable by not only students but also their educators.

For this thesis, the System Usability Scale (SUS), which was introduced by Brooke (1986), was used to measure the usability of the games. Evaluation of usability has been tested with SUS in many different technological systems, such as websites, applications, and games. Bangor, Kortum & Miller (2008) indicated that the reliability of SUS was high for different applications after they analyzed more than 2000 SUS surveys.

CHAPTER 3

SERIOUS GAMES DEVELOPED FOR SPECIFIC LEARNING DIFFICULTY

In this chapter, five serious games that were presented for the training of children with SpLD are presented in detail.

The games were developed in Unity (version 2018.3.0f2), which is a cross-platform game engine that allows its users to develop their games in both two-dimensional (2D) and three-dimensional (3D) settings. Unity also enables building applications for Android devices so that an ASUS Zenpad 8 tablet is used during the data acquisition.

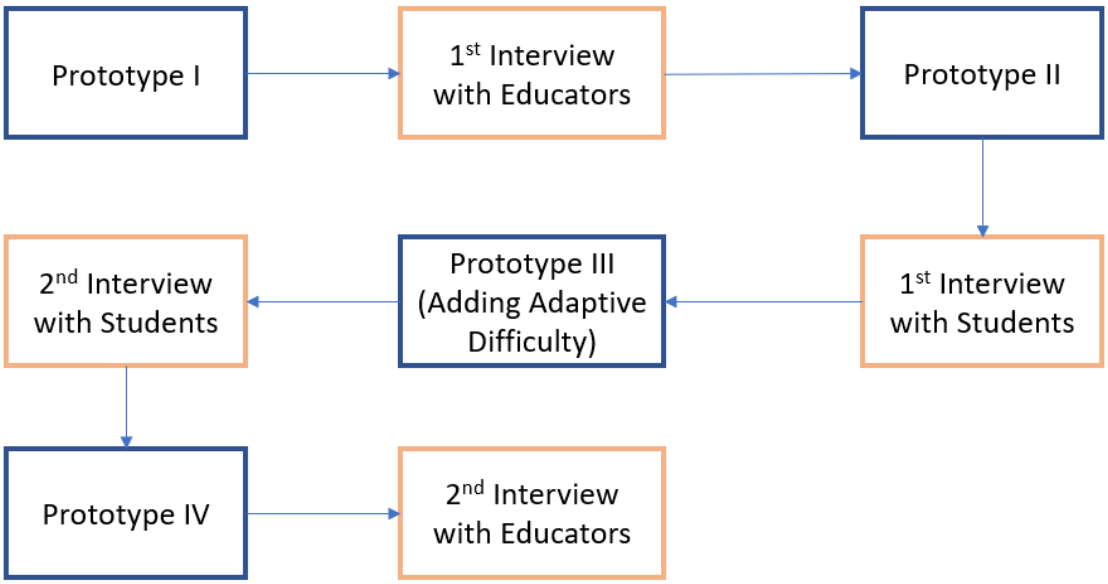


Figure 1: The workflow diagram of this study.

3.1.Design Criteria

At first, design criteria were defined to eliminate future problems with incompatibility between children's educational background and games with the help of special educators. The first criterion is about the target groups of the games, which include children who know reading and writing, with specific learning disabilities (SpLD). Also, the age of children is limited to 11, since designing a game that is appropriate for both new readers and secondary school students could be an unrealistic goal even there will be different difficulty levels. The second criterion is that games should be easy-to-use and understand due to the struggle of children with SpLD. To make that, games should not include two or more independent tasks such as picking up stars in the sky while solving multiplication problems on the ground.

Moreover, the games should not include any distracting elements which can be visual or auditory because of children's low attention level. To encourage children during gameplay, the games should have positive feedback and almost non-negative feedback. Finally, games are designed for children who will play under the supervision of their teachers or parents since these games are thought of as supportive educational material.

All these criteria should be taken into consideration for every game regardless of the theme or mechanics of the games.

3.2.Themes of the Games

The themes of the games should be decided very carefully due to the fact that games will be played by children with SpLD. Games should not focus on a single type of SpLD but should cover a variety of different types of SpLD. The first game's theme is decided to be the spelling for children with Dyslexia and/or Dysgraphia. Simple mathematical operations are the second game's theme for children with Dyscalculia. The third game's theme is writing examples of specific categories, which are chosen for children with Dysgraphia and/or Dyslexia. Also, two more themes are added for different purposes — i.e., a memory game targeting to improve visual memory and a space game requiring time management and constant concentration.

While choosing themes, opinions of educators were also taken. They showed what kind of practices they do in special education and suggested game ideas.

Limitations of the number of themes were due to the time limitations of the children. The children had only 30 minutes per week, and they had to play all the games and answer the study questions in two weeks. After deciding these five themes, other possible game ideas were discarded, and the development procedure of these five games started.

3.3.Developed Games – Prototype I

Four different versions of the five games are used during the study.

The first version of the games was unfinalized and was developed as prototypes, but all the themes and rules were well-defined at that stage. Educators examined the unfinished version of the games in order not to lose any time for developing ineffective games so that the first interview with the educators can be considered as a quality check of the first prototype (Prototype I).

3.3.1 Word Game

The Word game is a game focusing on spelling for children with Dyslexia and/or Dysgraphia. In the game, a child character asks ten questions about a single topic that is pre-defined. Answers were limited to four to six letters in Turkish. The player needs to drag and drop the letters to the appropriate yellow squares, which represents a letter in the answer. If the letter is correct, it replaces the yellow square, and if it is a false letter, the letter does not fill the yellow square and goes back to its initial position at the bottom left corner. After entering all the correct letters, the child avatar says, “It is correct!” and asks the next question.

The background image of the game was chosen according to the topic to keep the players focused on the topic. For instance, “Kitchen” was the topic of the first designed level, and the players were asked to answer questions about cutlery or fruits, such as “What is the name of the sweet fruit which has red, yellow, and green kinds?” and “What do you use to eat your soup?”. The player gets 10 points for each correct answer.

Figure 2: Screenshot of the Word game from the Prototype I.

3.3.2 Memory Game

The Memory Game was created to improve the visual memory and attention of children. In the game, there are two panels; on the left side of the screen, there are four colors (or four arrows each faced to another direction), and a counter. On the right side of the screen, nine boxes are aligned as a three by three grid.

When the game starts, three different or the same colors appear on three of the nine boxes randomly, and the counter starts to go down. After the counter hits zero, from five, colors on the right side disappear. The goal of the player is to refill the three boxes with correct colors as quickly as possible. To do that, the player first needs to tap the correct color on the left panel and then tap the correct grid on the right panel. To erase the grids, the player can tap on a filled grid. Whenever the player fills all the three grids correctly, "CORRECT!" feedback shows up at the counter's position. After three seconds, a new, randomly generated memory question is asked. For each correct answer, the player gets 10 points.

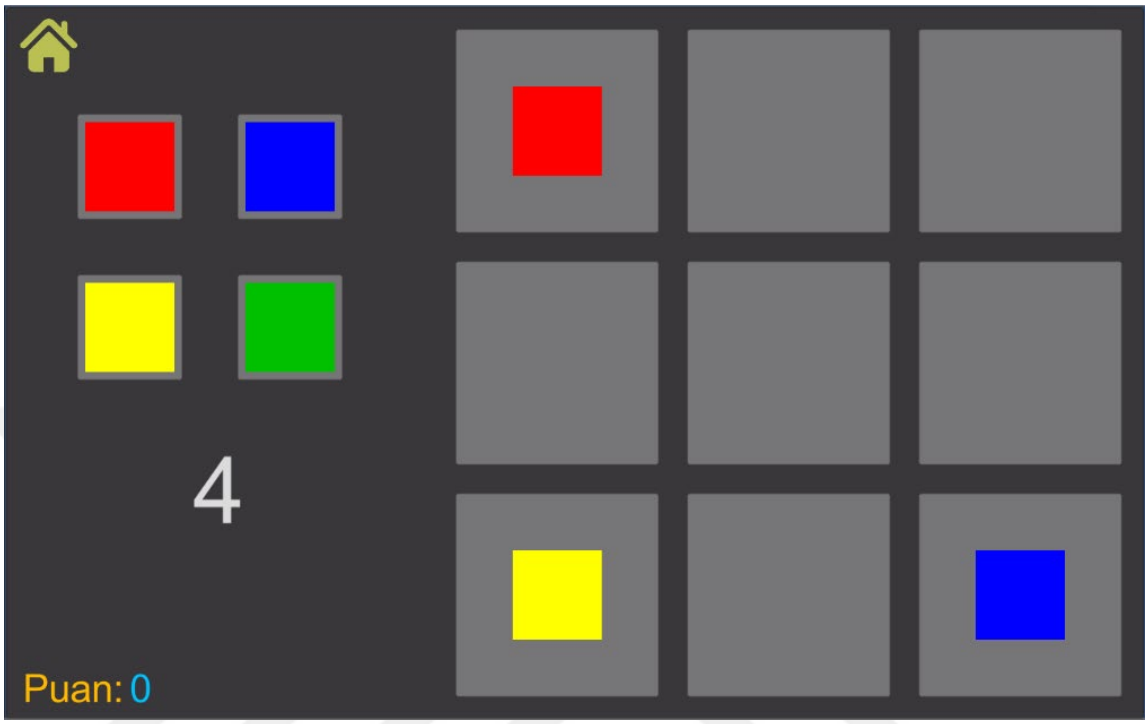


Figure 3: Screenshot of the Memory game from the Prototype I.

3.3.3 Category Game

The Category game is developed in order to improve thinking and writing —not handwriting but correctly ordering letters in a word— skills of the children with Dysgraphia and/or Dyslexia.

During the game, a couple of categories are given, and the players are expected to find words related to that category. For example, in Figure 4, the category is "Things that we closed," and pre-defined words are "Window," "Door," "Box," etc. In each category, there are 15 to 20 pre-defined answers. If the player writes down one of them, it appears on the screen with a random color and position, and the player gains 10 points. If the answer is wrong or if it is misspelled, the word does not appear on the screen.

The player uses the keyboard of the tablet in order to write the words. If the player gets stuck at some point, the skip button on the top right corner can be used to pass on to the new category. There were ten different categories in the first version of the game.

Figure 4: Screenshot of the Category game from the Prototype I.

3.3.4 Space Game

The main intention while designing the Space game is to test time management and concentration level of children with a theme that is not causally related to their lessons. The main goal of the player is collecting gems, which is steady while trying not to hit small spaceships that are programmed to go straight into a random direction.

The initially available time slot is 90 seconds, and the timer counts down. Whenever the player taps into the screen, a spaceship rotates into that point and moves towards the point until it reaches the tapped position. A gem appears in every five seconds at a random place. If the player cannot collect the gem in ten seconds, the gem disappears. For every 0.4 seconds, an enemy spaceship spawns off-screen and flies straight through the screen in a random direction. The player gains two, four, or six seconds randomly, after collecting a gem. Whenever a player collides with another spaceship, the player loses three, five, or seven seconds randomly. The game is over when the time is up.

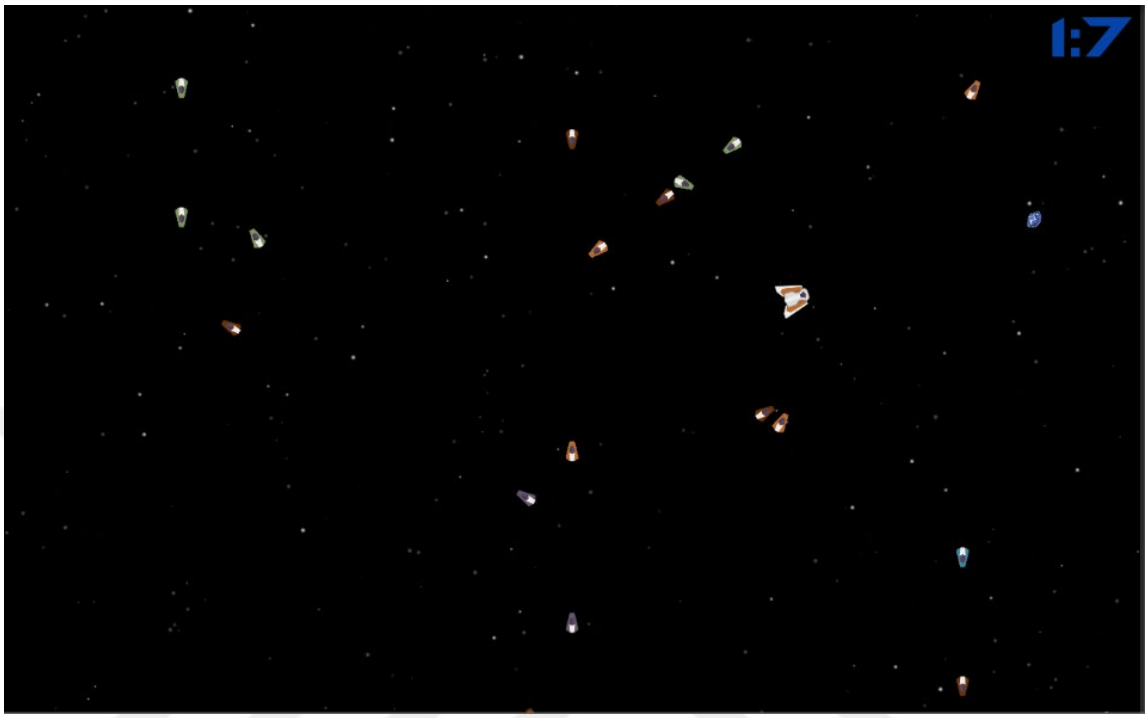


Figure 5: Screenshot of the Space game from the Prototype I.

3.3.5 Math Game

Math Game is a simple addition and subtraction game. In the first version, the problem appears below the screen, and four kites float up in the air. Problems are created randomly with a set of selected numbers between 0 to 15. The main goal of the player is tapping on a kite, which contains the correct answer before it exits from the top of the screen. There was no penalty for wrong answers to see the tapping strategy of the children. They could tap all four to see the next question and still get full points —i.e., 10 points.

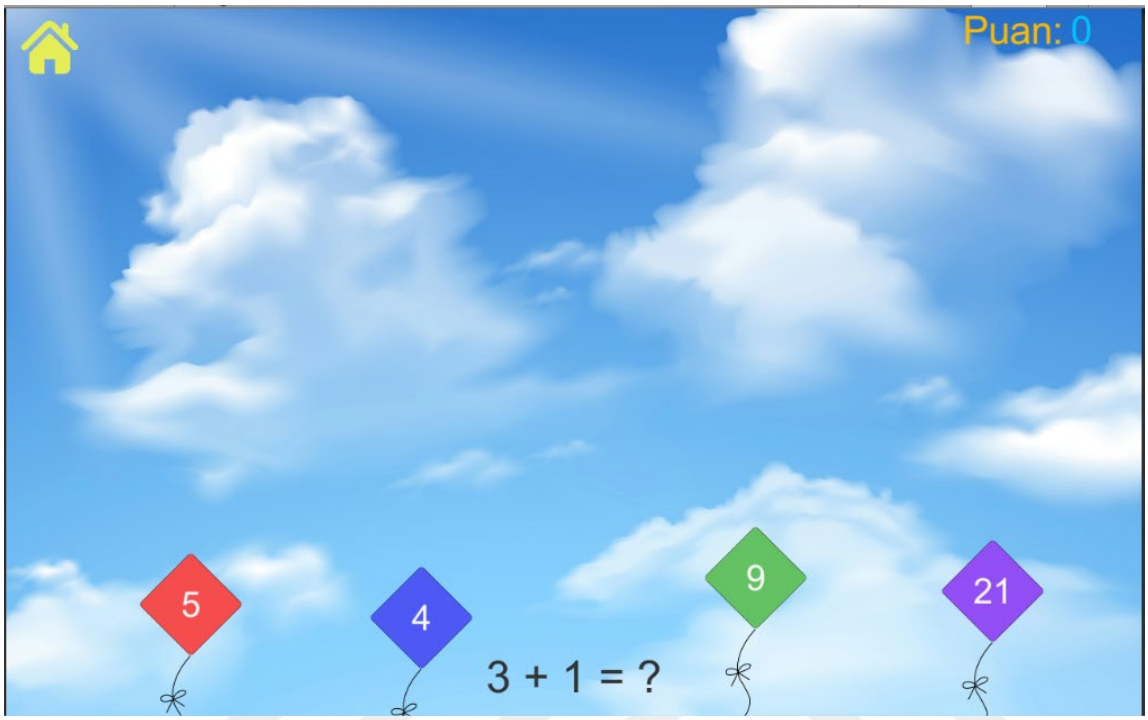


Figure 6: Screenshot of the Math game from the Prototype I.

3.4.Developed Games – Prototype II

After the first interview with the educators, the games were finalized with a couple of changes, and a local database was added to record the gameplay statistics of the players.

3.4.1. Word Game

A skip button was added to the screen for being able to pass unsolved questions. In the database, the number of correct answers, the number of false letters, the number of skipped questions, and the total time were recorded.

3.4.2. Memory Game

A hint button was added to the screen to show the colors or arrows once again to the players. Also, a timer was added to display the countdown. Both the difficulty and timer can be set just before the start of the game by the educators or parents. The difficulty setting affects the number of grids to be remembered during the game. The number of correct answers, number of chosen false grids (i.e., filling a grid which should be empty), number of grids which were filled with false-color or direction, difficulty level, and session time were recorded in the database of the Memory game.

3.4.3. Category Game

The category game was separated into two parts: In the first part, the player must choose five correct words about a given category, between 20 available options. After the fifth correct answer, the category changes. When five categories are completed, the second part

starts. This part is the same as the previous Category game. The player gives examples of the same categories in the first part. Moreover, a skip button is added to the screen.

Figure 7: Screenshot of the first part of the Category game from Prototype II.

In the database, the number of correct answers, the number of false answers, and session time were saved for the first part of the game. For the second part, the number of correct answers, number of typos, number of false words, which are irrelevant to the given topic, and session time were recorded.

3.4.4. Space Game

No change was applied in the Space game except adding the database for statistics. The number of collected gems, the number of crashes with the enemy, and session time were adjusted to be kept in the database.

3.4.5. Math Game

A timer was added as a countdown, and both the timer and difficulty could be set before the game starts. The difficulty level changed the speed of the kites, and a maximum limit of numbers for the math operations could also be set. The number of correct answers, the number of false solutions, the number of unanswered questions, session time, and difficulty level were kept in the database.

Figure 8: Screenshot of the second part of the Category game from Prototype II.

3.5.Developed Games – the Third Prototype

After analyzing both first interviews with the students and educators, two games — Memory game and Space game— were updated and enhanced with the adaptive difficulty level. The design or the mechanics of the games did not change since the purpose of the second interview with the students was to find out the behavior and success of the children in games that have adaptive difficulty levels. The analysis of the interviews is presented thoroughly in the Results section.

3.6.Developed Games - Prototype Four

The Prototype four is the final version of games, which were developed according to the feedback of both participant groups, and the observation of the children while playing the games and the statistics of the children. As it was mentioned before, all the analyses of these statistics and feedback will be presented in the Results section of this thesis.

3.6.1. Word Game

No change was applied to this game.

3.6.2. Memory Game

To represent the alternative versions of this game, a version with pictures was made. In this version, sky objects were used, such as Sun, Moon, cloud, and lightning. Moreover,

the appearing time can be set as 3, 4, or 5 seconds before the game starts —which also increases and varies the difficulty. Thus, this game has four versions (colors, arrows, letters, and pictures), five difficulty options —one grid, two grids, three grids, four grids, and adaptive (two grids with variable appearing time)— with three appearing time options (3, 4 and 5 seconds) and an adjustable timer.

3.6.3. Category Game

In this game, the input system of the open-ended questions was changed. The player does not use the keyboard of the tablet keyboard anymore but uses an on-screen keyboard aligned in alphabetical order.

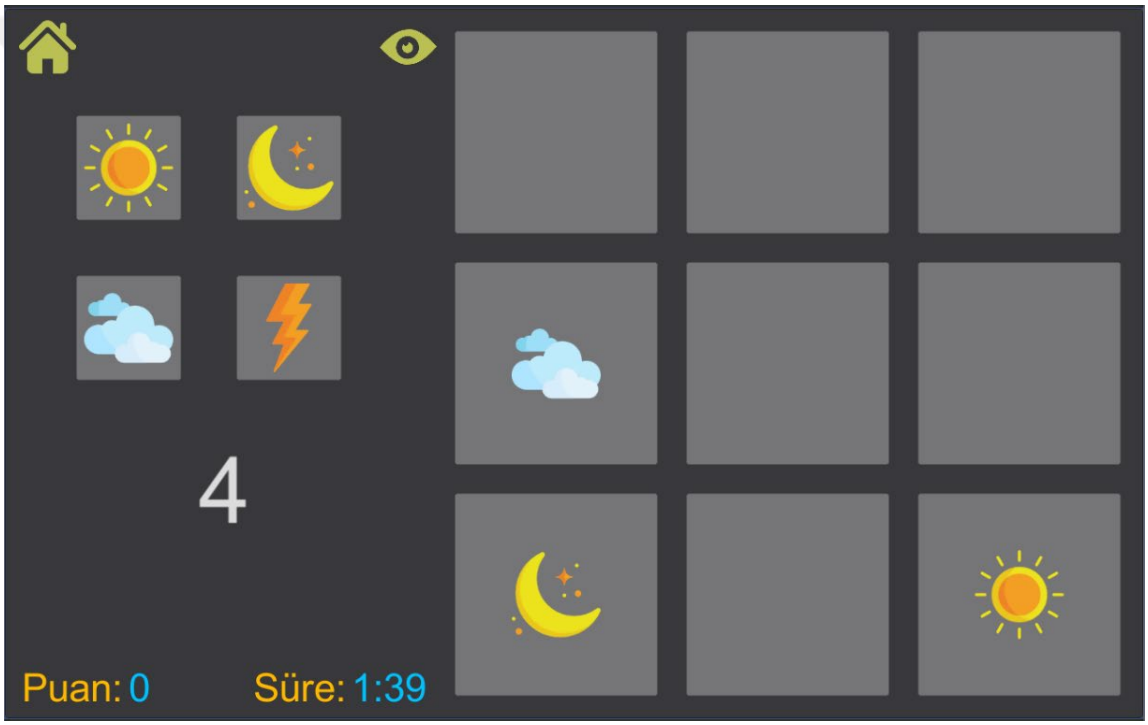


Figure 9: Screenshot of the picture version of the Memory game from Prototype IV.

3.6.4. Space Game

All the objects in the scene, except UI, were enlarged. The Control mechanics were changed with a joystick, which was added to the bottom right of the screen. Also, the default enemy spawn frequency was decreased. The waiting time between two enemies spawn was changed to 0.7 seconds. Moreover, visual feedback on collecting gems and crashing into enemies was added. Players can see how much time they gained (in blue) or lost (in red) during the game.

3.6.5. Math Game

Multiplication operation was added to the game as an option. The scoring system was changed so that the player does not gain full points after the first guess anymore. In the

second guess, 8 points, in the third guess 4 points, at the last guess, only 1 point is given to the player.

Figure 10: Screenshot of the second part of the Category game with an alternative keyboard layout from the Prototype IV.

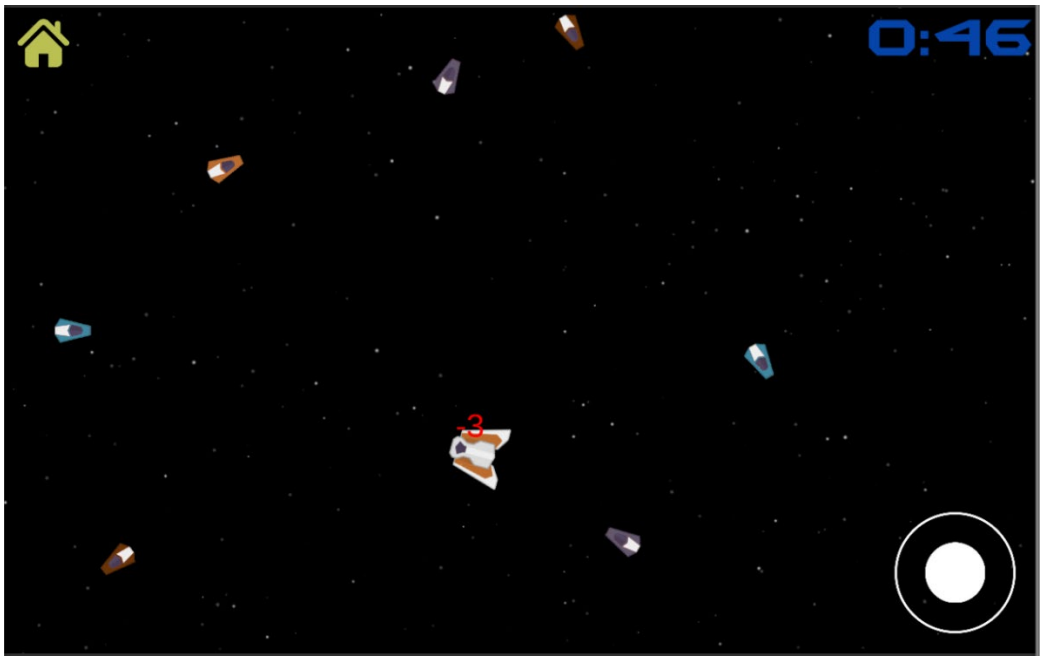


Figure 11: Screenshot of the Space game with an alternative control system from Prototype IV.

CHAPTER 4

ADAPTIVE DIFFICULTY

4.1 Adaptive Difficulty Mechanism

The adaptive difficulty is designed to check consecutive successful or unsuccessful moves by the player and adjust the game difficulty according to the gameplay level of the player. A simple adaptive difficulty system was developed for the Memory game and Space game. The number of games decreased to two since these games will be played by students twice, and they had time constraints. These games were selected rather than the other three because these games had repeatability, and they were liked by children the most. This adaptive difficulty is added to these games to keep the players in 'golden path' or 'optimal gameplay corridor,' which prevents the players from getting bored or frustrated by the difficulty of the game as it is rooted in the Flow study of Csikszentmihalyi (1990) and enhances in Thomas & Young (2010) (Figure 12).

For the adaptive difficulty, a rule-based method was selected due to the time constraints of the students. They could only play both games twice, once each for the standard and adaptive versions. Since repetitive testing is no longer an option, methods that use training to give a result could not be used. Also, Bayesian Networks could not be used for a couple of reasons. First, datasets of the first interview were too small to determine more than two inputs, achieving the goal —giving the correct answer or collecting the gems— and making mistakes —giving false answers or crashing into enemies— for adaptive difficulty. The number of touches to the screen or the time between two touch inputs could be variables to change the difficulty, but without training and with small datasets defining probabilities to these actions could not cause meaningful outcomes. Secondly, since there will be only two inputs and no training for the Bayesian Network, a well-defined ruled-based method could also give similar results. In other words, the rule-based difficulty method is chosen to change the difficulty certainly for a couple of actions, instead of for every single action.

These two alternatives would give the same results for large sample sizes in theory; however, the rule-based method is chosen to make sure every student gets the same results for their successful or unsuccessful set of actions. Also, since there will be only one gameplay session per student, the adaptive difficulty should affect the game while the

game is still continuing, unlike the study of Hendrix, Bellamy-Wood, McKay, Bloom & Dunwell (2018), where players' previous game sessions change the game difficulty.

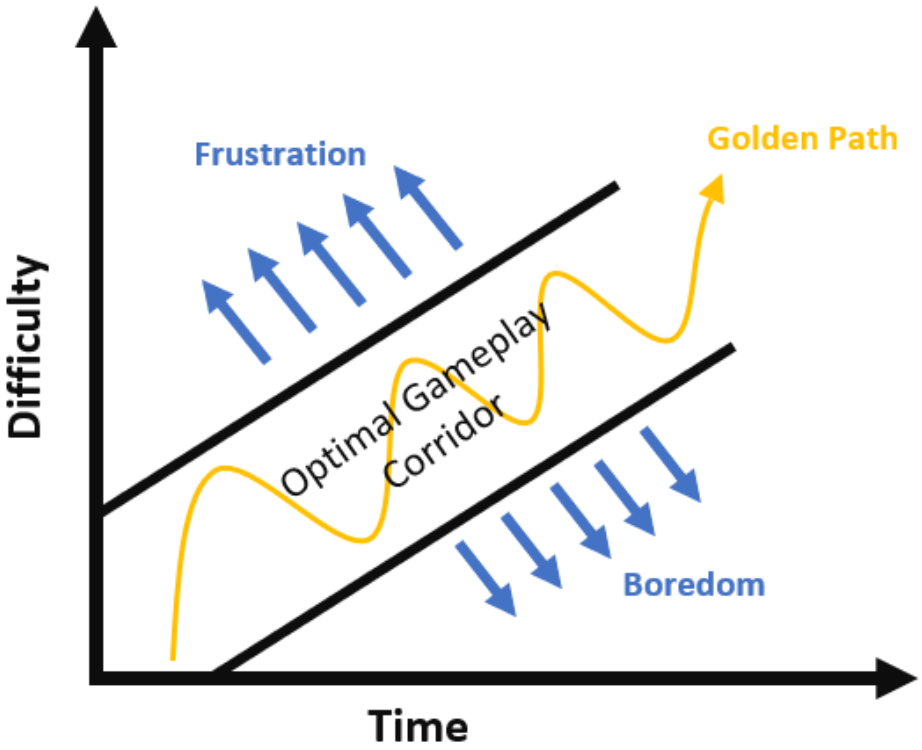


Figure 12: The optimal gameplay corridor and golden path, adapted from Thomas 2010.

To be able to get meaningful results from the adaptive difficulty levels, the system was designed to change only one variable in games. For the Memory game, the display time of the correct answer is the selected variable, while for the Space game, enemy spawn frequency is selected. These variables affect games directly but cannot be noticed by the player directly, unlike the number of letters in the Memory game or the type of enemies in the Space game, so that the students can give their honest opinion about how they feel about the game difficulty without any a priori clue. The flow of adaptive difficulty change is shown in Figure 10. The rules of the performance check and reasons for rules are elaborated in the sections below. Since the students could play these games only for a limited time, a single change in the variables was adjusted about 10% so that the change in these variables could affect the outcome quicker after the game gets easier or more difficult.

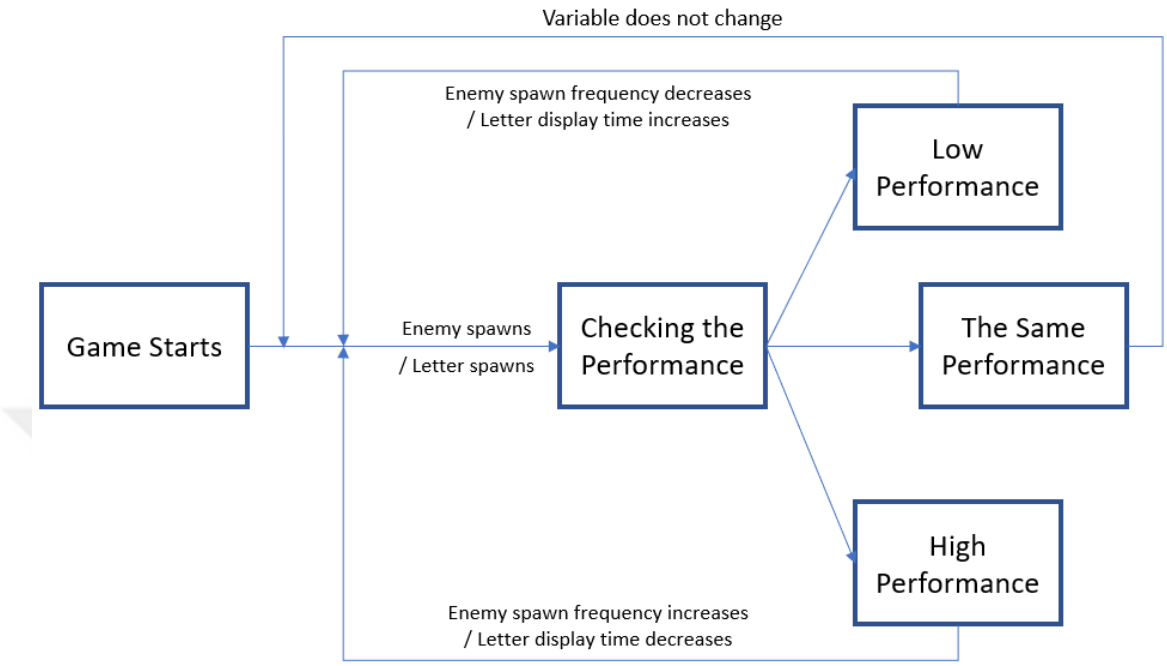


Figure 13: Visual diagram of the adaptive difficulty system.

4.1.1 Memory Game

For all the students with Dyslexia or Dysgraphia, a letter version of the Memory game was developed. The letter version of the Memory game contains only four consonants, 'p,' 'q,' 'b,' and 'd.' These are the most commonly confused letters for Dyslexic children when reading and writing. Although the Turkish alphabet does not contain the letter 'q,' children are familiar with it from the keyboard.

The adaptive system changes the first appearing time of the letters on the screen. The timer counts down from five ostensibly. However, it counts faster or slower according to the success of the players in the game. For instance, the timer goes from five to zero in three seconds if the player quickly answers earlier questions or goes in six seconds if the player struggles.

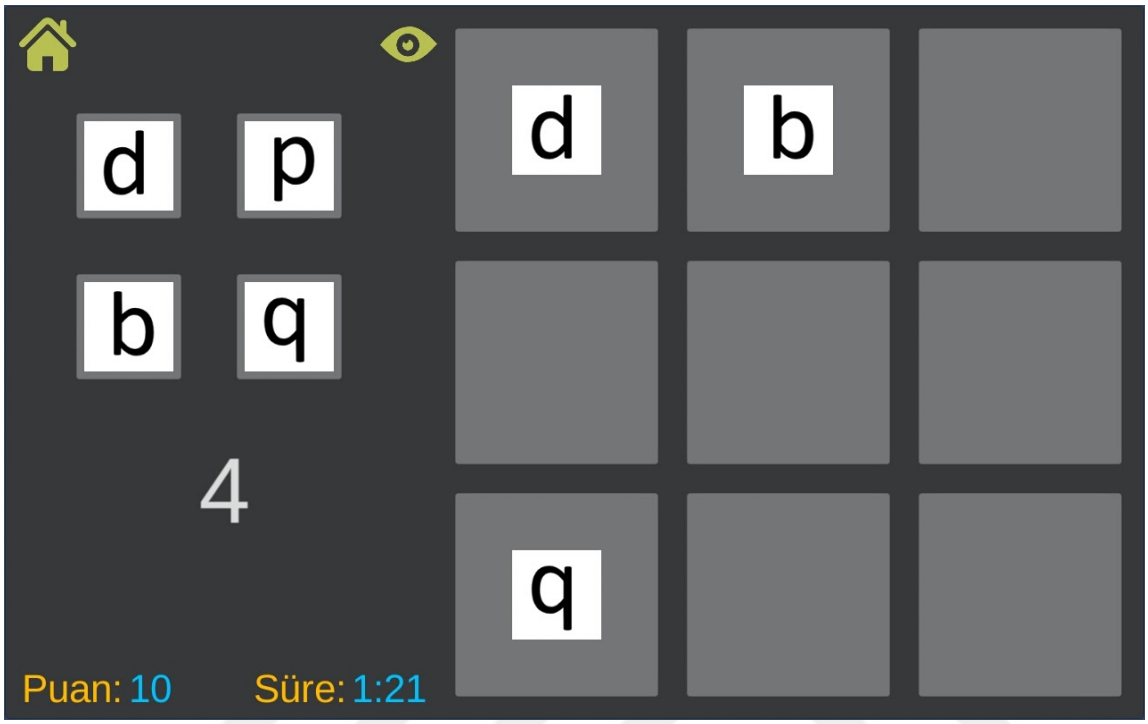


Figure 14: Screenshot of the letter version of the Memory game from Prototype III.

The rules of the adaptive difficulty are simple as they are explained as pseudocode in Algorithm 1. If the player can give two correct answers (C_{CCA}) in a row, appearing time (t_{LA}) decreases by half of a second. If the player provides more than one false solution, false grid, or incorrect letter, in a row (C_{CFA} is used for all), appearing time increases by half of a second (A_g is used for 'given answer'). Changes in the emerging time are applied in the next question.

The results of the first interview show that students give an average of 4.45 correct answers and 2.35 false answers (in total, details are presented in the Results section) for two grid difficulties in 60 seconds, which led to the selection of two consecutive correct or false answers as the adaptive difficulty change condition.

According to the results, the expected average difficulty change was more than two for the increment and more than one for the decrement for 100 seconds playtime. Due to these initial outcomes, the amount of the change was set as 0.5 seconds so that neither students can realize the difference directly nor they face massive difficulty changes, which could cause another difficulty change after the first one but in the opposite direction of that immediately.

Algorithm 1 AdaptiveDifficultyMemory

```
1:  procedure ADAPTIVEDIFFICULTYMEMORY (Ag, timer)
2:      CCCA ← 0
3:      CCFA ← 0
4:      tLA ← 5
5:      while timer != 0 do
6:          if Ag == true then
7:              CCCA ← CCCA + 1
8:              CCFA ← 0
9:              if CCCA == 2 then
10:                 tLA ← tLA - 0.5
11:                 CCCA ← 0
12:             end if
13:         else
14:             CCFA ← CCFA + 1
15:             CCCA ← 0
16:             if CCFA == 2 then
17:                 tLA ← tLA + 0.5
18:                 CCCA ← 0
19:             end if
20:         end if
21:     end while
22: end procedure
```

The number of correct and false answers before a change in appearing time and number of changes in appearing time was added to the database.

4.1.2 Space Game

Adaptive difficulty in the Space game only affects the spawn frequency of the enemy spaceships (Algorithm 2). If the player is successful, the frequency increases, and vice versa. The default waiting time (T_{ES}) between the spawn of two enemy spaceships is 0.4 seconds. The waiting time increases by 0.03 seconds when the number of crashes (C_{CE}) is three more than the number of collected gems (C_{CG}). Waiting time decreases by 0.03 seconds when the number of collected gems is two more than the number of crashes. Both counters reset when the spawning frequency changes (E_{CE} is used for the crashing enemy event, and E_{CG} is used for the collecting gem event).

According to the results of the first interview, students collected gems with an average of 0.997 times and crashed the enemies with an average of 1.317 times per 10 seconds, and their average playtime was 94.06 seconds. Due to these results, the number of events (crashing enemies or collecting gems) in the game is stored as one variable. Instead of the

consecutive events, the total number of crashes or collecting gems is expected to be more than the other one in order to change the difficulty.

To decrease the enemy spawn period, the number of collisions is expected to be three more than collecting gems instead of two, since the students were more likely to crash an enemy rather than collecting a gem. The change amount of the enemy spawn period is selected as 0.03 seconds because the Space game is a more dynamic game than the Memory game, and it needs more attention. Because of that, the change amount was kept below 10% in order to avoid a profound change in the difficulty.

Algorithm 2 AdaptiveDifficultySpace

```
1:  procedure ADAPTIVEDIFFICULTYSPACE (E_CE, E_CG timer)
2:      C_CE ← 0
3:      C_CG ← 0
4:      T_ES ← 0.4
5:      while timer != 0 do
6:          if E_CE == true then
7:              C_CE ← C_CE + 1
8:              if C_CE == (C_CG + 3) then
9:                  T_ES ← T_ES + 0.03
10:                 C_CE ← 0
11:                 C_CG ← 0
12:             end if
13:         else if E_CG == true then
14:             C_CG ← C_CG + 1
15:             if C_CG == (C_CE + 2) then
16:                 T_ES ← T_ES - 0.03
17:                 C_CE ← 0
18:                 C_CG ← 0
19:             end if
20:         end if
21:     end while
22: end procedure
```

CHAPTER 5

EXPERIMENTAL SETUP AND DATA ACQUISITION

5.1 Participants

The study is performed by two participant groups, students and their educators. Ethical Approval of Research has been approved by the Middle East Technical University Human Subjects Ethics Committee (ID of the permission). All participants, or their parents, if they are children, were informed both with verbal & written communication about the purpose of the study, possible benefits of the study to dyslexic children, and the voluntary nature of participation.

Educators include people with different professions, such as special education teachers, occupational therapists, and psychologists. Ten educators participated in study —7 of them are special education teachers, two of them are occupational therapists, and one of them is a psychologist. The mean age of educators is 33 (SD = 14.37), and seven of the educators are female. Children participants are selected among the students from Albatros Special Education Center (Albatros Özel Eğitim Merkezi, in Turkish), which provides education for children with SpLD. The student group includes 28 students having an age group of 7-11 ages. Permission of their parents is taken with both verbally and written. Also, children's consent is taken verbally. However, three of the children changed their school before the end of the study, so they could not complete the study. The mean age of the 25 children participants was 8.6 years (SD = 1.13). Most of the children were male (80%). Twenty of the children were diagnosed with the SpLD by a doctor. In contrast, the rest of the participant children had taken the Albatros Special Education Center's test, and they are being educated, given that the psychologists in the institution evaluated them as having SpLD.

5.2 Procedure

Two interviews were done with both participant groups in different months —started in October, ended in March. Firstly, educators played the unfinalized versions of the games. The concepts of games were clear, but rules, such as losing conditions or how many points a player gets after a correct answer, were not fully implemented yet at that time. Gamer profile, System Usability Scale (SUS), and Technology Acceptance Model (TAM) questionnaires were taken after the participants played the games. In this phase, the suggestions of educators were also collected via open-ended questions.

After the first interview with the educators, games were finalized. Children played that version of games in two 25 minutes sessions, which were held one week apart. Sessions with children were done in a classroom on a one-to-one basis, before or after their lessons. During the first week, the children played Math, Space, and Memory games at least twice. Gameplay duration for one turn was 100 seconds in the Math game, 60 seconds for the Memory game, and the children played the Space Game until they ran out of time. The average playtime was 94 seconds for the Space game. These games were played more than once because they are repeatable, and playtime is short. The Memory game was played three times since it had two versions in that interview.

During the second week, the children played Word and Category games, which took different times for each child, depending on their reading skills. Children's behaviors were observed by the authors, and notes were taken in the written format while the children were playing the games. Also, their scores, playing times, and game-specific statistics, such as the number of correct actions or collected gems, were taken. After playing, the same questionnaires —i.e., SUS and TAM— were filled by the children. Moreover, game-design related questions like "What do you want to change in this X game" or "What types of games can be added besides these games?" was asked to the children. In the end, all the students played all of the games regardless of their SpLD type.

Before the second interview, an adaptive system, which was explained in detail in chapter 4, was added to both Space and Memory games. Space and Memory games were chosen due to their game mechanics, game rules, and children's preferences. The second interview with the children was only one session where they played the Space and the Memory games twice —the first version and the adaptive version. Children played the Space game until they ran out of time (Average of 98 seconds) and played the Memory game 100 seconds per turn. To eliminate familiarization differences in statistics, the playing order of two versions was random, and children were unaware that the difficulty of the games might vary. In these interviews, children's behaviors were also observed by the authors, and notes were taken in written format. After children played each game's both versions, they answered four open-ended questions. The questions are, "Do you think there are any differences between those two versions of the same game? If any, what was that?" "What version makes you feel better? Why?" "Which version was easier? Do you think the enjoyment of games differs because of that?" and "If you should make a choice, which version would you choose to play a couple of times? Why?".

The final interviews were the second interviews of educators. After the first three phases, games were updated according to the observations and feedback from both children and educators. Educators tested the very final version. Finally, they asked final opinions, and their final questionnaires —i.e., SUS and TAM— were taken.

CHAPTER 6

RESULTS

In this chapter, data obtained from the interviews are analyzed. Throughout this thesis work, two interviews with both participant groups were done.

Answers to the following questions are examined with questionnaires, think-aloud method, and open-ended questions:

- Are proposed games suitable for educational purposes?
- Are children with SpLD comfortable with different types of games?
- Can adaptive difficulty system provide better game experience for children with SpLD?

In the following subsections, the results of participant groups are examined separately.

6.1 The Questionnaire Results of the Students

There are ten questions in the System Usability Scale (SUS), where half of the questions consist of positive statements, and the other half of them consist of negative statements. The questionnaire's results for students can be seen in Table 1. These questions were used to understand how usable the proposed games have been.

Table 1: System Usability Scale Questionnaire - Mean and standard deviation results of the first interview with students. Answers of the questionnaire are on a 1 to 5 scale; a higher score indicates a stronger agreement on statements.

	First Interview
Positive Statements	4.49 ± 0.39
Negative Statements	1.50 ± 0.36

Technology Acceptance Model Questionnaire (TAM) includes 19 questions about games to understand the acceptance level of participants on the proposed games. Table 2 shows the results of the TAM Questionnaire for students.

Table 2: Technology Acceptance Model Questionnaire - Mean and standard deviation results of the first interview with students. Answers of the questionnaire are on a 1 to 10 scale; a higher score indicates a stronger agreement on statements.

	First Interview
Positive Statements	9.07 ± 0.56

Participants were also asked to give ratings to each game on a 0 to 10 scale. Table 3 shows how much the students liked the games.

Table 3: Game Ratings - Mean and standard deviation results of the first interview with students. Ratings of the games are on a 0 to 10 scale; a higher score indicates a more liked game.

	First Interview
Math Game	8.52 ± 2.26
Word Game	8.36 ± 2.46
Memory Game	8.44 ± 2.32
Space Game	8.84 ± 2.29
Category Game	8.08 ± 2.92

The playtime of the games was recorded, and the average playing times are shown in Table 4.

Table 4: Average Playing Time of Games - Average and standard deviation of playing time in the first interview with students.

	Average Play Time in a Session (seconds)
Math Game	99.57 ± 2.00
Word Game	543.27 ± 210.81
Memory Game	725.63 ± 273.09
Space Game	93.94 ± 24.61
Category Game	58.35 ± 9.10

Open-ended questions were asked to students during their second interview to understand how they felt about the two versions of the game —adaptive and standard. Table 5 and Table 6 show the percentages of the answers to the open-ended questions.

Table 5: Open-Ended Questions - Percentage of answers on the adaptive and standard version of the Space game in the second interview with students.

	Standard Version	Adaptive Version	The Same/Does not matter
Do you think that there were any differences while playing the second time?	76%	24%	-
Which version made you feel better?	40%	28%	32%
Which version was harder?	8%	40%	52%
If you could access it one more week, which version would you prefer?	24%	28%	44%

*: For the first question, this column is used for the answer 'No.'
**: For the first question, this column is used for the answer 'Yes.'

Table 6: Open-Ended Questions- Percentage of given answer about the adaptive and standard version of the memory game in the second interview with students.

	Standard Version	Adaptive Version	The Same/Does not matter
Do you think that there were any differences while playing the second time?	72%	28%	-
Which version made you feel better?	28%	36%	36%
Which version was harder?	24%	12%	64%
If you could access it one more week, which version would you prefer?	28%	24%	44%

6.2 The Comments and Suggestions of the Students

Comments of children will be handled in three sections: comments on current games, future game ideas, general feedbacks; for the first part, children asked to comment on current game ideas, the difficulty level of games, visuals, and all related things to the games.

Most of the kids said that the Math game is too simple in terms of visualization and question type diversity. They all said that there should be other operations (multiplication & division), and there should be more themes besides kite themes such as planes and cars. Multiplication was added to the final version of the game.

Most of the students complained about the number of enemies and their small sizes in the Space game, which was fixed in the last version. Also, most of the students wanted to play Space game level by level rather than endlessly. Also, they said that they wanted to see "boss creatures" at the end of each level. Moreover, the majority of children suggested alternative control mechanics for space games such as buttons, touch joystick, or dragging. Touch joystick was applied in the last version after these suggestions.

There are not many common comments on the Word game, but there are a couple of creative suggestions. One child suggested that letters can be on pies or cakes instead of boxes that are unrelated to the environment. One other kid suggested that there can be riddle instead of direct questions. Only two of the kids said that it would be better to tapping on letters, like telephone's/tablet's keyboard inputs, in word game instead of dragging. Another student suggested that after every ten-question in the Word game, the background image can change, and new questions can be asked related to a new

background. It was the main goal of the word game. If there would be any other level, there will be another background (school, playground, or children's room), and questions would change relative to the background.

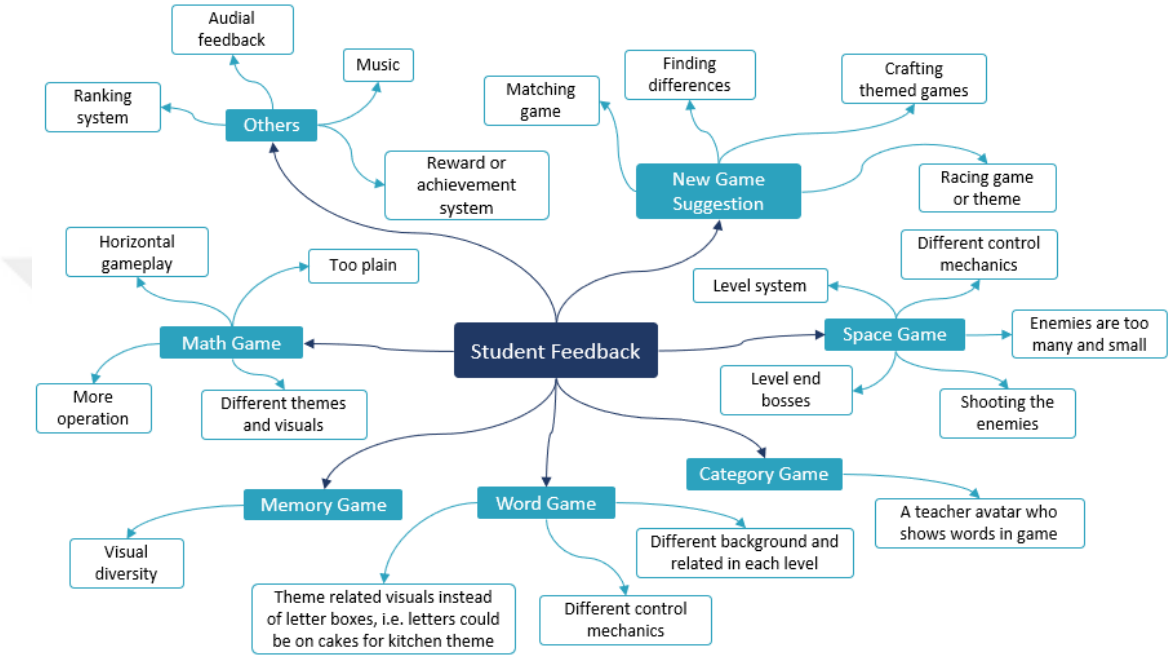


Figure 15: Student Feedback Illustration Map.

The most common comment for the Memory game was about a variety of images. Students want to play this game with images such as animals, shapes, cars, etc. Letters and sky objects versions of the game were created after these comments.

The Category game was the least commented game. One of the children suggested that since the game has a blackboard as a background, there could be a teacher in the game, which can show words and leads the game.

Children were asked to suggest a game idea, which can be the sixth game in this system later. Many of them did not suggest anything, or they wanted to add some existed games such as Minecraft, Brawl Stars, or Happy Glass. Other than that, most of the remaining kids want a car racing game or racing theme in existed games. For instance, they said the player could collect correct answers with the car and gets a boost in the Math game. Finding the difference between two pictures was suggested as the sixth game. Finding a match of pictures while many cards were faced down was also suggested. Finally, children wanted to play crafting themed games. One of the students wants a game in which players can collect some items in a given area —especially a room— and crafts some other items to escape from there. Also, another kid suggested that after correct answers, items can be gained, and producing new items with them like in Minecraft, will be joyful to play.

Finally, students are asked if they have any other comments in general. Three of them suggested that music or audial feedback in-game will make more fun. Most of the kids suggested an in-game achievement system. Finally, one of the children suggested a ranking system so that players can understand if they won something or not.

Figure 16 shows the most common words or phrases in the comments of students. The most common five words are: 'Colorful,' 'Ship,' 'Car,' 'Control,' and 'Enemy' respectively.

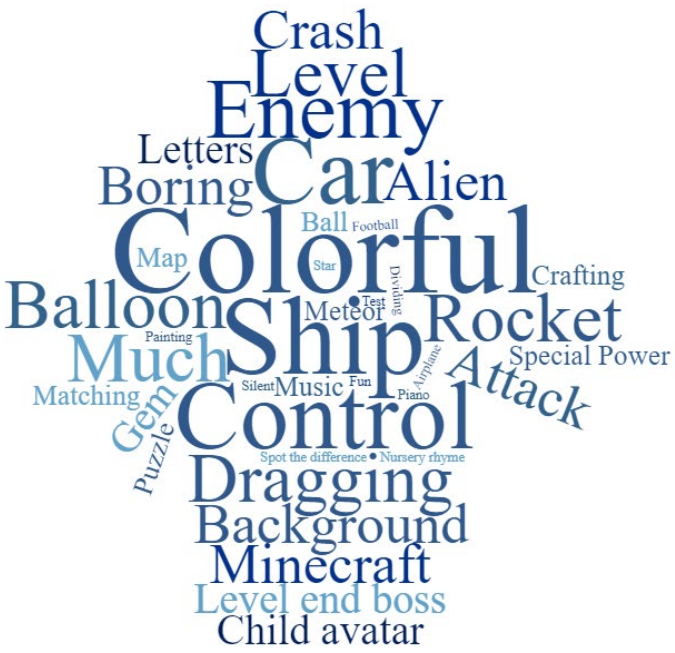


Figure 16: Student Feedback Word Cloud.

6.3 Game Statistics of the Students

In this subsection, statistics of the games held in the database are shown.

Table 7 and Table 8 show data of students in Memory Game from the first interview. 'Number of false grids' refers to when the player fills a grid which is supposed to be empty. 'Number of false-color/arrow markings' refers to when the player marks a grid with the wrong color or arrow option. The difficulty of Memory Game changes with the number of grids that should be memorized to be filled. Students played three times this

game; the first color version with low difficulty to get used to the game, then again color version with greater difficulty, and finally the arrow version. The level of the difficulty was set according to the age of students.

Table 7: Memory Game Statistics - Mean and standard deviation results of data of the first session with students according to how many times they played in that session.

Order of Play Sessions	Number of Correct Answers	Number of False Grids	Number of False Color/Arrow Markings
First Time	4.88 ± 1.24	0.16 ± 0.46	0.88 ± 1.37
Second Time	4.64 ± 0.93	0.24 ± 0.51	1.92 ± 3.49
Third Time	3.91 ± 1.10	0.74 ± 1.59	2.26 ± 2.23

Table 8: Memory Game Statistics - Mean and standard deviation results of data of the first session with students according to difficulty and version.

Version - Difficulty	Number of Students	Number of Correct Answers	Number of False Grids	Number of False Color/Arrow Markings
Color - 1 Grid	20	5.10 ± 1.26	0.20 ± 0.51	0.85 ± 1.42
Color - 2 Grids	13	4.15 ± 0.53	0.23 ± 0.42	0.77 ± 0.80
Color - 3 Grids	17	3.76 ± 1.11	0.88 ± 1.78	2.12 ± 2.49
Arrow - 1 Grid	3	5.67 ± 0.47	0.00 ± 0.00	0.00 ± 0.00
Arrow - 2 Grids	16	4.69 ± 0.98	0.19 ± 0.53	3.25 ± 3.98
Arrow - 3 Grids	5	4.20 ± 0.98	0.40 ± 0.80	1.40 ± 1.36

Table 9, Table 10, and Table 11 show the game statistics of the students from the first interview for Space Game, Math Game, and Word Game, respectively.

Table 9: Space Game Statistics - Mean and standard deviation results of data of the first session with students according to how many times they played in that session.

Order of Play Sessions	Number of Collected Gems	Number of Crashed Enemies	Average Play Time (sec)
First Time	7.36 ± 5.07	11.36 ± 2.68	88.44 ± 23.22
Second Time	11.58 ± 6.16	13.46 ± 4.82	99.67 ± 24.71

Table 10: Math Game Statistics - Mean and standard deviation results of data of the first session with students according to how many times they played in that session.

Order of Play Sessions	Number of Correct Answers	Number of False Answers	Number of Unanswered Questions	Number of Total Answers
First Time (Summation Only)	11.75 ± 4.15	1.46 ± 2.68	0.25 ± 0.52	13.46 ± 3.62
Second Time (Summation and Subtraction	10.46 ± 3.07	2.46 ± 3.21	0.29 ± 0.61	13.21 ± 3.98

Table 11: Word Game Statistics - Mean and standard deviation results of data of the first session with students in per 60 seconds and per 1 correct.

	Number of Correct Answers out of 10	Number of False Answers out of 10	Number of Unanswered Questions out of 10	Total Time (in seconds)
Per 60 Seconds	1.09 ± 0.48	0.61 ± 0.60	0.01 ± 0.04	60.00 ± 0.00
Per 1 Correct	1.00 ± 0.00	0.56 ± 0.49	0.01 ± 0.05	54.80 ± 21.49

Statistics of both parts of Category Game are shown in Table 12 and Table 13. For Part 2, an irrelevant answer to the given category is counted as 'False Answer' while misspelled words are counted as 'Typo.'

Table 12: Category Game (Part 1) Statistics - Mean and standard deviation results of data of the first session with students in per 60 seconds and per 1 correct.

	Number of Correct Answers out of 25	Number of False Answers out of 25	Total Time (in seconds)
Per 60 Seconds	4.98 ± 3.17	1.01 ± 0.90	60.00 ± 0.00
Per 1 Correct	1.00 ± 0.00	0.20 ± 0.23	12.05 ± 4.74

Table 13: Category Game (Part 2) Statistics - Mean and standard deviation results of data of the first session with students in per 60 seconds and per 1 correct.

	Number of Correct Answers	Number of False Answers	Number of Typos	Number of Total Answers	Total Time (sec)
Per 60 Seconds	1.58 ± 0.79	0.03 ± 0.06	0.15 ± 0.15	1.76 ± 0.78	60.00 ± 0.00
Per 1 Correct	1.00 ± 0.00	0.02 ± 0.04	0.10 ± 0.10	1.13 ± 0.11	38.00 ± 15.09

Table 14 shows the data of the students in Memory Game from the second interview, where they played each version of the game once.

Table 14: Memory Game Statistics - Mean and standard deviation results of data of the second session with students.

	Number of Correct Answers (per 10 seconds)	Number of False Grids (per 10 seconds)	Number of False Letters (per 10 seconds)
Normal	0.70 ± 0.16	0.08 ± 0.11	0.51 ± 0.45
Adaptive	0.72 ± 0.18	0.07 ± 0.08	0.45 ± 0.34

Space Game statistics from the second interview with the students are shown in Table 15.

Table 15: Space Game Statistics - Mean and standard deviation results of data of the second session with students.

	Collected Gems (per 10 seconds)	Crashed Enemies (per 10 seconds)	Average Play Time (sec)
Normal	1.20 ± 0.29	1.32 ± 0.44	103.92 ± 20.47
Adaptive	1.09 ± 0.24	1.35 ± 0.32	94.36 ± 17.63

Table 16 shows how many times the adaptive difficulty system adjusts the difficulty while the students were playing the games.

Table 16: Adaptive Difficulty Changes During Game - Mean and standard deviation results of in-game adaptive difficulty changes for both games in the second session with students.

	Number of Difficulty Up	Number of Difficulty Down
Memory Game	2.58 ± 1.41	1.25 ± 0.97
Space Game	1.52 ± 1.20	1.56 ± 0.98

Wilcoxon Signed Ranks Test results can be seen in Table 17. This test was applied to compare the standard and adaptive versions of the games since data were not normally distributed. Column 'Z' refers to the z-score, which is closer to 0 when the groups are evenly distributed. A p-value of ~0.05 is approximately equal to a z-score of 1.96. The asymptotic significance column shows the significance of the difference. A value of less than 0.05 is considered significant.

Figure 17 to Figure 20, box plots of game data from second interviews are displayed.

Table 17: Wilcoxon Signed Ranks Test - Results of test between data of two versions of both games from the second session with students.

	Z	Asymp. Sig. (two-tailed)
Memory Game – Correct Answers	-0.567	0.571
Memory Game – False Grid	-0.138	0.890
Memory Game – False Letter	-0.536	0.592
Space Game – Collected Gem	-2.173	0.030
Space Game – Crashed Enemy	-0.924	0.355
Space Game – Play Time	-1.884	0.060

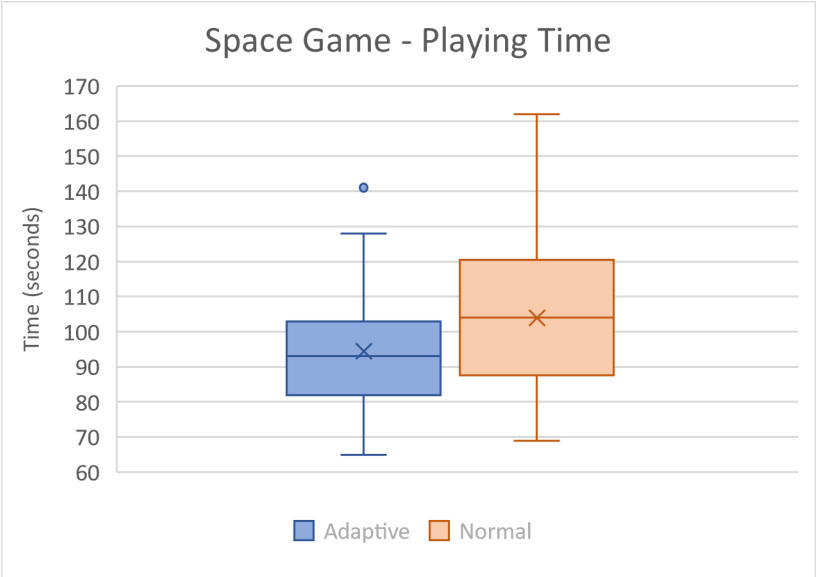


Figure 17: Box plot of playing time for Space game from the second session with students.

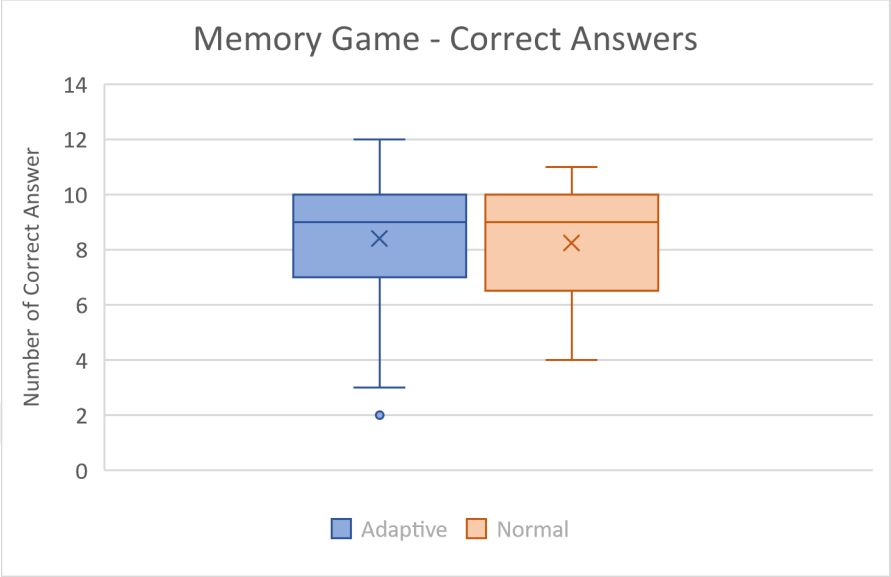


Figure 18: Box plot of correct answers for Memory game from the second session with students.

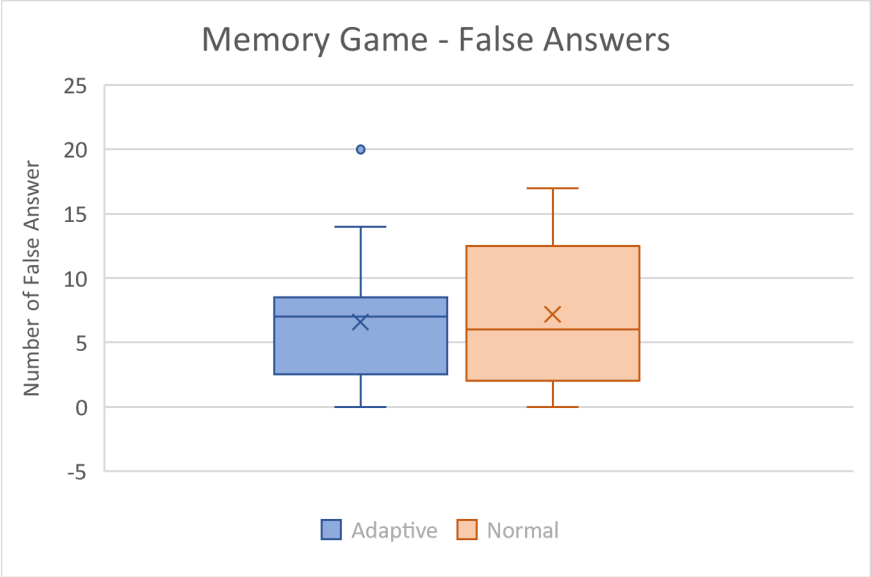


Figure 19: Box plot of false answers for Memory game from the second session with students.

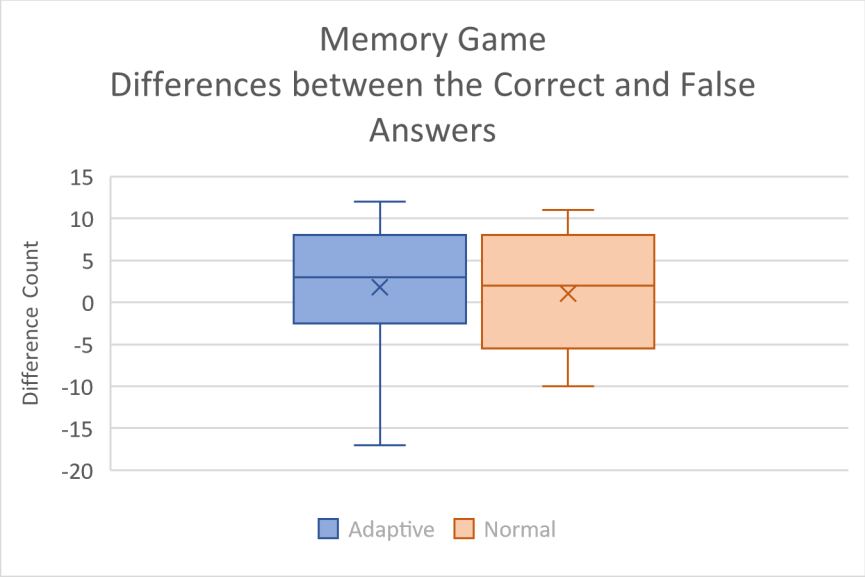


Figure 20: Box plot of the difference between the correct and false answers for Memory game from the second session with students.

6.4 The Results of the Educators

SUS and TAM Questionnaire results of educators from both interviews are shown in Table 18 and Table 19.

Table 18: System Usability Scale Questionnaire - Mean and standard deviation results of both interviews with educators. Answers of the questionnaire are on a 1 to 5 scale; a higher score indicates a stronger agreement on statements.

	First Interview	Second Interview
Positive Statements	4.54 ± 0.10	4.66 ± 0.08
Negative Statements	1.58 ± 0.29	1.32 ± 0.32

Table 19: Technology Acceptance Model Questionnaire - Mean and standard deviation results of both interviews with educators. Answers of the questionnaire are on a 1 to 10 scale; a higher score indicates a stronger agreement on statements.

	First Interview	Second Interview
Positive Statements	8.41 ± 0.87	8.72 ± 0.64

Table 20 shows the Wilcoxon Signed Ranks Test results of educators for both Questionnaires.

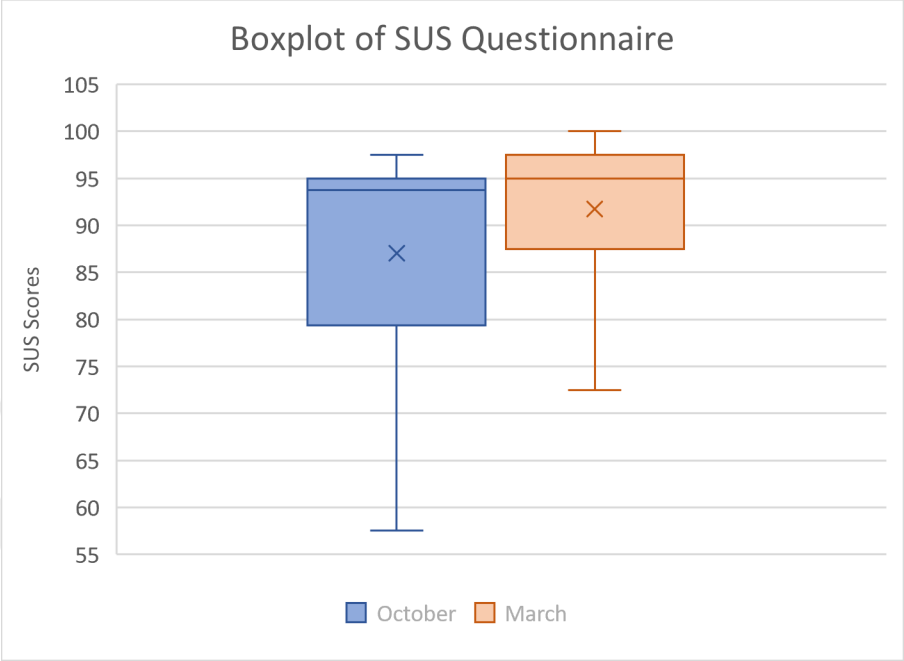
Table 20: Wilcoxon Signed Ranks Test - Results of test between two interviews of educators on both questionnaires.

	Z	Asymp. Sig. (two-tailed)
SUS Scores	-2.060	0.039
TAM Answers	-2.310	0.021

Educators were also asked how many points they would give to the games out of ten. Table 21 shows the ratings of educators from both interviews.

Table 21: Game Ratings - Mean and standard deviation results of both interviews with educators. Ratings of the games are on a 0 to 10 scale; a higher score indicates a more liked game.

	First Interview	Second Interview
Math Game	8.00 ± 2.19	8.00 ± 2.19
Word Game	8.50 ± 1.57	8.60 ± 1.43
Memory Game	7.80 ± 2.93	8.50 ± 2.06
Space Game	7.90 ± 1.51	8.80 ± 1.40
Category Game	7.80 ± 2.09	7.90 ± 1.97



From Figure 21 to Figure 24, box plots and individual plots of SUS and TAM Questionnaires can be seen.

Figure 21: Box plot of System Usability Scale scores of the educators from the first (October) and second (March) sessions.

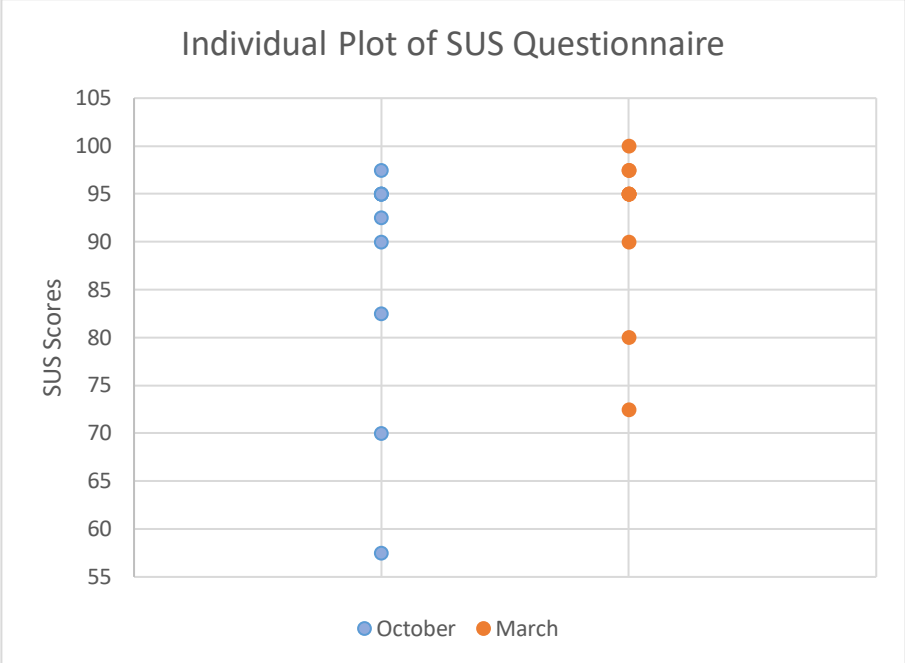


Figure 22: Individual plot of System Usability Scale scores of educators from the first (October) and second (March) sessions.

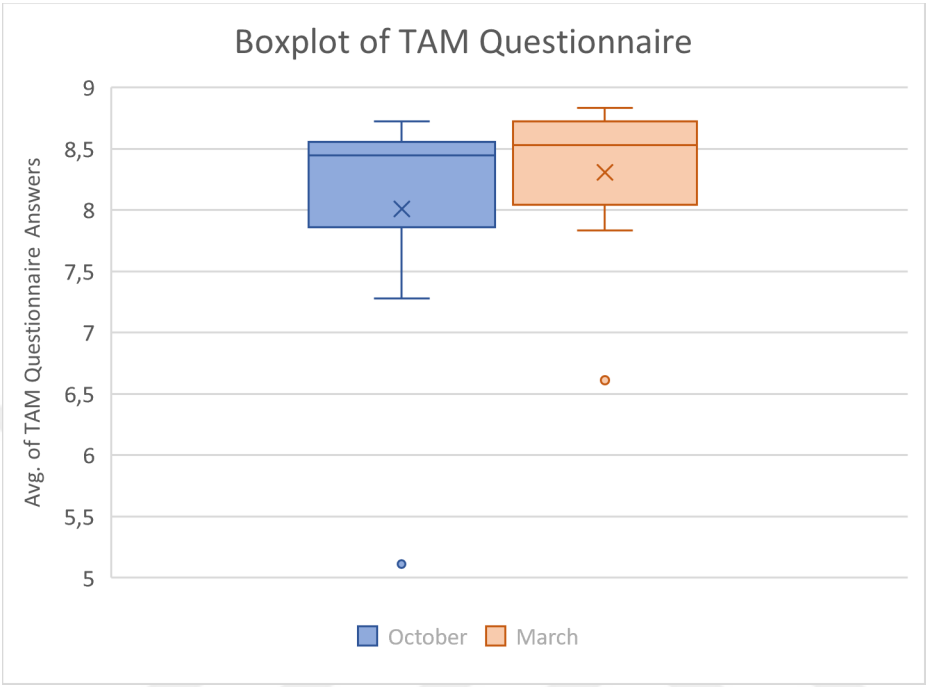


Figure 23: Box plot of Technology Acceptance Model questionnaire answers of educators from the first (October) and second (March) sessions.

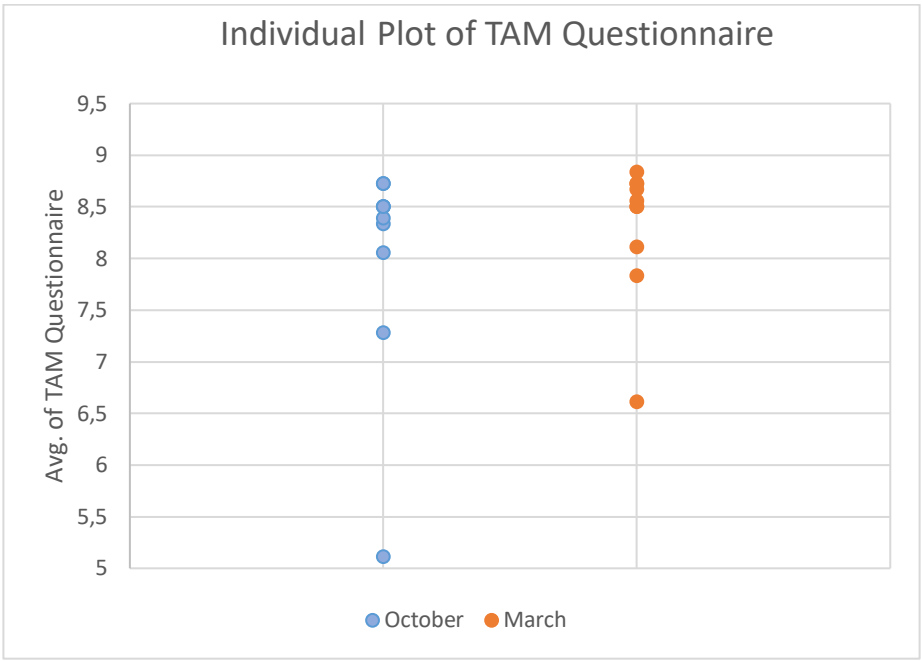


Figure 24: Individual plot of Technology Acceptance Model questionnaire answers of educators from the first (October) and second (March) sessions.

6.5 The Comments and Suggestions of the Educators

Comments of the educators will be handled in three sections: comments on current games, future game ideas, general feedback; for the first part, educators were asked to comment on present game ideas, the difficulty level of games, visuals, and all related things to the games.

One of the educators suggested that for the first part of the category, game pictures should be used instead of words. Another educator suggested that math games should be enriched by visuals without creating a distraction. Other teachers suggested that creating a shape by collecting a gem in the Space game would support the tracing ability of children. She also said that star or diamond should be gained instead of points since they correspond a visual for children.

Educators were also asked that is there any type of game that they want to see in this system. A game that enables the players to write letters by tracing a line or an area with fingers suggested by one of the educators. Another educator said that a game based on finding synonyms or antonyms would be beneficial for dyslexic children. Rhythmic counting is suggested as another version for Math game. "Spot 7 differences" type of puzzles and jigsaw puzzles are also suggested. Finally, an educator said that a game for improving the auditory memory of players would be beneficial and unique.

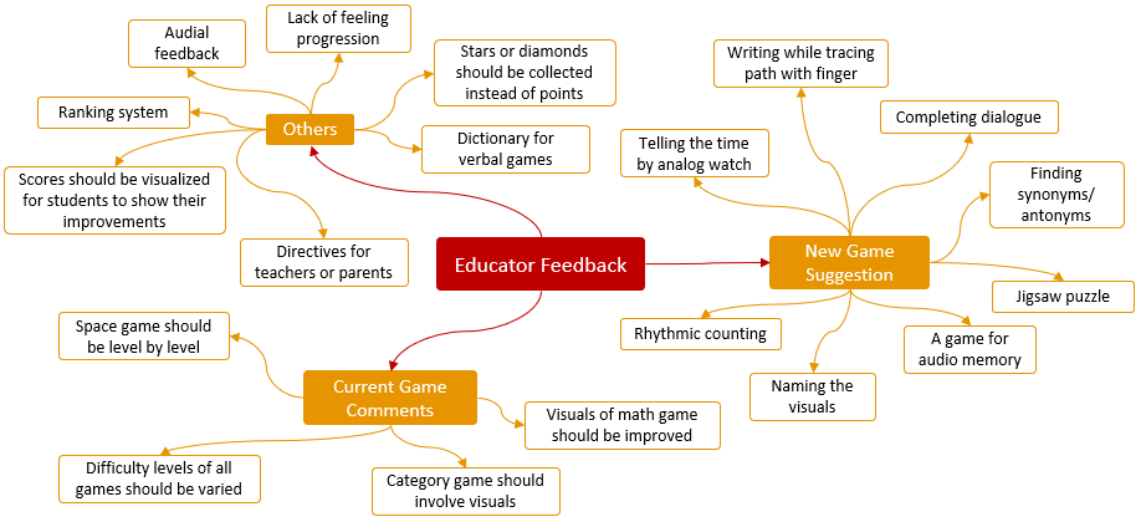


Figure 25: Educator Feedback Illustration Map.

As general feedback, three educators said that there is a lack of feedback so that the feeling of making progress does not occur. Also, two educators said that difficult levels should be diversified. Another two educators suggested keeping score as a leaderboard list and showing a graph that contains earlier scores would be beneficial and a source of motivation for children. One of the educators suggested that it would be better if there will be a guide for only educators or parents. Finally, although the majority of educators had a

favorable opinion about using this system in the course or as a homework or as auxiliary material, only one educator indicated that she would not use this system until parents have enough consciousness level about both the system and benefits of this system.

Figure 26: Educator Feedback Word Cloud.

Figure B shows the most common words or phrases in the comments. The most common five words are: 'Feedback,' 'Audial Feedback,' 'Well-thought,' 'Visual Elements,' and 'Reward,' respectively. The only negative feedback is "Will not use," and this was said due to the educator's concern about the unconsciousness of parents about tablet usage in special education.

CHAPTER 7

DISCUSSION

Results on system usability and technology acceptance of the students show that children evaluated the games as easy-to-use and easy-to-learn. “Easy to use” was a design criterion, and it can be said that it was accomplished. Since the game was designed with a single control mechanic and a single goal, children did not feel uncomfortable, even when the game itself was challenging (like the Space game).

Children with SpLD who are 7 to 11 years old were comfortable with different input types (dragging, tapping, joystick, etc.) when they were using only one of them during the game. The single goal does also matter because most of these children did not experience bigger games yet. To habituate the children with any type of SpLD to serious games, not only difficulty levels but also the complexity of games should increase stepwise. Also, in questionnaires, the results of feedback related questions are relatively low. There was not any audio feedback, and visual feedback was reduced to prevent any possible distraction. After this result, the final version of the games was changed to have slightly more visual feedback. The questionnaire results of the educators support these initial findings. Educators remarked that the games were easy to learn and play, but there was a lack of feedback even in the final version of the game, although it includes more visual feedback.

The average points of the games given by the children indicate that verbal games (Word and Category games) have the lowest scores (8.36 and 8.08 out of 10, respectively). The main reason for that is the duration of the games. Both Word and Category game took more than four times longer than other games while playing. Since playtime does not include any leveling in these games, it became boring for children. Also, children could give a correct answer per 54.8 –Word game– and 38 –the second part of the Category game– which can be accepted as an exceedingly long time for students with Dyslexia. The point difference between the Category and Word games is based on the visual difference. Word game has a child avatar and a kitchen illustration, which is a more child-oriented visual theme. Thus, the Word game was not the least liked game instead of the Category game, which only includes a blackboard and some writings.

This result conflicts with Shabbir and Bhatti’s rules (Shabbir & Bhatti, 2019) for serious games for Dyslexic children. Shabbir and Bhatti claim that not using a plain background can cause a distraction for children. However, both questionnaires and comments of the

children show that game-related static background images or avatars increase immersion. The kitchen image on the background or child avatar that asks questions was not mandatory for the Word game, but children stated that they loved both. A couple of them suggested a similar visual design for the Category game.

The average points by both participant groups are between 7.9 and 8.84 out of 10. When these results combine with the questionnaire results, it can be easily said that both participant groups liked all games, and they can accept all parts of the system as an educational source. Unlike other studies on serious games for children with SpLD, which are mentioned in the Related Work chapter, this study proposes a collection of games to be used in the education of children with any type of SpLD, and all of the games can be played by any children regardless of the type of their SpLD. A specially designed set of games for these children can provide a complete educational source and better experience rather than only playing a game that is directly related to their disorder.

The average for the number of days when students want to play these games is 3.6 per week. This average can be accepted as high since the students' permission to play games is limited. Almost half of the kids indicate that they are allowed to play games only during the weekend or every other day, and they said that they would play these games in that limited time.

A student stated that she liked the inclusion of the questions related to the theme in the Word game and the background to the theme, and said, *"Let the questions come in different themes in the new sections, and the background should change accordingly"* — a suggestion that exactly matches the design of the game. Despite the possibility of showing the background as mixed up, the idea of creating a theme by using visuals was welcomed by the students.

There is a difference between statistics of color and arrow version of the Memory game. Children almost had the same average correct answer and false grid answer. However, the false-color answer (0.137) is way less than the incorrect direction answer (0.501) per 10 seconds. This can be related to Dyslexia, which led to developing a letter version of the game.

In this thesis, a rule-based adaptive difficulty system is proposed to be used in serious games which adjusts difficulty during the game, while other studies in literature used a system which changes difficulty between two play sessions (Hendrix et al., 2018) and used predefined levels which is limited (Cuschieri et al., 2014).

In the Space game, the adaptive version of the game did not change the results considerably for the students. Their success was affected barely. Moreover, the speed change in the game was practically equal, which means that they nearly had the same statistics at the same difficulty level. It can be interpretable in a couple of ways. The first possible reason for this result can be bad adjusted adaptive rules or adaptive effects. If adaptive rules are too hard to achieve, the players will usually play in normal difficulty,

or if adaptive changes affect the game difficulty, too many players probably increase or decrease back to the standard difficulty level. The other possible reason for this result can be too much-balanced game difficulty. The game could have such a difficulty that players can collect the gem and crash enemies at almost the same rate. However, this is not a desirable outcome since children should encounter this situation after a successful game period to feel the progress.

Results of the Wilcoxon Test for Memory game showed that children had better performance in the adaptive version of the game, but it is not statistically significant. Since the adaptive version increases its difficulty while players show good performance, children played a more difficult version and showed better performance. It means that the flow strategy worked. Due to the increased difficulty level, the performance of children could not be statistically different, but still better.

Especially for adaptive games, more tests should be done for different variables. This study includes only one variable change —enemy spawn frequency for Space game and question screen time for Memory game— which gives promising results for the future, especially in the Memory game. Children did not realize they were playing faster, and they showed slightly better performance in adaptive versions. More tests can be applied with changing grid numbers or changing the order of the pre-defined letters in the left panel.

During the test of the adaptive games, only one student was able to say that he understood the change in the Memory game and explained the correct reason, while others could not explain the change they felt in both games, misjudged them, or did not feel any change.

Many students have shown that they are influenced by the games by using the cries of sadness, joy, or surprise during the games —mostly in the Space game. Except for one, none of the students objected during the replay of the games and even demanded to play more than they wanted. A single student stated in the first interview that she did not want to play the Space game for the second time and that it was not necessary. The same student did not make such a request in other games or in the second interview.

The fact that the students wanted different visual themes in the Math game showed that the game could be diversified visually, not mechanically. The students saw the problem visually in terms of repeatability; they did not see any problem in solving the repeat operations.

One of the words that students talked about the most was boss creatures. It may be good to add them as follows, as positive feedback in the games that are designed in levels instead of endless games. In addition, while the game ends with the mistake of the player in endless games, in level-based games, the success of the player leads to a new stage, which changes the feeling at the end of the game. Boss creatures can be good for strengthening the feeling at the end of the episode.

Comments of students on serious games were usually game-related things; they did not interest in the educational purpose of games. This shows that children did not see these games as educational material, but they accept these games as normal games, which is the basic purpose of the concept of the serious game.

CHAPTER 8

CONCLUSION AND FUTURE WORK

With increasing awareness of serious games and SpLD, studies in this area became varied in recent years. Many types of research suggest design rules or test a game on children with SpLD. However, those studies deal with only one type of SpLD. This thesis evaluates five games that were specially designed for children with SpLD, all four types. Games were played by both educators and students, and their reviews were taken. The suggested "all in one" type system was welcomed by both participant groups. Moreover, different design criteria, which were indicated in earlier studies, are examined in this thesis. Some new design criteria are suggested with an alternative to existing ones.

This study also includes an adaptive difficulty level, and it was found that the performance of children may increase with the adaptive difficulty even though it gets more challenging than the standard games. It was found that most of the children did not notice any difference in the difficulty if it was done gradually. These small changes made Dyslexic children focus on the games all the time and increased their immersion.

Targeting different SpLD types using serious games has been one of the contributions of this study. Each child participant had more than one type of SpLD, and instead of just playing games that focus on Dyslexia or Dyscalculia has been useful to them. In addition, playing games such as Memory and Space games, together with the educational-oriented question-and-answer games, increased the students' positive feedback and enthusiasm in general. In addition, the students said that they were happier to play in a general system instead of just playing games in the area where they had difficulty.

Although we used a simple adaptive system, it was good for the students to see how they would face this challenge. In the future, the adaptive system can be diversified with different variables without the need for complex systems. In addition, the fact that the adaptive system is in-game is a positive step because the games can be played in 3 or 4-minute sessions by the teachers, and the students are very likely to be disappointed or bored, so the in-game variability is a good move to prevent this boredom. Since only the same student will continue to play with this system for a long time, this variability must be provided in the process, and incremental difficulty must be established accordingly.

The general thoughts of the teachers are quite positive; most of them plan to use this system when they are seated. They already have simple math applications that they use for practice now, and if they switch to a single application where they can give homework and track performance that can also be used in the classroom, they said that they would use it with every student.

REFERENCES

Annetta, L. A. (2010). The "I's" have it: A framework for serious educational game design. *Review of General Psychology*, *14*(2), 105-113.

Avila-Pesantez, D. F., Vaca-Cardenas, L. A., Avila, R. D., Padilla, N. P., & Rivera, L. A. (2018, August). Design of an Augmented Reality serious game for children with Dyscalculia: A case study. In *International Conference on Technology Trends* (pp. 165-175). Springer, Cham.

Balcı, E. (2019). Disleksi hakkında öğretmen görüşleri ve karşılaştıkları sorunlar. *Ege Eğitim Dergisi*, 20(1), 162-179.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, 24(6), 574-594.

Bartolomé, N. A., Zorrilla, A. M., & Zapirain, B. G. (2012, July). Dyslexia diagnosis in reading stage though the use of games at school. In *2012 17th International Conference on Computer Games (CGAMES)* (pp. 12-17). IEEE.

Bellotti, F., Ott, M., Arnab, S., Berta, R., de Freitas, S., Kiili, K., & De Gloria, A. (2011, October). Designing serious games for education: from pedagogical principles to game mechanisms. In *Proceedings of the 5th European Conference on Games Based Learning*. University of Athens, Greece (pp. 26-34). Academic Publishing Limited.

Bernhaupt, R. (Ed.). (2015). *Game user experience evaluation*. Springer.

Brooke, J. (1996). SUS: a "quick and dirty'usability. *Usability evaluation in industry*, 189.

Calderón, A., & Ruiz, M. (2015). A systematic literature review on serious games evaluation: An application to software project management. Computers & Education, 87, 396-422.

Cuschieri, T., Khaled, R., Farrugia, V. E., Martinez, H. P., & Yannakakis, G. N. (2014, September). The iLearnRW game: support for students with Dyslexia in class and at home. In *2014 6th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)* (pp. 1-2). IEEE.

Edvardsen, F., & Kulle, H. (2010). Educational games: Design, learning and applications.

El Kah, A., & Lakhouaja, A. (2018). Developing effective educative games for Arabic children primarily dyslexics. *Education and Information Technologies*, 23(6), 2911-2930.

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Doctoral dissertation, Massachusetts Institute of Technology).

Di Tore, P. A., Di Tore, S., Ludovico, L. A., & Mangione, G. R. (2014, September). Madrigale: a multimedia application for dyslexia and reading improvement gamifying learning experience. In *2014 International Conference on Intelligent Networking and Collaborative Systems* (pp. 486-491). IEEE.

Flanagan, R. (1995). A review of the Behavior Assessment System for Children (BASC): Assessment consistent with the requirements of the Individuals with Disabilities Education Act (IDEA). *Journal of School Psychology*, 33(2), 177-186.

Franceschini, S., Gori, S., Ruffino, M., Viola, S., Molteni, M., & Facoetti, A. (2013). Action video games make dyslexic children read better. *Current Biology*, 23(6), 462-466.

Freilich, R., & Shechtman, Z. (2010). The contribution of art therapy to the social, emotional, and academic adjustment of children with learning disabilities. *The Arts in psychotherapy*, 37(2), 97-105.

Gaggi, O., Palazzi, C. E., Ciman, M., Galiazzo, G., Franceschini, S., Ruffino, M., ... & Facoetti, A. (2017). Serious games for early identification of developmental Dyslexia. *Computers in Entertainment (CIE),* 15(2), 1-24.

Garrett, J. J. (2010). *The elements of user experience: user-centered design for the web and beyond*. Pearson Education.

Hendrix, M., Bellamy-Wood, T., McKay, S., Bloom, V., & Dunwell, I. (2018). Implementing adaptive game difficulty balancing in serious games. *IEEE Transactions on Games*.

Ludovico, L. A., Di Tore, P. A., Mangione, G. R., Di Tore, S., & Corona, F. (2015). Measuring the Reading Abilities of Dyslexic Children through a Visual Game. *iJET*, 10(7), 47-54.

Mayer, I., Bekebrede, G., Harteveld, C., Warmelink, H., Zhou, Q., Van Ruijven, T., ... & Wenzler, I. (2014). The research and evaluation of serious games: Toward a comprehensive methodology. *British journal of educational technology*, 45*(3)*, 502-527.

McLoughlin, J. A., & Wallace, G. (1990). *Learning disabilities: Concepts and characteristics*. Prentice Hall.

Moll, K., Kunze, S., Neuhoff, N., Bruder, J., & Schulte-Körne, G. (2014). Specific learning disorder: Prevalence and gender differences. *PLoS one*, *9*(7), e103537.

Moreno-Ger, P., Torrente, J., Hsieh, Y. G., & Lester, W. T. (2012). Usability testing for serious games: Making informed design decisions with user data. *Advances in Human-Computer Interaction, 2012*.

Rauschenberger, M., Rello, L., Baeza-Yates, R., & Bigham, J. P. (2018). Towards language independent detection of Dyslexia with a web-based game. *In Proceedings of the Internet of Accessible Things* (pp. 1-10).

Rauschenberger, M., Rello, L., & Baeza-Yates, R. (2018, September). A tablet game to target dyslexia screening in pre-readers. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (pp. 306-312).

Rello, L., Bayarri, C., Otal, Y., & Pielot, M. (2014, October). A computer-based method to improve the spelling of children with Dyslexia. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility* (pp. 153-160).

Shabbir, N., & Bhatti, Z. (2019, December). Serious Game User Interface Design Rules for dyslexic children. In *2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)* (pp. 1-6). IEEE.

Thomas, J. M., & Young, R. M. (2010). Annie: Automated generation of adaptive learner guidance for fun serious games. *IEEE Transactions on Learning Technologies*, 3(4), 329-343.

Venkatesh, V., & Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision sciences*, *27*(3), 451-481.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, *46*(2), 186-204.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, *46*(2), 186-204.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, *7*(3), 225-240.

Zyda, M. (2005). From visual simulation to virtual reality to games. Computer, 38(9), 25-32.

APPENDICES

QUESTIONNAIRES – TECHNOLOGY ACCEPTANCE MODEL

Technology Acceptance Model (Answers from 1 to 10, 1 being most negative)
I liked the idea that creating a game platform about Dyslexia.
Was playing the games fun?
During the game, I felt pain and/or discomfort.
The games are challenging and exciting.
It's easy to learn to play proposed games.
The game reacts readily to my movements.
I did not find it hard to interact with the virtual world finger.
I found the graphical interface clear and explanatory.
The interface is simple to use.
I liked the interface design.
The instructions of the game are clear. I understood what to do in the game and how.
The program helps me to understand if and when I'm acting incorrectly.
I like the theme.
I like the cartoon-like graphics style.
The tablet screen is big and close enough.
I found it easy to reach targets.

I would like to keep training on Dyslexia heritage with this system in the future.
If I had the option to keep using the system at home, I would play with them often.
If I could use the system at home, I would play (select a number) _ _ _ _ _ days per week. (Scale 1 to 7)

QUESTIONNAIRES – SYSTEM USABILITY SCALE

System Usability Scale (1: Strongly Disagree 5: Strongly Agree)
I think that I would like to use this system frequently.
I found the system unnecessarily complex.
I thought this system was easy to use.
I think I would need assistance to be able to use this system.
I found various functions in this system were well integrated.
I found the system unnecessarily complex.
I would imagine that most people would learn to use this system very quickly.
I found this system very cumbersome/awkward to use.
I felt very confident using this system.
I needed to learn a lot of things before I could get into this system.

QUESTIONNAIRES – GAME RATINGS

How much did you like the ____ __ on a scale from 1 to 10 (1 being most unliked)	
	Math Game
	Word Game
	Memory Game
	Space Game
	Category Game

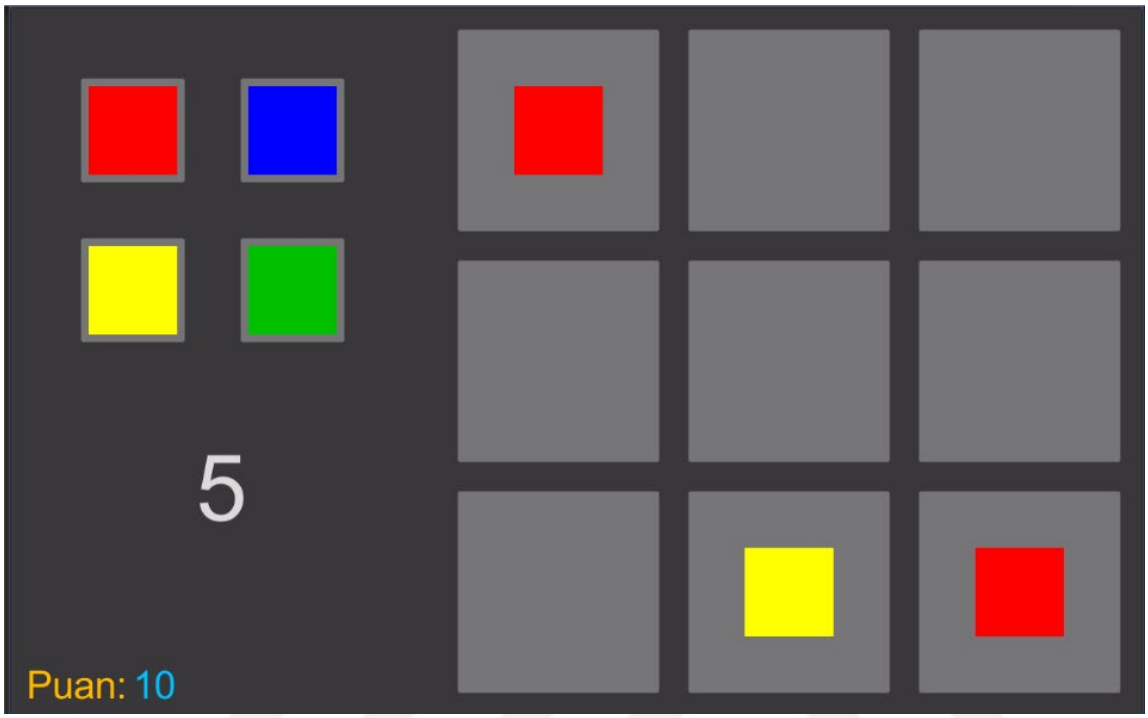
QUESTIONNAIRES – SUGGESTIONS

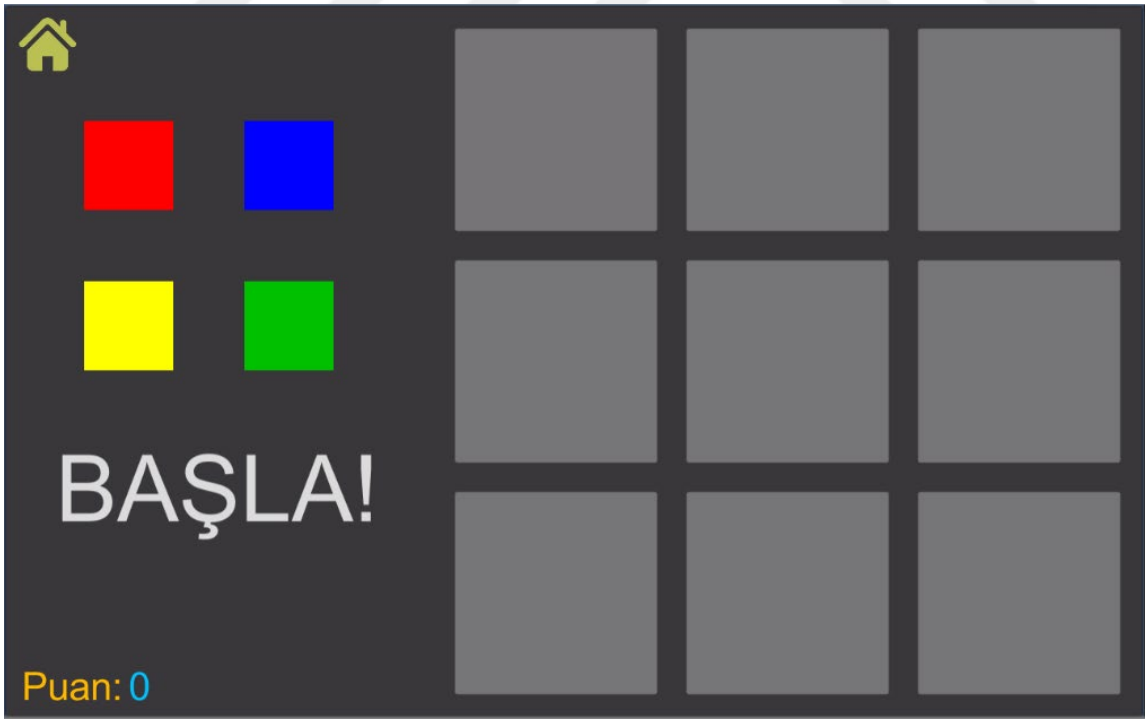
Suggestions on the Games
What are your suggestions for us to improve the Math Game?
What are your suggestions for us to improve the UI of Math Game?
What are your suggestions for us to improve the Space Game?
What are your suggestions for us to improve the UI of Space Game?
What are your suggestions for us to improve the Memory Game?
What are your suggestions for us to improve the UI of Memory Game?
What are your suggestions for us to improve the Category Game?
What are your suggestions for us to improve the UI of Category Game?
What are your suggestions for us to improve the Word Game?
What are your suggestions for us to improve the UI of Word Game?
Suggestions on the System
What kind of game would you like to play as sixth game of this system?
Which game did you like most? Why?
Insert here all your additional comments.

QUESTIONNAIRES – OPEN-ENDED QUESTIONS FOR ADAPTIVE DIFFICULTY

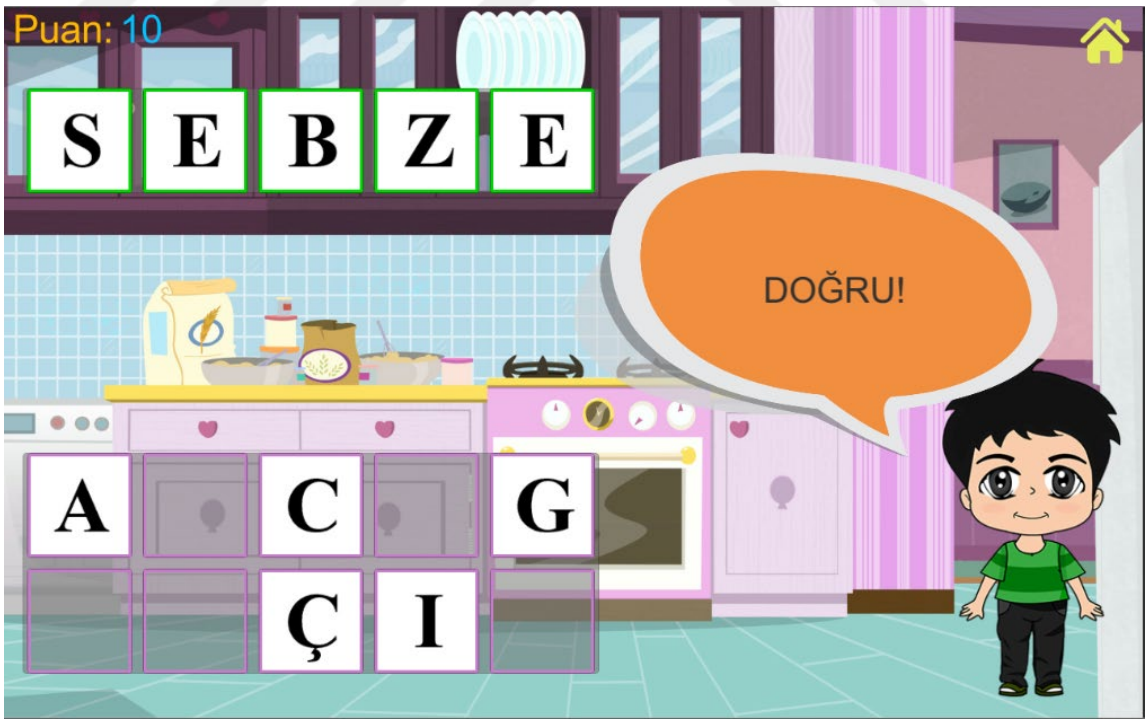
Do you think that there were any differences while playing twice?
Which version made you feel better?
Which version was more difficult?
If you can access it one more week, which version would you prefer?

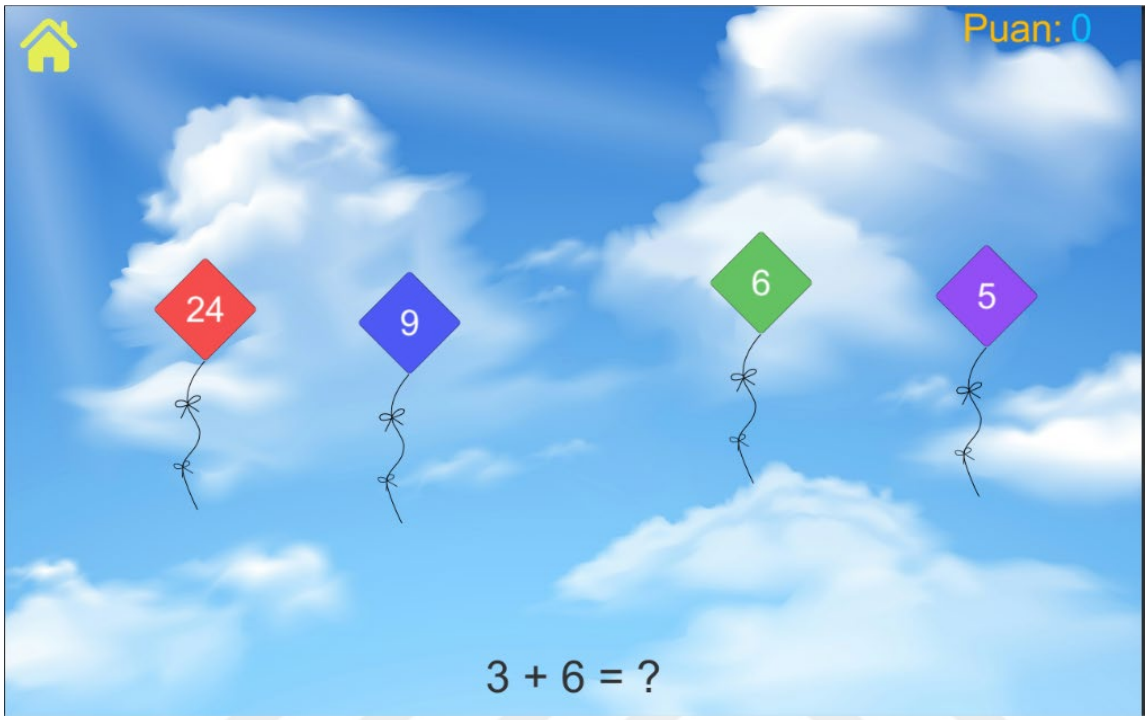
Additional Game Visuals from Different Phases of the Data Acquisitions

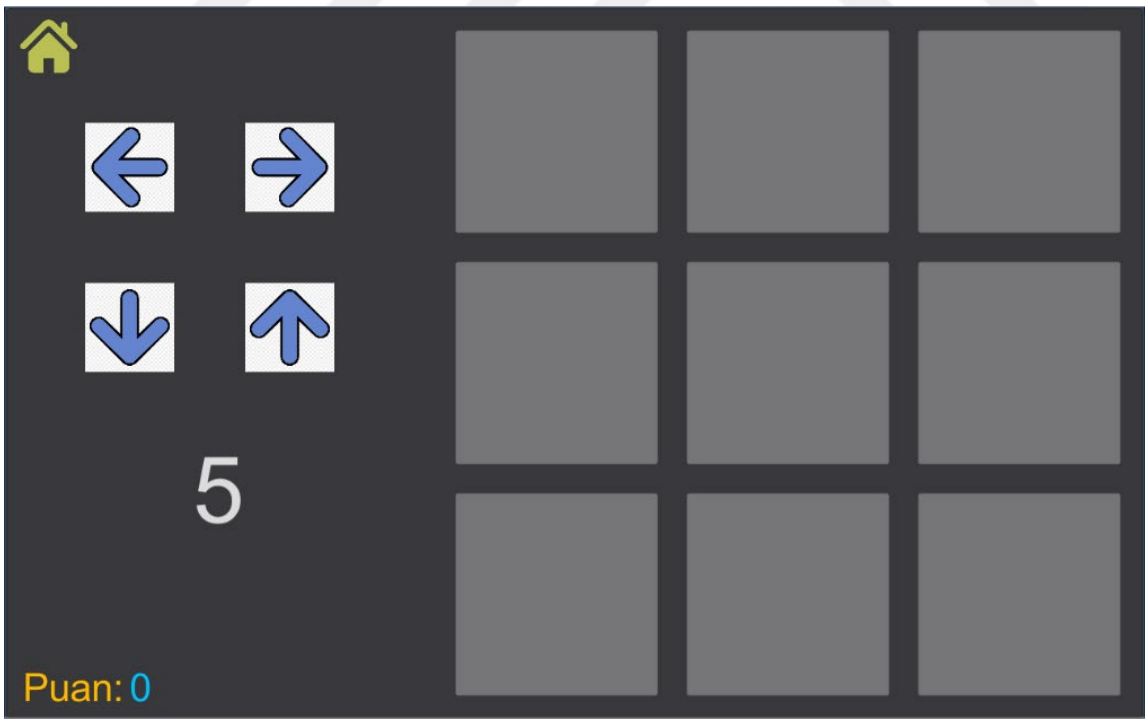

5
Puan: 10

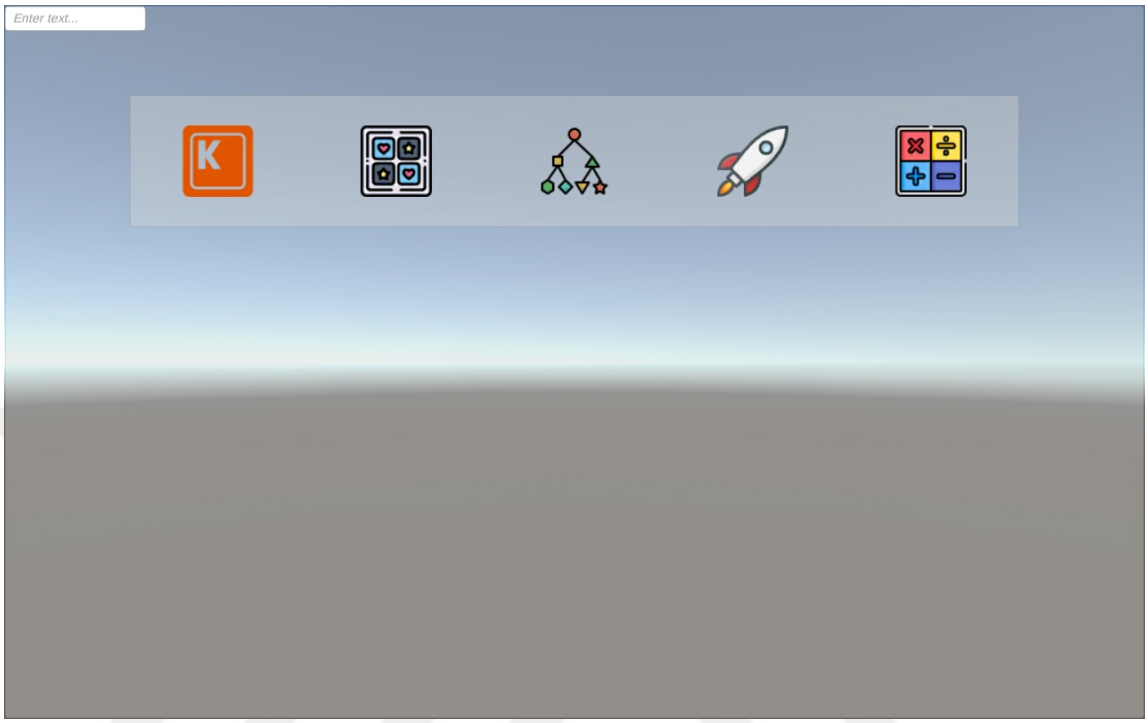

BAŞLA!
Puan: 0

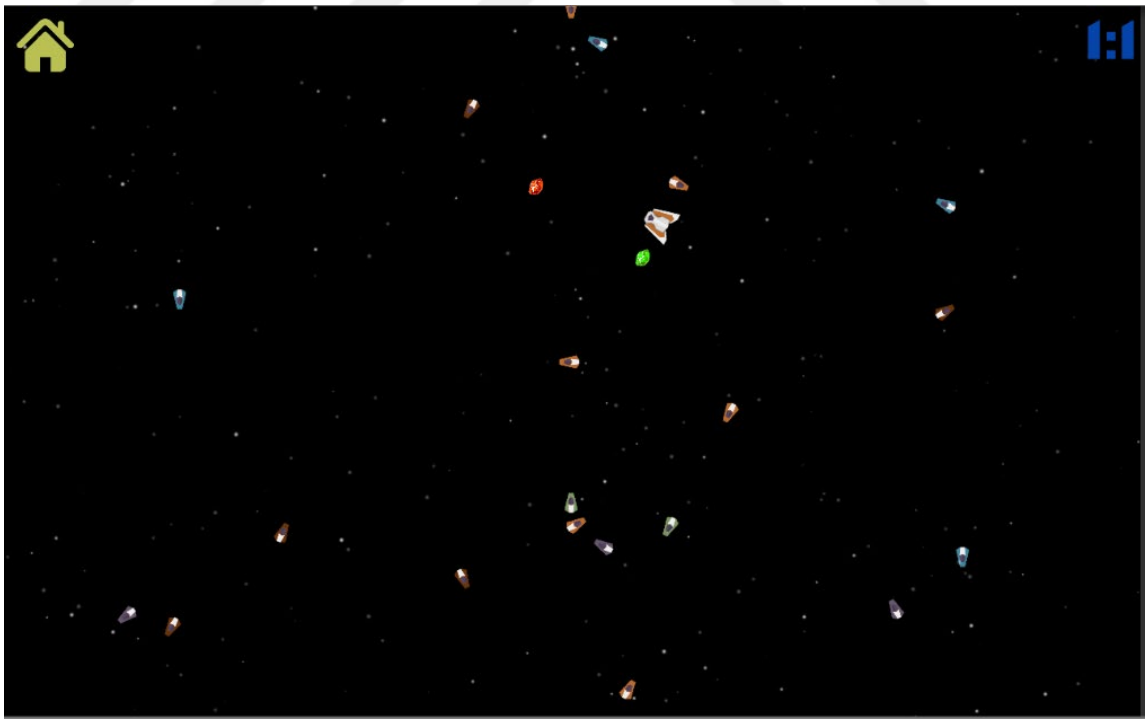

Puan: 0
S E B
Yemeklerde en çok kullandığımız havuç, ıspanak gibi yiyeceklerin genel adı nedir?
A C G
E Z Ç I


Puan: 10
S E B Z E
DOĞRU!
A C G
Ç I


Puan: 0
24
9
6
5
3 + 6 = ?


5
Puan: 0


Enter text...


1:1