



# Intelligent gamification for individualized dyslexia remediation and in-home therapy

Hossein Jamshidifarsani

## ► To cite this version:

Hossein Jamshidifarsani. Intelligent gamification for individualized dyslexia remediation and in-home therapy. Psychology. Université Paris-Saclay, 2021. English. NNT : 2021UPASW002 . tel-03585122

**HAL Id: tel-03585122**

**<https://theses.hal.science/tel-03585122>**

Submitted on 23 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Intelligent gamification for individualized dyslexia remediation and in-home therapy

## Gamification intelligente pour une remédiation à domicile et individualisée de la dyslexie

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°566 : Sciences du Sport, de la Motricité et du Mouvement Humain (SSMMH)

Spécialité de doctorat: Psychologie

Unité de recherche : Université Paris-Saclay, UVSQ, Inserm, END-ICAP, 78000, Versailles, France

Référent : Université de Versailles -Saint-Quentin-en-Yvelines

Thèse présentée et soutenue à Montigny-le-Bretonneux, le 22/02/2021, par

**Hossein JAMSHIDIFARSANI**

### Composition du Jury

**Pr. Saida BOUAKAZ**

Professeure des Universités  
Université Claude Bernard Lyon 1

Présidente du jury

**Pr. Rui CORTESAO**

Professeur des Universités  
University of Coimbra

Rapporteur

**Dr. Domitile LOURDEAUX**

Maîtresse de Conférences  
Université de Technologie de Compiègne

Rapporteuse

**Pr. Pascal GUITTON**

Professeur des Universités  
Université Bordeaux

Examineur

**Pr. Jean-Claude MARTIN**

Professeur des Universités  
Université Paris Sud

Examineur

**Pr. Saeid NAHAVANDI**

Professeur des Universités  
Deakin University

Examineur

### Direction de la thèse

**Dr. Samir GARBAYA**

Maître de Conférences  
Arts et Métiers ParisTech

Encadrant de thèse

**Pr. Pierre BLAZEVIC**

Professeur des Universités  
Université Versailles St-Quentin-en-Yvelines

Directeur de recherche

**M. Lawrence HOWARD**

Hands-Free Computing Ltd.

Invité



## Acknowledgements

---

*I would like to thank many individuals whose kind support and help turned my work into this thesis.*

*I would like to express my deepest gratitude to my supervisor Dr. Samir Garbaya that guided me throughout my research. Finishing this thesis would not have been possible without his continuous support and advice. I was lucky to be Samir's student, as he always generously shared his knowledge that allowed me to grow both as a scientist and as a person.*

*I warmly thank my co-supervisor Pr. Pierre Blazevec for giving me valuable feedback and guidance throughout the PhD studies. I will always remember the kindness he showed me. I also owe my gratitude to Dr. Theodore Lim for supporting me with his valuable advice that increased the quality of this research.*

*I also express my warmest gratitude to the managing director of Hands-Free Computing, Mr. Lawrence Howard for giving me this wonderful opportunity, as the research reported in this thesis would not have been possible without the financial support of Hands-Free Computing Ltd. I owe Lawrence not only my career growth but also I learned a lot, from how he managed people and I will not forget the kindness he showed to the people he worked with. I would like to thank all the personnel of Hands-Free Computing that helped me throughout the years, notably Mr. Neil Judd that helped in conducting a usability study.*

*I would like to thank Dr. Luis Garcia for hosting me in the END-ICAP laboratory, which provided a perfect working environment for carrying out the research.*

*I express my gratitude to Dr. Domitile Lourdeaux and Pr. Rui Cortesao for reviewing this thesis and writing the detailed reports, giving me valuable feedbacks. In addition, I would like to thank other members of the thesis defence jury, Pr. Saida Bouakaz, Pr. Pascal Guitton, Pr. Saeid Nahavandi, and Pr. Jean-Claude Martin for refereeing my thesis. I enjoyed answering their relevant and interesting questions.*

*I also thank Mrs. Marie-Pierre Richoux and Pr. Isabelle Siegler from the Doctoral School who helped me finding my way throughout the administrative procedures.*

*I would like to thank Mrs. Roghayeh Torki for helping me in conducting a usability study that provided valuable user feedback and led to a significant improvement in the designed system.*

*I express my gratitude to my close friends, Hossein Badami, Payman Mehdizadeh, Dr. Mahraz Amini, Mojtaba Hassani, and Dr. Alireza Banisharif for always being available and offering their sincere help, and a special thanks to Dr. Olga Petrova for supporting me during the difficult period of thesis writing.*

*I owe my deepest gratitude to my mother and father for always offering their love and support, and my brothers and sister that graced my life with their kind and compassionate presence.*



# Contents

|  |           |
|--|-----------|
| <b>I. Introduction .....</b>   | <b>3</b>  |
| 1. Importance of Reading .....   | 3         |
| 2. Objective.....  | 3         |
| 3. Organization of the Chapters .....                                    | 6         |
| 4. References .....  | 6         |
| <b>II. Technology-Based Reading Interventions: State of the Art.....</b> | <b>10</b> |
| 1. Introduction .....  | 10        |
| 2. Methodology.....  | 11        |
| 3. Results .....   | 14        |
| 4. Analysis of the Reviewed Research .....                               | 33        |
| 5. Conclusion .....  | 38        |
| 6. References .....  | 39        |
| <b>III. Scientific Approach .....</b>                                    | <b>50</b> |
| 1. Introduction .....  | 50        |
| 2. Objective.....  | 51        |
| 3. Problem Statement.....  | 51        |
| 4. Scientific Ground.....  | 52        |
| 5. System Architecture .....   | 57        |
| 6. The Novelty of the Approach .....                                     | 58        |
| 7. Conclusion .....  | 59        |
| 8. References .....  | 60        |
| <b>IV. Training Model for Automaticity Acquisition .....</b>             | <b>64</b> |
| 1. Introduction .....  | 64        |
| 2. Automaticity .....  | 65        |
| 3. Key Attributes of Automaticity Acquisition .....                      | 66        |
| 4. Automaticity Stages.....  | 76        |
| 5. Proposed Model for Automaticity Acquisition Training.....             | 77        |
| 6. Conclusion .....  | 81        |
| 7. References .....  | 81        |

|   |            |
|---|------------|
| <b>V. Validation of the Gamification Models for Motivation and Automaticity Acquisition ...</b> | <b>92</b>  |
| 1. Introduction .....   | 92         |
| 2. State of the Art on the Gamification and Serious Game Models .....                           | 93         |
| 3. Game Design Based on Motivation Theories .....   | 97         |
| 4. Game Design for Automaticity Acquisition .....   | 105        |
| 5. Conclusion .....   | 111        |
| 6. References .....   | 112        |
| <b>VI. Validation of the Individualized Reading System.....</b>                                 | <b>118</b> |
| 1. Introduction .....   | 118        |
| 2. The Implementation of the Proposed Approach.....   | 119        |
| 3. Methodology and Data Collection.....   | 135        |
| 4. Experiments and Results .....  | 138        |
| 5. Conclusion .....   | 153        |
| 6. References .....   | 154        |
| <b>Conclusion and Perspectives of Research .....</b>  | <b>157</b> |
| <b>Annex.....</b>   | <b>161</b> |

## CHAPTER I

### Introduction

#### 1. Importance of Reading

In modern societies, the role of reading is becoming increasingly crucial. Hence, any impairment to the reading ability such as dyslexia can seriously limit a person's aspirations. Despite the enormous importance of reading skills, still, a considerable percentage of students are left behind. In 2010, a large-scale study evaluated the reading skills of around 800,000 young French adults (age 17 or older) in terms of reading automaticity, vocabulary knowledge, and comprehension [1]. The results showed that 20.4 percent were not effective readers, among them, 5.7 percent had very weak reading skills and another 5.1 percent had severe reading difficulties. The alarming statistics shown in such studies has encouraged many researchers to search for more effective reading acquisition approaches. Technology-based approaches represent a potential solution. However, despite all of the advantages, constructing a complete and effective technology-based reading program is a great scientific and development challenge. It is a multi-disciplinary work that requires the knowledge of educational psychology, linguistics, computer science, and interactive graphics. This complexity is the reason why current approaches did not deeply address one or multiple aspects. Undertaking this challenge is the main purpose of this research.

#### 2. Objective

The objective of this research is *to facilitate and enhance the reading acquisition of opaque orthographies in young dyslexic children*. For meeting this objective, an intelligent gamified in-home approach is proposed, which focuses mainly on the acquisition of automaticity in reading. Home-based approach is proposed to increase the access to instructional content. Intelligent component is developed to enable the home-based approach that lacks the presence of an instructor. Gamification is developed to increase the engagement and adherence to the remedial program. The training program focuses on automaticity acquisition because multiple theories point to the automaticity deficit in dyslexics [2]–[6], as well as the fact that working on automaticity is highly neglected in schools [7]. Each part of this approach is selected to address a current scientific gap, which hinders reaching the objective of this research. Dissecting different facets of this objective can shed light on the research methodology developed in this work. In the following sections, each aspect of the objective is briefly introduced.

##### 2.1. Dyslexia

Dyslexia is a “specific learning disability” that is characterized by problems in accurate and/or fluent reading [8]. It is the most prevalent form of learning difficulties, with researchers differing on its prevalence from 5 to 20 percent [9]. It can have serious negative impacts on a



person's self-esteem and academic aspirations [10]. The most widely accepted theory for the cause of dyslexia is the phonological deficit theory [11]. However, research on dyslexia is extensive and complex, and one single theory such as phonological deficit fails to cover all this complexity [12]. One simple reason is that reading is “a multifaceted construct” [13], and any deficit to each aspect of this complex process can cause a problem in reading. Therefore, it is natural to have different profiles of individuals with reading difficulties. Ultimately, dyslexia is an umbrella term for all of these different profiles and this is the reason multiple scientific theories of dyslexia coexist at the same time. Multiple theories of dyslexia are related to a deficit in automaticity acquisition and procedural learning [2]–[5]. Whether problems in automaticity acquisition are the cause or the consequence of dyslexia, it is open to debate. However, poor automaticity in naming remains as a long-term and universal symptom of dyslexia even for transparent orthographies [14].

## **2.2. Reading Acquisition**

Reading is an automatic skill. When the sight of a known word comes to the visual field of a skilled reader, the word will be immediately read in the brain, and the reader cannot consciously suppress the reading process. Automaticity in reading frees up the cognitive load for higher-order information processing required for comprehension. Acquiring this automatic skill requires extensive practice. However, for maximizing the effect of the practice, the training principles of automaticity acquisition should be considered in the training design process. For example, consistency is paramount in automaticity acquisition. Some orthographies such as English are highly inconsistent, but still, the skilled readers of these orthographies manage to reach automaticity level. This shows that when the consistency does not exist in the letter-sound level, the higher-level consistencies will be used for automatizing the reading process. For example, the letters in “igh” that exist in words such as “high” and “thigh” are not consistent individually, but they are consistent as a group of letters. Sometimes this consistency exists only at the whole word level, and even worse, in case of heteronyms such as present and past tense of the verb “read”, the consistency can only be found in context. For nurturing automaticity, it is required to identify these common consistencies for each language and allow the learners to practice them extensively in a training program designed specifically for automaticity acquisition.

## **2.3. Young Children**

It has been shown that earlier reading interventions are more effective [15], [16], and also, more cost-effective [17]. In case of lack of any effective intervention, a phenomenon similar to Matthew effect will happen [18]. The gap between the good readers and the poor readers becomes wider over time. Good readers feel confident in reading and engage more in reading activities. Hence, their reading skills improve even further and their vocabulary knowledge grows. Poor readers on the other hand, find the reading process hard and laborious, and feel demotivated to get engaged in reading activities, and therefore, they remain further behind. For preventing this gap, early evidence-based interventions are necessary. This is why this study is focused on the development of a remediation system dedicated for young children.

## **2.4. Opaque Orthography**

The remediation of dyslexia is not the only challenge in the process of reading acquisition. Irregularities of some writing systems can create a considerable delay in the acquisition of reading skills [19]. Such writing systems in which the letters often do not correspond to the sounds are called opaque (or deep) orthographies, which is contrasted to transparent (or shallow) orthographies. These irregularities are mostly the consequence of the natural dynamism of all languages, and the inertia of the writing systems to follow the sound changes. Orthographically opaque languages such as English and French lay a heavy burden on learners. It is not only the letters and the corresponding sounds that they should master; they should also master many more common and uncommon graphemes-phoneme associations to become a skilled reader. One computational study found 461 grapheme-phoneme associations in British English [20], and for French language, 346 grapheme-phoneme associations are listed in the Lexique Infra database [21]. These numbers are many more than the 26 letters existing in the writing system of these two languages. This is why some cross-language studies have suggested that reading programs should take into account the characteristics of the language's orthography [22]. Furthermore, in opaque orthographies, the learner frequently faces words that do not follow the simple decoding rules, and for guessing their pronunciation, the vocabulary knowledge and the understanding of the context plays a big role [19]. This means that vocabulary knowledge plays a bigger role in the reading acquisition of an opaque orthography.

## **2.5. Facilitating Reading Acquisition**

The term “facilitate” is used because dyslexics need additional instructional time. Some studies have pointed out that for treating dyslexia the instruction duration should be between 80 to 100 hours, while their healthy peers need 30 to 60 hours [23], [24]. This extra amount of the necessary instruction time is often not provided in schools. Some researchers have suggested that it is not dyslexia, it is “dysfacilitia” [25], highlighting the fact that the current schools do not provide an effective environment for students. The evidence points to the effectiveness of intensive training programs for struggling readers [16], and researchers suggest extensive vocabulary interventions for bridging the gap [26]. These intensive and extensive interventions are not often possible with traditional schools. Ubiquitous technology has the potential to overcome this lack of sufficient instruction. It offers several advantages such as unlimited access to instructional material, the possibility of individualization, and relieving the burden from the shoulders of the teachers. Another big advantage of utilizing ubiquitous technology is the opportunities offered by home-based training. With a cross-platform implementation, the reading system can be accessible to every student at any time and location. For example, the COVID-19 pandemic with unprecedented lockdowns highlighted the importance of such home-based educational tools. However, unlimited access to instruction does not guarantee the engagement and the adherence to training programs. The Gamification concept as a facilitatory mechanism has the potential of allowing positive impact on motivation and the change of attitude [27]. Therefore, a technology-based gamified in-home approach was developed in this study.

## 2.6. Enhancing Reading Acquisition

The term “enhance” is used to emphasize the fact that the approach needs to be based on the evidence in the scientific literature. There is a history of ineffective approaches in the instruction of reading, and it went as far that some researchers used the term “dysteachia” instead of dyslexia [28]. However, not all the blame could be put on the teachers. For example, scientific evidence suggests that reading intervention should be systematic [29]; however, being systematic in the instruction of an opaque orthography with its numerous grapheme-phoneme associations is an onerous task to expect from teachers who have limited time and linguistic knowledge. For this reason, we suggested in this research to use an intelligent computation model with access to linguistic databases in order to provide a truly systematic approach in the instruction of reading.

## 3. Organization of the Chapters

The following chapters of this dissertation are organized as follows. The next chapter provides an in-depth and analytical look into the current state of the art of the technology-based or technology-assisted reading programs. The third chapter lays out the detailed scientific approach for reaching the objective of this research, it is focused on the research gaps and the disadvantages of the current approaches identified in the second chapter. The fourth chapter provides a comprehensive overview of automaticity training principles, as well as proposing a model for designing training programs that target automaticity acquisition. The fifth chapter proposes two gamification models, the first model is based on motivation theories of self-determination and self-efficacy and the second model is based on the automaticity training model proposed in chapter 4. Two experimental studies are carried out and the results are reported to evaluate the efficiency of these gamification models. Chapter 6 provides a detailed description about the implementation of the remediation system. In addition, it reports on the results of iterative usability studies carried out. Furthermore, this chapter provides linguistic analysis on word difficulty using artificial neural networks and multiple linear regression models. The final chapter summarizes the main findings of the research carried out and suggests the future research perspectives.

## 4. References

- [1] F. De La Haye, J.-E. Gombert, J.-P. Rivière, and T. Rocher, “Les évaluations en lecture dans le cadre de la Journée Défense et Citoyenneté. Année 2010,” 2011.
- [2] R. I. Nicolson and A. J. Fawcett, “Development of Dyslexia: The Delayed Neural Commitment Framework,” *Frontiers in Behavioral Neuroscience*, vol. 13, p. 112, May 2019, doi: 10.3389/fnbeh.2019.00112.
- [3] R. I. Nicolson, A. J. Fawcett, R. L. Brookes, and J. Needle, “Procedural learning and dyslexia,” *Dyslexia*, vol. 16, no. 3, pp. 194–212, Jul. 2010, doi: 10.1002/dys.408.
- [4] L. Blomert, “The neural signature of orthographic–phonological binding in successful and failing reading development,” *Neuroimage*, vol. 57, no. 3, pp. 695–703, 2011.
- [5] R. I. Nicolson and A. J. Fawcett, “Automaticity: A new framework for dyslexia research?,” *Cognition*, vol. 35, no. 2, pp. 159–182, May 1990, doi: 10.1016/0010-0277(90)90013-A.
- [6] L. Blomert and D. Froyen, “Multi-sensory learning and learning to read,” *International journal of psychophysiology*, vol. 77, no. 3, pp. 195–204, 2010.

- [7] H. Jamshidifarsani, S. Garbaya, T. Lim, P. Blazevic, and J. M. Ritchie, "Technology-based reading intervention programs for elementary grades: An analytical review," *Computers & Education*, vol. 128, pp. 427–451, Jan. 2019, doi: 10.1016/j.compedu.2018.10.003.
- [8] G. R. Lyon, S. E. Shaywitz, and B. A. Shaywitz, "A definition of dyslexia," *Ann. of Dyslexia*, vol. 53, no. 1, pp. 1–14, Jan. 2003, doi: 10.1007/s11881-003-0001-9.
- [9] R. K. Wagner *et al.*, "The Prevalence of Dyslexia: A New Approach to Its Estimation," *J Learn Disabil*, p. 002221942092037, May 2020, doi: 10.1177/0022219420920377.
- [10] E. Lithari, "Fractured academic identities: dyslexia, secondary education, self-esteem and school experiences," *International Journal of Inclusive Education*, vol. 23, no. 3, pp. 280–296, Mar. 2019, doi: 10.1080/13603116.2018.1433242.
- [11] M. Snowling, "Dyslexia as a phonological deficit: Evidence and implications," *Child Psychology and Psychiatry Review*, vol. 3, no. 1, pp. 4–11, 1998.
- [12] A. Castles and N. Friedmann, "Developmental Dyslexia and the Phonological Deficit Hypothesis," *Mind Lang*, vol. 29, no. 3, pp. 270–285, Mar. 2014, doi: 10.1111/mila.12050.
- [13] E. Reynor, "How important is language development for skilled reading? Exploring the relationship between language processes and reading skills for children with dyslexia," *REACH Journal of Special Needs Education in Ireland*, vol. 31, no. 1, pp. 38–53, 2018.
- [14] S. Araújo and L. Faisca, "A Meta-Analytic Review of Naming-Speed Deficits in Developmental Dyslexia," *Scientific Studies of Reading*, vol. 23, no. 5, pp. 349–368, Sep. 2019, doi: 10.1080/10888438.2019.1572758.
- [15] B. R. Foorman, D. J. Francis, S. E. Shaywitz, B. A. Shaywitz, and J. M. Fletcher, "The case for early reading intervention," in *Foundations of reading acquisition and dyslexia: Implications for early intervention*, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 1997, pp. 243–264.
- [16] J. Wanzek, E. A. Stevens, K. J. Williams, N. Scammacca, S. Vaughn, and K. Sargent, "Current Evidence on the Effects of Intensive Early Reading Interventions," *J Learn Disabil*, vol. 51, no. 6, pp. 612–624, Nov. 2018, doi: 10.1177/0022219418775110.
- [17] R. I. Nicolson, A. J. Fawcett, H. Moss, M. K. Nicolson, and R. Reason, "Early reading intervention can be effective and cost-effective," *British Journal of Educational Psychology*, vol. 69, no. 1, pp. 47–62, 1999, doi: 10.1348/000709999157563.
- [18] K. Cain and J. Oakhill, "Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development," *Journal of learning disabilities*, vol. 44, no. 5, pp. 431–443, 2011.
- [19] J. C. Ziegler and U. Goswami, "Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory," *Psychological bulletin*, vol. 131, no. 1, p. 3, 2005.
- [20] P. F. D. Gontijo, I. Gontijo, and R. Shillcock, "Grapheme—phoneme probabilities in British English," *Behavior Research Methods, Instruments, & Computers*, vol. 35, no. 1, pp. 136–157, Feb. 2003, doi: 10.3758/BF03195506.
- [21] M. Gimenes, C. Perret, and B. New, "Lexique-Infra: grapheme-phoneme, phoneme-grapheme regularity, consistency, and other sublexical statistics for 137,717 polysyllabic French words," *Behav Res*, May 2020, doi: 10.3758/s13428-020-01396-2.
- [22] M. Aro and H. Wimmer, "Learning to read: English in comparison to six more regular orthographies," *Applied psycholinguistics*, vol. 24, no. 4, p. 621, 2003.
- [23] F. R. Vellutino, J. M. Fletcher, M. J. Snowling, and D. M. Scanlon, "Specific reading disability (dyslexia): What have we learned in the past four decades?," *Journal of child psychology and psychiatry*, vol. 45, no. 1, pp. 2–40, 2004.
- [24] H. Lyytinen, J. Erskine, M. Aro, and U. Richardson, "Reading and reading disorders.," 2007.
- [25] D. Thornburg, *From the campfire to the holodeck: Creating engaging and powerful 21st century learning environments*. John Wiley & Sons, 2013.
- [26] W. Nagy, "Why vocabulary instruction needs to be long-term and comprehensive," in *Teaching and learning vocabulary: Bringing research to practice*, Routledge, 2005, pp. 27–44.
- [27] S. Dreimane, "Gamification for Education: Review of Current Publications," in *Didactics of Smart Pedagogy*, L. Daniela, Ed. Cham: Springer International Publishing, 2019, pp. 453–464.

- [28] E. B. Hurst and M. R. Hurst, *Why Can't My Son Read?: Success Strategies for Helping Boys with Dyslexia and Reading Difficulties*. Sourcebooks, Inc., 2014.
- [29] L. C. Ehri, S. R. Nunes, S. A. Stahl, and D. M. Willows, "Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis," *Review of educational research*, vol. 71, no. 3, pp. 393–447, 2001.

## Contents

|  |    |
|--|----|
| <b>CHAPTER II Technology-Based Reading Interventions: State of the Art</b> ..... | 10 |
| <b>1. Introduction</b> .....   | 10 |
| <b>2. Methodology</b> .....  | 11 |
| 2.1. Literature Research Methodology .....                                       | 11 |
| 2.2. Inclusion Criteria.....   | 13 |
| 2.3. Coding Procedure.....   | 13 |
| <b>3. Results</b> .....  | 14 |
| 3.1. Phonological Awareness .....  | 14 |
| 3.2. Phonics .....   | 16 |
| 3.3. Vocabulary .....  | 20 |
| 3.4. Reading Comprehension .....   | 21 |
| 3.5. Fluency .....   | 24 |
| 3.6. Multi-Component .....   | 27 |
| 3.7. Additional Research .....   | 29 |
| <b>4. Analysis of the Reviewed Research</b> .....                                | 33 |
| 4.1. Measures.....   | 33 |
| 4.2. Interventions.....  | 34 |
| 4.3. Duration of Interventions .....   | 35 |
| 4.4. Technologies .....  | 36 |
| 4.5. Grade Levels .....  | 36 |
| 4.6. Game-Based and Gamified Interventions .....                                 | 37 |
| 4.7. Languages.....  | 37 |
| 4.8. Study Limitations .....   | 38 |
| <b>5. Conclusion</b> .....   | 38 |
| <b>6. References</b> .....   | 39 |

## CHAPTER II

### Technology-Based Reading Interventions: State of the Art

#### 1. Introduction

Reading is the essence of academic life, and its importance can hardly be overestimated [1]. Its impairment can cause a life-long disability that affects the quality of life in numerous ways. However, reading is a complex and multifaceted process and can be challenging for some individuals to master [2]. Despite all of the attempts to raise reading instruction standards over the years, many students still fail to achieve grade-level reading when they reach the upper elementary grades. This achievement gap tends to widen during the following grades [3], [4]. The report from the National Assessment of Educational Progress [5] shows that 64% of fourth-graders and 66% of eighth-graders read below their grade level. The case for children with learning disabilities is much worse (88% Grade 4 and 92% Grade 8).

The National Reading Panel [6] has identified five core components essential for a comprehensive reading program: **phonological awareness**, **phonics**, **vocabulary**, **comprehension**, and **fluency**. However, the conventional methods of instructing these components are truly time-consuming, and to be effective, they need to be carried out intensively and explicitly by an instructor [7]. Some studies have pointed out that for treating dyslexia, the instruction duration should be between 80 to 100 hours, while their healthy peers need 30 to 60 hours [8]. Furthermore, the growing public awareness of learning disabilities and governments' inclusion policies of disabled learners, as well as the emergence of new ubiquitous technologies, have encouraged many researchers to investigate and propose innovative, more engaging, and more effective approaches for facilitating the literacy acquisition of young learning disabled children [9], [10].

Incorporating technology into instructional intervention can have several benefits. First of all, learning in a playful and engaging digital environment can increase motivation, leading to enhanced acceptance, concentration, and persistence in learning [11], [12]. The second benefit can be the capacity of technology-based instructions in reducing the cognitive load and increasing the retention [13]–[15]. Third, it can provide personalized and adaptive tutoring with no or reduced instructor involvement, which is truly beneficial when there are not enough human resources available [16], [17]. Finally, without an instructor's time limit, it can allow the users to reach mastery levels by letting them train at their own pace [18].

There are some studies in the literature that tried to review the technology-assisted approaches in teaching literacy. MacArthur et al. [19] wrote a critical review of 14 studies applying technology to literacy instruction for school-age students with literacy problems, from 1985 to 2000. Blok et al. [20] reviewed 42 studies of computer-assisted reading instruction for early literacy learners, from 1990 to 2000. Cheung and Slavin [21] provided a meta-analysis of k-12 educational technology's effectiveness on reading achievement, using 85 studies published from 1970 to 2010. Grant et al. [22] assessed the content and quality of 30 commercially available reading software for preschool, kindergarten, and first grade. Edwards Santoro and Bishop [23] evaluated 21 popular beginning reading software targeting pre-

kindergarten to third grade, based on their interface design, instructional design, and content. Cidrim and Madeiro [24] reviewed 21 studies that applied information and communication technology (ICT) to dyslexia, from 2010 to 2015.

This chapter is dedicated to the technology-based intervention programs for reading instruction of elementary grades. This chapter has the following objectives:

- Provide a comprehensive review of studies which applied technology to their reading intervention, from the year 2000 to 2017
- Introduce the reading components of phonological awareness, phonics, comprehension, fluency, and vocabulary, as well as their common instructional approaches
- Describe the content and instructional mechanisms of the identified programs, to provide a basis for researchers and developers new to this field
- Analyze reviewed studies from various aspects.

In order to study the state of the art and analyze the published results, forty-two studies have met the inclusion criteria and 32 technology-based reading programs have been identified. This chapter is structured as follows. Section 2 describes the methodology used to create this review chapter. A brief introduction of each reading component, and description of intervention programs addressing that reading component, as well as the details and characteristics of the reviewed studies, are brought in Section 3. An analytical review is presented in Section 4. Conclusion and the future research directions is presented in Section 5.

## 2. Methodology

This section is dedicated to the methodology followed in this review chapter. The literature research procedure is outlined, then the inclusion criteria is presented and the coding procedure used for the reviewed published research is discussed.

### 2.1. Literature Research Methodology

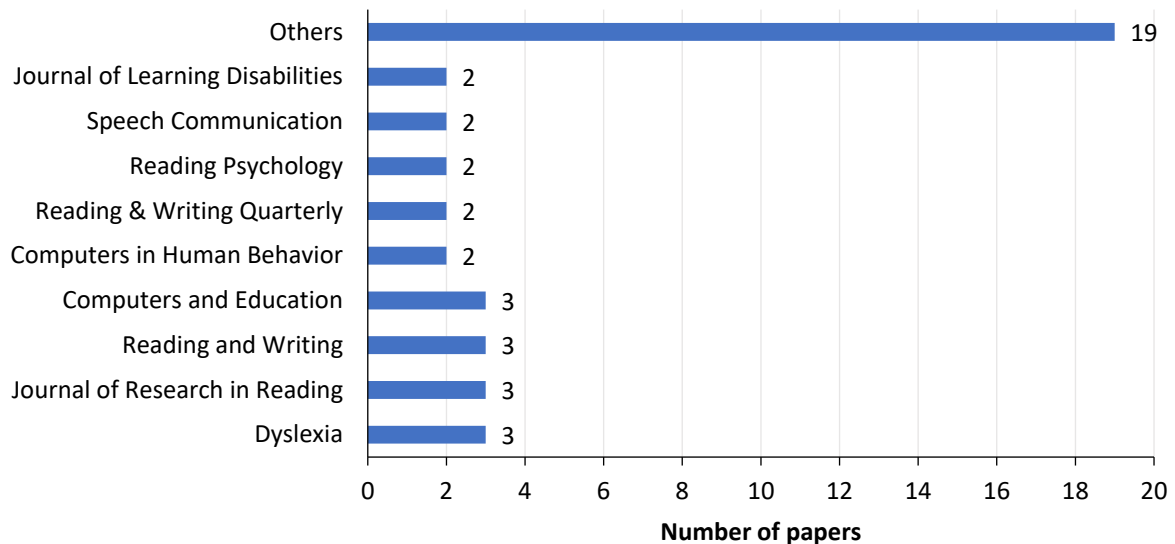
This research was conducted using the Google Scholar database. Various keywords were used to find targeted papers. The keywords are presented in table 1. Combinations of these keywords were used to search for the targeted articles. Based on this research, 187 studies were selected for further investigation, and from these set of articles, 42 have met all of the inclusion criteria, which resulted in 32 different intervention programs. However, in the end, when it appeared that only one study focused on vocabulary, another search specifically for vocabulary interventions was carried out to ascertain that the result is not biased by the used keywords. Surprisingly, no additional vocabulary study that meets the inclusion criteria was found.

Finally, 41 articles were identified for this review. Figure 1 depicts the distribution of published papers based on their journals. Category *Others* indicates the number of journals from which only one study was included in this review. Note that *Computers and Education*, *Reading and Writing*, *Journal of Research in Reading*, and *Dyslexia* are the most represented journals. This figure shows a high dispersion of the publications over journals dealing with different research domains such as education, psychology, and technology.



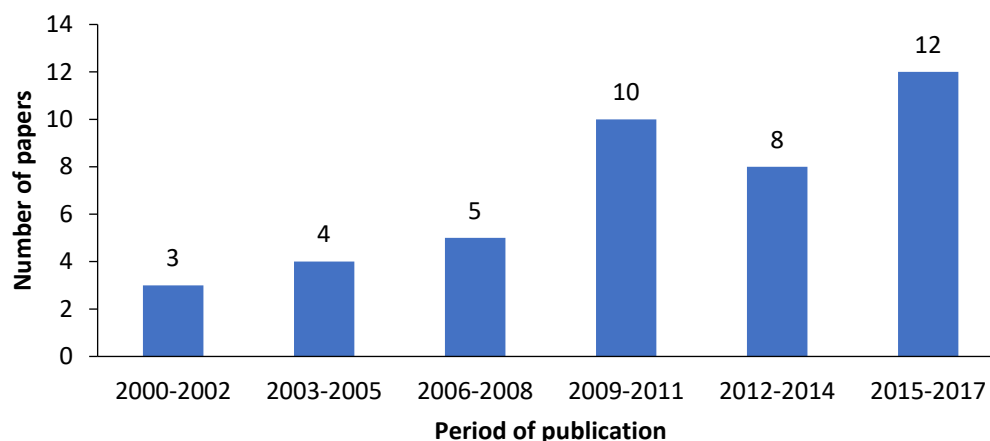
**Table 1.** Keywords used for the research procedure

| Category 1   | Category 2  | Category 3   | Category 4  |
|--|---|--|---|
| <ul style="list-style-type: none"> <li>• Technology-Assisted</li> <li>• Technology-Based</li> <li>• Computerized</li> <li>• Computer-Assisted</li> <li>• Computer-Based</li> <li>• Tablet</li> <li>• Mobile</li> <li>• Smartphone</li> <li>• Virtual Reality</li> <li>• Augmented Reality</li> </ul> | <ul style="list-style-type: none"> <li>• Reading</li> <li>• Literacy</li> <li>• Dyslexia</li> </ul> | <ul style="list-style-type: none"> <li>• Intervention</li> </ul> | <ul style="list-style-type: none"> <li>• Elementary</li> <li>• Primary</li> </ul> |



**Figure 1.** Distribution of reviewed papers by journals

Figure 2 summarizes the reviewed publications and shows their distribution over periods of three years. This figure shows the growing interest in using technology to remediate the reading difficulties of early readers. Given the importance of the matter and the fact that various technologies with the potential to be utilized for educational purposes have become more accessible and ubiquitous over time, it is not surprising to observe this upward trend.



**Figure 2.** Distribution of reviewed papers by year

## 2.2. Inclusion Criteria

The principal inclusion criteria for this review are:

- All participants should be in elementary grades or between the ages of 6 to 12 years old.
- The purpose of intervention completely or partially should be about improving reading acquisition.
- The intervention should be based on or assisted by technology.
- The intervention should be carried out on the reading acquisition of the first language, and studies focusing on second language learning are excluded.
- The intervention should be based on explicit reading instructional approaches.
- The study should include at least five participants in the intervention.
- The study should be published between 2000 and 2017.

## 2.3. Coding Procedure

In Table 2, the characteristics of the reviewed studies are presented. Articles were categorized based on their intervention types aligned with the National Reading Panel [6], which means the main categories were phonological awareness, phonics, vocabulary, comprehension, and fluency. In addition, the multi-component category has been added to represent the intervention programs that target multiple of these key reading categories. Furthermore, the phonological awareness training programs that heavily involved written letters were put in the phonics category. For the design type of the studies, they are classified into three categories of Treatment vs. Comparison groups (TC), Multi-Treatment groups (MT), and Single-Treatment group (ST). In addition, the number of participants in each treatment or control group is written separately. A comma separates multiple treatment or control groups, and groups from multiple experiments in one study are separated with a dash. The groups that received traditional teaching instruction were considered as control groups. Finally, if the total time of intervention (hours) was not explicitly mentioned in the study, it was estimated by

simply multiplying the number of weeks, frequency of sessions per week, and duration of each session (if available).

### 3. Results

Overall, 42 studies have met the inclusion criteria. These studies' reading programs are classified based on their intervention types, including phonological awareness, phonics, vocabulary, reading comprehension, fluency, and multi-component. Each category begins with a brief introduction to each intervention type without considering the use of technology. Then, each technology-based intervention program existing in that category is described alongside the interventions' results. Moreover, additional research that did not meet the inclusion criteria is discussed at the end of this chapter. The important details and characteristics of these studies are brought in Table 2.

#### 3.1. Phonological Awareness

Phonological awareness (PA) is “the understanding of different ways that oral language can be divided into smaller components and manipulated”[25]. That means having the capabilities such as isolating, identifying, segmenting, blending, deleting, adding, or substituting the sounds of the smaller units of language such as word, syllable, onset, rime, and individual phonemes. Phonemic awareness is one of the building blocks of phonological awareness, which is the ability to attend to and manipulate individual phoneme sounds. The other components of phonological awareness are syllable awareness and onset-rime awareness [26]. In the literature, sometimes the term phonemic awareness has been used to signify phonological awareness.

Phonological awareness at early ages has been proven to be a strong predictor of reading proficiency at later years [27], [28]. Furthermore, one of the most widely accepted theories of dyslexia describes this disability as a phonological deficit disorder [29], [30]. Particularly, the phonemic awareness subset is identified as the key to reading success [31]. According to the reports [6], teaching phonological awareness in small groups is more effective than the classroom. Additionally, focusing on one or two phonological awareness skills throughout the intervention will result in larger effect sizes than teaching three or more. Furthermore, it is suggested that involving written letters to manipulate phonemes, will result in better outcomes, especially for older children [32].

The technology-based interventions that target phonological awareness are summarized, and their effects are briefly mentioned below.

##### 3.1.1. LiPS

Torgesen, Wagner, Rashotte, Herron, & Lindamood (2010) used *Lindamood Phoneme Sequencing Program for Reading, Spelling, and Speech (LiPS)* [34], which aims at improving the phonemic awareness of children explicitly by teaching them the articulatory gestures of different phonemes. The user could also do other activities such as tracking phonemes in words to reinforce their phonemic awareness by using mouth-form images, color blocks, and letters associated with different phonemes. Once children's phonemic awareness skills are solidified,

they can go on with advanced activities such as reading and spelling. Throughout the immediate post-test and the follow-up, the intervention group performed significantly better on measures of phonemic awareness, phonemic decoding, and rapid automatized naming. Pokorni et al. [35] have also examined LiPS and found significant PA gains, especially for blending and segmenting phonemes, but no transfer to reading or language measures was observed. Finally, it was more effective than the other treatment groups who used Fast ForWord or Earobics.

### **3.1.2. Fast ForWord**

Pokorni et al. (2004) evaluated three different literacy programs focusing on phonological awareness teaching. One of those programs was *Fast ForWord* (FFW) [36], an instructional program based on the internet and CD-ROM, which includes various activities using acoustically processed speech and speech sounds. These activities target various skills, such as phoneme discrimination, listening comprehension, working memory, and auditory word recognition. The gains of this intervention were limited to phonological awareness, but no and a transfer to reading or language measures was not found. It was also less effective than the other two treatment groups that used other PA programs called *LiPS* and *Earobics*. Cohen et al. [37] tested FFW with children suffering from severe mixed receptive-expressive specific language impairment, but it showed no more benefits than the comparison groups. The effectiveness of FFW has been evaluated throughout many other studies. A meta-analysis on its effectiveness concluded that there is no evidence that FFW is an effective treatment for children's oral language or reading difficulties [38].

### **3.1.3. Earobics**

Pokorni et al. [35] evaluated a gamified phonological awareness program called *Earobics* [39], which had two sets of gamified activities for teaching Phonological Awareness systematically. However, just the second set of activities were tested in that study, consisting of games for auditory memory, sound recognition, segmenting sounds, blending sounds, discrimination of vowel and consonant sounds, and recognizing word endings and beginnings. However, the intervention gains were limited to phonological awareness, especially in segmenting phonemes, and a transfer to language or reading measures was not found.

### **3.1.4. PLAY-ON**

Magnan et al. [40] tested a computer software called *PLAY-ON* [41], consisting of several gamified phonological awareness training activities. In their study, they evaluated only one audio-visual exercise of the software, called basket game, which aims to help children discriminate the sounds of similar phoneme pairs such as /p/-/b/, /t/-/d/, and /k/-/g/. The sound of a CV syllable (e.g., /ba/) is played to participants, and then after they listened to the sound, a basketball falls from the top of the screen and then, they should choose the basket with the right orthographical representation of it (ba or pa). After training with this game, they found positive effects on a word recognition test. Their result was consistent with studies suggesting that phonological awareness training with letters is more effective than speech-only approaches [32].

### 3.1.5. COMPHOT

Gustafson et al. [42] evaluated a computerized phonological training program named *COMPHOT* [43]. It includes four sections, which are Rhyme (four exercises), Position (eight exercises), Addition (five exercises), and Segmentation (three exercises). The exercises are mainly phonological and sound-based, along with many pictures and limited use of written letters and words. Whenever the child clicks on a picture, the corresponding word will be played by a natural, recorded voice. An example task is when the participants hear a word, and then they should choose the picture that rhymes with the heard word, or for instance, they should choose a (picture that has the same initial phoneme as the heard word. In addition, game-like elements such as showing high scores are incorporated into exercises. After the intervention, the post-test results revealed that it had large effects on reading comprehension, word decoding, and sight word reading. It also had a large to moderate effect on passage comprehension and a moderate to small effect on pseudoword reading. However, post-test and follow-up results showed that intervention was more effective and persistent when it was accompanied by reading comprehension instruction [44]. Furthermore, another intervention study on COMPHOT revealed large effects on word decoding and text reading and moderate to large effects on phonological awareness [45].

## 3.2. Phonics

Phonics refers to “various approaches designed to teach children about the orthographic code of the language and the relationships of spelling patterns to sound patterns” [46]. It is recommended to use a systematic and explicit phonics approach for teaching early literacy skills [6]. Also, it has been indicated that interventions relying heavily on phonics are more effective for dyslexia remediation [47]. There are different phonics approaches, including synthetic phonics (blended phonics), analytic phonics, embedded phonics, analogy phonics, onset-rime phonics, and phonics through spelling [48].

Synthetic Phonics, which is the most widely accepted approach in English-speaking countries, is “an approach to the teaching of reading in which the phonemes associated with particular graphemes are pronounced in isolation and blended together (synthesized)” [49], [50]. However, in analytic phonics, phonemes associated with graphemes are not pronounced in isolation, and letter-sound associations are taught after the word has been recognized [48], [50]. Embedded Phonics is an implicit approach that teaches letter-sound relations through the context of reading comprehension [51], [52]. Analogy phonics is an approach that uses parts of already learned words to acquire and decode new words [48]. Onset-rime phonics, as its name suggests, is learning letter-sound associations through detecting the sound of a letter or a cluster of letters before the initial vowel (onset) and the sound of the rest of the word (rime) [50]. Finally, in phonics through spelling, students learn letter-sound associations by segmenting words into phonemes and writing the individual phonemes' letters to build words [48].

Technology-based phonics interventions found in this review are summarized below, alongside their intervention outcomes.

### 3.2.1. Lexia

Macaruso et al. [53] used a software program called *Lexia* [54] for their Literacy teaching intervention. It consists of two different programs: Phonics Based Reading (PBR) and Strategies for Older Students (SOS). PBR consists of 3 levels, 17 skill activities, and 174 units. Phonics-based activities were structured and systematic. They include multisensory tasks like audio-visual matching and kinesthetic responses. The teacher has to set the initial level of the program to prepare the student to work independently on the activities. After finishing PBR activities, students can go on to SOS activities with five levels, 24 skill activities, and 369 units. It starts by building on the user's phonics knowledge, and it advances throughout the levels. Finally, the intervention results showed that children in the treatment group had improved their reading skills more than the control group, but the difference was not significant. However, if the comparison is limited to 'at risk' students, the treatment group's improvement is significantly more than the control group.

### 3.2.2. GraphoGame

In their computer-assisted reading intervention, Saine, Lerkkanen, Ahonen, Tolvanen, & Lyytinen (2011) used a phonics-based Finnish program called *GraphoGame* (also known as *Ekapeli* or *Literate*) [56]–[58]. It consists of gamified practices ranging from pre-reading to fluency. Its primary objective is to build an automatic phonological-orthographical binding by focusing on matching speech sounds to their written counterparts [59]. It starts with letter-sound relations, and then it progresses to the syllable level and the word and pseudoword levels. It is an adaptive program, and it adjusts the difficulty level of the activities to the user's performance. This study showed significant gains in letter knowledge, decoding, accuracy, fluency, and spelling.

In another study, Kyle et al. [60] compared GraphoGame Phoneme (GG Phoneme) and Graphogame Rime (GG Rime) which are two components of this program in English. Each component, systematically trains the integration of speech sounds to their written forms, though, one at the phoneme level and the other at the rime level. Both interventions led to significant gains in reading, spelling, and phonological skills. However, the effect size did not differ significantly between the two interventions. Furthermore, Rosas et al. [61] evaluated the GraphoGame program with Spanish speaking children from low and high Socio-Economic Status (SES). Children with low SES showed improved letter-sound knowledge, while the high SES children in the treatment group improved their Rapid Automatized Naming (RAN). However, no significant improvement was found in word reading, pseudoword reading, and phonological awareness for low and high SES children in the treatment group.

### 3.2.3. Phonological Analysis

Wise et al. [62] used a computer-assisted intervention named *Phonological Analysis*. It consists of four sub-programs called Phonological Analysis with Letters, Nonword Choice, Marvin, and Spello. Phonological Analysis with Letters is designed to help children practice and learn letter-sound relations. Nonword Choice is an exercise in which the computer pronounced a non-word, and then children should choose the right non-word between multiple choices that matched the pronunciation. Marvin is a similar task in which an animated mouth

pronouncing a non-word is shown, and then the user should decide whether it matched the displayed non-word or not, and if it did not match, what was the difference. Spello is a task in which the user should enter the right spelling of a word after hearing its pronunciation by the computer. In the process, the user can listen to the pronunciation of the entered word at any time to see if it matches the one pronounced by the computer. It aims at teaching the orthographical-phonological relations to children. All the intervention tasks are adaptive by automatically increasing or decreasing the difficulty level, depending on the user's performance. They compared this program with a computerized reading comprehension program (Accurate Reading in Context). The results showed that the Phonological Analysis program was more effective at improving phonological skills. Furthermore, the effect remained significant after the two-year follow-up. However, for word decoding measures, the difference between the two programs was not significant.

#### **3.2.4. ABRACADABRA**

Savage et al. [63] investigated *ABRACADABRA*, a web-based tool for improving literacy that can be accessed freely. *ABRACADABRA*'s flexibility and customizability allowed them to test and compare the effectiveness of two different approaches to phonics, which are Synthetic Phonics and Analytic Phonics. Synthetic Phonics intervention is aimed at building the skills for blending and segmenting words at the phoneme level. Students were introduced to six letter-sounds each week. Then they were able to build on their acquired letter-sound knowledge for developing the blending and segmenting skills by doing these activities: "Auditory Blending (blending sounds and choosing a matching picture); Blending Train (identifying a word by blending its letter sounds); Basic Decoding (sounding out and reading words); and Auditory Segmenting (matching words to their segmented sounds)." Each of These activities contained different difficulty levels, and as students progressed, the more demanding levels were introduced to them.

Analytic Phonics is aimed at improving skills in distinguishing and manipulating the onset and rime units of words. Letter-sound associations were presented to students at a slow pace, to allow them enough time to practice and learn sound patterns thoroughly. This intervention approach consists of several activities which are: "Same Word (identifying similar words based on their sound); Word Matching (matching word cards by their beginning sounds); Rime Matching (matching words that rhyme); Word Families (making words from the same word family by changing the first letter); and Word Changing (manipulating the onsets or other letters of words in rhyme families to form a new word)". Finally, this study revealed that both of the interventions had significant impacts on literacy scores. Analytic phonics had a marked impact on letter-sound knowledge, but synthetic phonics influenced more phonological awareness skills in both post-tests and fluency in the second post-test. In 2010, another study evaluated different implementation styles of *ABRACADABRA*. They concluded that the Adaptation group, which applied technology to broader learning themes, benefited the most from the intervention [64]. In 2015 a meta-analysis investigated *ABRACADABRA*'s effectiveness, and they found out that it was more effective on phonemic awareness, phonics, listening comprehension, and vocabulary. However, it was less effective for fluency and reading comprehension [65].

### 3.2.5. RITA

Nicolson et al. [66] presented and evaluated a computer-assisted reading program for low readers, called *Reader's Interactive Teaching Assistant* (RITA). The teacher has a central role in this program by choosing and planning children's activities. It used an alphabetically arranged pad of buttons as an input, and output was in the form of text, picture, graphics, and synthesized or digitized speech. It contained various phonics activities from individual phonemes to the word level. All of these activities, along with a computerized version of books and sounds, are available in the Resource Library component. Significant improvement in standard literacy scores was observed in the RITA group compared to the control group. However, compared to the other group who received traditional instruction, it was slightly less effective.

### 3.2.6. Trainertext

Messer and Nash [67] examined a computer system called *Trainertext* [68], which uses visual mnemonics to teach grapheme-phoneme relations. This method incorporates embedded picture mnemonics that illustrate an object whose name starts with the target letter or depict an object whose shape is similar to the target letter. In this study, for each English phoneme, a picture representing the phoneme is shown above it. For example, for the phoneme /a/ in the word 'gas', the Ant in Pink Pants' visual mnemonics is shown. Therefore, wherever the children struggle to decode, they can click on the letter to see the relevant visual mnemonics. As the children progress in decoding, the program increases the number of words that should be read to establish the mastery of letter-sound knowledge. To increase children's motivation, the system proposes some simple decoding related games, before and after the main activities. The intervention results showed that treatment had significant positive impacts on decoding, phonological awareness, naming speed, phonological short-term memory, and executive loaded working memory. However, they failed to find any meaningful effect on spelling.

### 3.2.7. DOT

Gustafson et al. [45] examined a computerized orthographic training program, named *DOT* [69], which includes four different sections containing eleven exercises. These sections are word reading (four exercises), text reading (two exercises), word parts (two exercises), and building words (three exercises). The exercises are heavily based on written letters, morphemes, words, and texts and the link to their sounds. Therefore, users can click on written forms, and the computer sounds them out. Besides, game-like elements such as high score and happy or sad auditory response feedback are included in the program. The intervention results showed an improvement in reading-related skills, notably large effects on word decoding and text reading and moderate to large effect on phonological awareness. However, there was no statistically significant superiority to the comparison group who received ordinary special instruction.



### 3.2.8. Chassymo

Ecalte et al. [70] ran experiments involving a software called *Chassymo* [71], which focuses on training the grapho-syllabic relations in words. The student hears a syllable, then, 500 ms later, the written form of the syllable will be displayed, and then, 500 ms later, the sound of a word will be played. Now the student should choose if the syllable was present in the heard word or not, and if it was present, what was its position in the word (initial, median, or final). Corrective feedback is displayed after the trial by showing the word and highlighting the syllable in green. Its training set consists of 600 bi-syllabic and tri-syllabic words. The results of this study revealed that this grapho-syllabic intervention was more effective than a grapho-phonemic program, on measures of silent and aloud word reading, as well as reading comprehension.

### 3.2.9. Oppositions Phonologiques

Ecalte et al. [70] used a piece of software called *Oppositions Phonologiques* [72] in their reading intervention study. It aims at teaching the grapheme-phoneme relations by focusing on phonological oppositions such as p/b, t/d, and m/n. Ten phonological oppositions are chosen for the program, and for each one, 50 pairs of words are included. The child is presented with the pair of written words differing in one phoneme, and then, one of the words is played, and the child should decide which of the two words was heard. Then, the corrective feedback was displayed by highlighting the correct word in green. This intervention study's outcome data revealed that this grapho-phonemic approach was less effective on word reading and comprehension than a grapho-syllabic approach used by another treatment group.

### 3.2.10. 8 Great Word Patterns

Moser et al. [73] evaluated a software application called *8 Great Word Patterns* [74], which teaches the common word structures and patterns. It has eight levels of instruction, containing 88 lessons. It focuses on one-syllable words, and it teaches students the common patterns of consonants, vowels, digraphs, and morphemes in words. It allows the students to manipulate word structures by using activities such as blending, segmenting, substituting, and sequencing. After finishing these lessons, children went on to practice word identification in connected text. Finally, the intervention outcome showed no significant difference in the rate and accuracy of oral reading and motivation to read between the intervention and the comparison groups. However, the intervention group outperformed the control group on measures of spelling, vocabulary, and comprehension.

## 3.3. Vocabulary

Vocabulary is “the knowledge of meanings of words” [75]. However, vocabulary knowledge is not only knowing the definition of the words but also knowing how they fit into the world [76]. The correlational relationship between vocabulary and comprehension has long been established. Many studies have shown that vocabulary size at an early age is a strong predictor of reading comprehension later on [77]–[79].

Vocabulary size varies in individuals; even before entering the school, the gap between children can be large [80], [81], and unfortunately, this gap tends to become larger as they grow

[82], [83]. The kids who have larger vocabulary knowledge have better reading comprehension and tend to engage themselves more in reading activities, and hence, they will acquire more vocabulary through incidental acquisition. On the other hand, children with poor vocabulary knowledge avoid extensive, independent reading activities because of their poor comprehension, and thus, they acquire fewer new words [84]. Therefore, it is vital to help students extend their vocabulary knowledge at an early age through long-term and comprehensive instruction [84], [85].

Vocabulary can be classified into four categories of listening, speaking, reading, and writing. The first two construct the spoken vocabulary, and the latter two form the written vocabulary. In addition, vocabulary knowledge can be divided into two categories of receptive and productive. Receptive vocabulary is those that a person can recognize through listening or reading, while productive vocabulary is the words that one can utilize during speaking or writing. Finally, sight vocabulary is a subcategory of reading vocabulary that does not require explicit word decoding [6]. From the effective approaches for teaching vocabulary, we can mention direct or explicit instruction, using multimedia methods, teaching mnemonic strategies, and instructing morphemic analysis [86]. Indirect instruction or simply encouraging children to engage themselves in extensive independent reading or reading aloud to them, which causes the incidental acquisition of new words, is a vital part of vocabulary learning [87]. Effective instructions rely highly on multimedia aspects, the richness of context in which vocabulary is learned, active engagement of children, and multiple exposures to words [6].

The section below presents a summary of the vocabulary interventions found in the literature. Surprisingly, only one paper incorporating technology in instructing first language vocabulary was found.

### **3.3.1. The Great Quake of '89,**

Xin and Rieth [88] evaluated video-assisted vocabulary instruction by using a videodisc called *The Great Quake of '89*, which is created by ABC News and focuses on the 1989 San Francisco earthquake. It contains an hour of video content in 28 chapters. Thirty words were selected, which were depicted in the video and were suitable for teaching to target students with a learning disability. The students watched the videos, as well as some other activities to reinforce the acquisition of the target words. These activities included a reading comprehension task with six narrative texts of 150 words length, each containing five target words and ten comprehension questions. The study results showed that children in the video-assisted group statistically outperformed the non-video group on word meaning acquisition. However, this work did not reveal any significant difference in the word generalization and reading comprehension.

## **3.4. Reading Comprehension**

Reading Comprehension is defined as “the construction of meaning of a written or spoken communication through a reciprocal, holistic interchange of ideas between the interpreter and the message in a particular communicative context. Note: The presumption here is that meaning resides in the intentional problem-solving, thinking processes of the interpreter during such an interchange, that the content of the meaning is influenced by that person’s prior knowledge and

experience, and that the message so constructed by the receiver may or may not be congruent with the message sent.” [89]. The ultimate goal of acquiring different literacy skills from phonemic awareness to vocabulary and fluency is comprehending texts efficiently and producing understandable texts for others.

Since various skills are involved in proficient reading comprehension, a deficit to one or multiple of these skills can impair comprehension. A source of comprehension difficulty can come from a deficit to one or multiple of these skills: lexical processes, working memory, cognitive inhibition, attention allocation, inference making, comprehension monitoring, and background knowledge. Lexical processes include phonological skills, semantic skills, and visual word recognition. [90], [91]. Therefore, each individual with reading comprehension difficulty can have a different underlying problem, resulting in a different reading profile [92].

Different reading comprehension methods have been proposed in the literature. However, the most common interventions are those that try to teach readers some strategies to improve comprehension skills, such as comprehension monitoring, inference making, cooperative learning, question generating and answering, identifying the main idea, summarizing, predicting, and recognizing the structure [6], [93], [94]. However, these strategies may differ for the comprehension of narrative texts and comprehension of expository texts. For narrative texts, strategies such as using story maps, retelling the story, making predictions, and answering comprehension questions can be taught. On the other hand, for expository texts, strategies like recognizing the structure, summarizing, main idea identification, and graphic organizers can be used [93].

In the following section, we present a summary of the instructional mechanisms of the comprehension intervention programs, as well as their results and effects.

### **3.4.1. Comprehension Booster**

Horne [95] used a computerized intervention program called *Comprehension Booster* [96]. It is created to improve the reading and listening comprehension of 7 to 14 years old children. It consists of 70 fiction and 70 non-fiction passages. It has seven different difficulty levels. Multiple-choice comprehension questions will follow each passage. Depending on the student's answers, the program can decide to change the difficulty level or continue with the same difficulty levels. Images accompany the texts, and the reader can select each word to hear the pronunciation or see its definition. It also includes asking questions and answering alongside immediate corrective feedback. The results of the intervention showed significant improvement in reading accuracy and comprehension for the intervention group.

### **3.4.2. CASTLE**

Sung et al. [97] presented and investigated a computer-assisted tool for teaching reading comprehension strategies to sixth-grade children. It is named *CASTLE*, which is the abbreviation of Computer Assisted Strategy Teaching and Learning Environment. They extended the Selection-Organization-Integration (SOI) model of text comprehension [98] by proposing the model of Attention-Selection-Organization-Integration-Monitoring (ASOIM). This model made the basis of their computer-assisted strategy-teaching tool. For the Attention component, they used self-questioning and error detection strategies to improve the readers'

concentration. For the Selection component, concept map blank-filling and highlighting strategies are used to teach students to select important messages. For the Organization component, concept map correction and inference blank-filling strategies have been used to teach students how to organize the article's messages. For the Integration component, proposition-combining and summarization strategies have been incorporated to teach students how to integrate the knowledge gained from reading the texts. Finally, the monitoring strategy aimed at teaching readers to monitor their performances while performing previous strategies. An agent using voice instruction was used to guide the users through the interface for facilitating user interaction. Finally, the intervention results showed that the experimental group outperformed the control group in reading comprehension and applying comprehension strategies.

### **3.4.3. Omega-IS**

*Omega-IS* (Omega-Interactive Sentence) [99] is a reading comprehension training program tested by Gustafson et al. [42]. Its focus is on the word and sentence level of reading. It begins from two-word sentences (noun + verb) and three-word sentences (noun + verb + noun), and it reaches to the level of constructing stories by choosing between different characters and scenarios for increasing the engagement and motivation of the child. The participant clicks on text buttons containing words or phrases and constructs the sentence. Then, a pre-recorder human voice reads the sentence, and an animation illustrates it. The program contains more than 1900 possible sentences with speech and animation. The posttest results from the intervention revealed large effects on passage comprehension and sight word reading, large to moderate effects on word decoding and pseudoword reading, and moderate to large effects on reading comprehension. Post-test and follow-up results revealed that intervention was more effective and persistent when combined with phonological awareness instruction [44].

### **3.4.4. Accurate Reading in Context**

Wise et al. [62] evaluated a computer-assisted reading comprehension program, called *Accurate Reading in Context*. Stories were available in ten directories, which corresponded to ten difficulty levels, and children chose stories from their appropriate grade level. During the reading, if children could not read a word, they click on the target word, and at first, it highlighted the word, and if it was a regular word, it was broken into segments to help the child pronouncing it. If the child clicked on it again, the computer pronounced the word. After the reading, children had to answer comprehension questions, and if they missed a question, the program brought them back to the corresponding section in the story. In addition, through teachers, children were presented with different reading comprehension strategies, such as making predictions, generating questions, and summarizing. Then, they were encouraged to incorporate these strategies into their computerized reading sessions. Finally, they compared the results of this intervention with a phonics program (Phonological Analysis). They concluded that although it improved children's literacy skills, it was less effective than the other approach in improving phonological skills. Furthermore, there were not any significant differences between the two programs on word reading measures.

### **3.4.5. LoCoTex**

Potocki et al. [100] examined a computer-assisted comprehension program named *LoCoTex*. It consists of three modules with 36 narrative texts of varying lengths and each having up to three non-fictional characters. The first module aims at strengthening the literal comprehension of children. The child reads the text and answers the multi-choice comprehension questions. If the child struggles to answer correctly, the passage containing the response will be highlighted, and the child has the chance to reread it and answer the question again. The second module aims at promoting the coherence or text-connecting inferencing skills. It uses anaphoric resolution exercises, in which the child has to match the anaphoric substitute (e.g., “it”, “the little girl”) with its right referent (e.g., “the ball”, “Anna”). The third module aims at fostering the children’s knowledge-based or gap-filling inferencing skills. In this module, after reading the text, they have to answer gap-filling questions, and if the answer was correct, they have to click on the words that lead them to the answer. Otherwise, if the answer was incorrect, the clue words will be highlighted, and then the question will be asked again. The result of this study showed a lasting effect on listening and reading comprehension. However, the effects on vocabulary and comprehension monitoring were less distinct.

### **3.4.6. e-PELS**

Ponce et al. [101] investigated the effectiveness of a computer-based system called *e-PELS* (“Programa de Entrenamiento en Lectura Significativa” or “Program in Deep Reading Comprehension”). It teaches multiple reading comprehension strategies, including underlining, paraphrasing, self-questioning, text structure, summarizing, using interactive graphic organizers, and conceptualizing strategies. Children start with reading a short text, and then, with the help of the teacher, they apply different strategies sequentially and systematically. The system contains 30 texts, but it is also possible for the teachers to add their texts. The intervention results revealed that the intervention group improved their reading comprehension skills significantly more than the control group, and the intervention was more effective for low-achieving children.

## **3.5. Fluency**

Reading fluency is the ultimate level that reading instructions aim to reach, and it has a bidirectional relationship with reading comprehension [102]. Despite the apparent general familiarity with fluency, there have been several different definitions of it in the literature, which indicates that fluency is a complex and multifaceted construct [103]. A more recent definition of fluency is “Fluency combines accuracy, automaticity, and oral reading prosody, which, taken together, facilitates the reader’s construction of meaning. It is demonstrated during oral reading through the ease of word recognition, appropriate pacing, phrasing, and intonation. It is a factor in both oral and silent reading that can limit or support comprehension.”[104].

Reading fluency integrates every process, skill, and sub-skill in reading [105], but broadly, it consists of three components of accuracy, automaticity, and prosody [106]–[109]. Automaticity is defined as “fluent processing of information that requires little effort or

attention” [89]. On the other hand, prosody refers to “stress, rhythm, intonation, and pause structure in speech and serves a wide range of linguistic and affective functions”[110]

There have been various intervention approaches for reading fluency. However, according to the National Reading Panel [6], most of them fall into these two categories of repeated oral reading practice and those approaches that try to increase independent or recreational reading. The first category includes repeated reading [111], neurological impress [112], radio reading [113], paired reading [114], etc. The second category includes approaches such as sustained silent reading [115] and accelerated reader [116]. Finally, in their review of fluency interventions, the National Reading Panel has concluded that repeated oral reading approaches have proven to be more effective than other approaches. More recent review studies have come to the same conclusion by admitting the effectiveness of repeated reading approach [117]–[119].

The following section presents a summary, the results, and the instructional mechanisms of the computerized fluency interventions.

### **3.5.1. Accelerated Reader**

Nunnery et al. [120] evaluated *Accelerated Reader* (AR) [116], widely used in thousands of United States’ schools throughout the years. A computer-assisted program helps teachers and students to monitor, motivate, and personalize the practice of reading. It provides comprehension quizzes at the end of each reading material. Through this feedback, it determines the reading level (zone of proximal development) of the student. Then, other reading materials matching this level will be available to be selected for future reading.

The teacher has access to information such as the results of quizzes, reading level, and amount of words read by the student. Thus, at any time, if a student is struggling, the teacher can decide to intervene by monitoring the student more closely or providing more personalized instruction. The reading practices can be in three forms. The materials can be read to the students (reading aloud); it can be read with the students (paired reading), or it can be read independently by the students. Compared to a control group, Nunnery et al. found that students using accelerated reader showed significantly higher improvement rate in reading achievement. In addition, they concluded that the effect was higher among lower grades. Furthermore, it reduced the negative effect of learning disability among these students. Shannon et al. [121] also evaluated Accelerated Reader across some elementary grades and concluded that students engaged in this computerized approach showed a statistically significant positive impact on their reading achievement. Additionally, another intervention study on AR revealed that high AR users had significantly greater improvement in their reading comprehension than low or average AR users [122]. Finally, the report from National Literacy Trust [123] indicated that the students who used accelerated reader enjoyed reading more, they read more frequently, and they thought more positive about reading, when compared to the students who did not use accelerated reader.

### **3.5.2. Reading Acceleration Program**

López-Escribano [124] tested the *Reading Acceleration Program* (RAP) [125] with Spanish dyslexic children. It is aimed at improving reading fluency by increasing the reading rate. First,

the user reads a certain amount of text at self-pace and answer the corresponding comprehension questions. Then, it calculates the user's current reading rate, and next time, the text will begin disappearing letter by letter from the start of the sentence at a pace relevant to the reading rate of the user. Multiple choice comprehension questions will follow each reading task. If the user answers them 100% correctly, the program will increase the pace of text disappearance with a small increment. If the user answers comprehension questions wrongly (less than 80%), it will decrease text disappearance speed. Finally, if the user answers the questions between 80% and 100%, it will not change the pace of the text disappearance. In this study, López-Escribano concluded that this intervention could improve the reading rate of proficient comprehenders while maintaining their comprehension level. In addition, It can increase the comprehension level of poor comprehenders, albeit without improving their reading rate.

This approach was tested in a study involving English children with a reading disability, and it was successfully improved the fluency sub-processes of children. However, it was not more effective than the control group's traditional repeated reading method [126]. Another study using RAP concluded that intervention resulted in an improvement in the comprehension level of reading disabled Dutch children during fast-paced reading [127]. Furthermore, another study on using this text-fading approach with German children showed that children in the treatment group significantly improved their sentence reading fluency in a standardized reading test [128]. In another study, RAP was accompanied by the instruction of a set of self-regulation of attention and engagement strategies. The results showed that the group using RAP significantly improved their silent sentence reading rate and lexical comprehension [129].

### **3.5.3. Computerized Flashcard Training**

Steenbeek-Planting et al. [130] have tested a computerized flashcard training approach [131], which displays single words with phonological CVC structure, one at a time with limited exposure time, and students are instructed to sound out the words as fast as they can. The experimenter records the correctness of verbal responses and, the response times are recorded in the computer so that the system can automatically adapt the exposure time of the words. To keep the accuracy level approximately constant, it decreases the presentation time if the student responds correctly to a certain amount of words, and it increases the presentation time if the student struggles to read aloud the words correctly. At each session, the student works on 100 words that are randomly taken from the training set. In one treatment group, every misread word was eliminated from the training set, so that the next time students only work on their past successes in addition to new words. However, in the other group, they did the opposite, and it was the correctly pronounced words, which were eliminated from further training so that the students can focus on their past failures, and the new words. The intervention results showed that it was effective for improving the word reading fluency and the effect transferred to untrained words as well. There was not a significant difference between the groups focusing on their successes or their failures. However, the children with lower initial reading level benefited more from focusing on successes. Conversely, the children who had higher initial reading level benefited more from focusing on their failures. However, another study using computerized flashcard training showed that children with low initial reading levels improved

more when focused on their failures, and children with high initial reading levels improved more when focused on their successes [132].

#### **3.5.4. ReadN'Karaoke**

Patel and McNab [110] assessed an oral reading software called *ReadN'Karaoke*, which aimed at promoting fluency through a guided repeated reading approach but with particular attention to prosody. To increase reading expressivity, it manipulates text into different formats to represent each of the fundamental frequency (pitch), Intensity (loudness), duration (length), and a combination of these elements. The spacing between characters and words are used to represent the duration of a word and pause between words, respectively. Three shades of font color (black, grey, and light grey) is used to show the intensity variations of the text. Finally, the text is fitted to the fundamental frequency contours to represent pitch variations of the text. These manipulated text formats were extracted from the recordings of a fluent adult reader. During the training session, the student should read the text in the standard format, manipulated formats, and their combination. The students have to read texts after listening to the readings made by the samples of a fluent adult reader. Their voice will be recorded; they can listen to their reading; also, they obtain the necessary feedback from the experimenter. Finally, they can reread and rerecord all over again. The results of this brief one session training showed that manipulated text, presenting fundamental frequency variations, has produced the most expressivity for participants. However, no significant difference in prosodic variations of standard reading between baseline and post-training was found. The next version of this software replaced manipulated text formats with augmented text by overlaid cues of pitch, duration, and intensity [133]. The results of a longer three-session training with this new version showed that participants transferred pitch and word duration variations to the standard reading of post-training sessions.

### **3.6. Multi-Component**

It is recommended for the interventions to contain explicit instruction in all key reading categories, including phonological awareness, phonics, vocabulary, comprehension, and fluency [6]. There are multiple reasons for this. First, reading is a multifaceted skill that involves many different processes, and all of them are important for becoming a skilled reader. Any impairment to each of these processes can lead to difficulty in reading. Second, every individual has a different reading profile, and not all struggling readers suffer from the same issue [92]. Therefore, a powerful intervention will cover all the aspects of reading to ensure that no aspect is neglected. A more powerful reading intervention can assess each individual's weaknesses and strengths to provide personalized and tailored instruction.

In the following sections, the multi-component reading intervention programs found in the literature are summarized and their effects are briefly mentioned.

#### **3.6.1. READ 180**

Kim et al. [134] utilized a computer-assisted program called *READ 180*, which includes various reading activities. It contains four different parts: the Reading Zone, Word Zone, Spelling Zone, and Success Zone. The Reading Zone is a section, which provides basic phonics



instruction; The Word Zone tries to help children achieve fluent Word Reading; The Spelling Zone tries to improve spelling abilities for target words, and the Success Zone is a section in which children should answer comprehension questions, and it records their oral reading. The intervention results found no significant impact on word reading efficiency measures, reading comprehension, and vocabulary. However, in another study, Kim et al. [135] evaluated the READ 180 Enterprise. The results showed that the treatment group outperformed the control group on vocabulary and reading comprehension measures, but not on spelling and oral reading fluency.

### **3.6.2. Alphie's Alley**

Chambers et al. [136] tested a computer-assisted tutoring system named *Alphie's Alley* [137] as an intervention for improving the literacy skills of struggling readers. Its basic function is to assess the individual's literacy performance and create individualized plans tailored to their needs. It consists of 12 types of activities from various aspects of literacy. Here is the list of these activities: 1- Letter identification 2- Letter writing 3- Auditory blending 4- Auditory segmenting 5- Sight words 6- Word-level blending 7- Spelling 8- Story preparation 9- Tracking 10- Fluency 11- Comprehension Questions 12- Graphic Organizers.

Besides, it utilizes embedded multimedia such as animations, pictures, and videos. Furthermore, through video vignette or written suggestions, it provides performance support to the tutors for addressing students' specific problems. The study results indicated that first graders in the experimental group showed significantly more improvement in reading achievement compared to the control group who practiced one-to-one tutoring. However, the second graders in the experimental group did not show any significant difference to the control group. In another study, Chambers et al. [138] evaluated *Alphie's Alley's* effectiveness. The results revealed that students whose tutors fully implemented this computer-assisted tutoring program showed significant improvement in letter-word identification, word attack, and fluency. However, no significant difference was found for comprehension. Another study [139] showed that tutoring with *Alphie* had a substantial positive effect on reading measures, and the effect size was larger for second and third-grade participants than first graders.

### **3.6.3. Read, Write, and Type**

Torgesen et al. [33] evaluated computer-assisted instructions to prevent future failures in children at risk of dyslexia, using a program called *Read, Write, and Type* created by Dr. Jeanine Herron [139]. It is a program to teach basic literacy knowledge by practicing phonetic spelling and writing skills by using colorful animation, digitized speech, and an engaging storyline. It explicitly teaches phonological awareness, letter-sound association, and phonemic decoding. It also aims at improving the keyboard typing skill of young children. The post-test results and follow-up showed significant improvements in phonemic awareness, phonemic decoding, and rapid naming in the intervention group compared to the control group.

### **3.6.4. Letter Prince**

Van de Ven et al. [140] tested a mobile game for improving early literacy, called *Letter Prince* (Letterprins) [141] on measures of pseudoword reading, word decoding, fluency, and

reading motivation. Letter Prince is a reading game in which children have to help a character (Letter Prince) slay a dragon and enter a castle. For this goal, by doing some reading exercises, the player helps the character to collect necessary items such as a wooden sword or shield. There are four types of exercises in the game. The first is a grapheme-phoneme conversion, in which a letter is displayed, and the child has to say it out loud, and test assistant will decide if the child said it correctly or not. The second exercise is a semantic categorization task, in which the children have to decide whether a word belongs to a certain category or not. The third exercise is a sentence verification task, which presents a short sentence, followed by a short question, and the child has to decide whether the sentence is semantically plausible or not. Eventually, the fourth exercise is a flashcard training that presents a letter or a word for a short amount of time and then after it disappears, the child has to indicate to the test assistant which word or letter was presented. The game incorporates several reward types including stars after completing each level, showing an encouraging prerecorded video, and virtual stickers. In addition, the difficulty of the game is adapted to children's ability level, to be neither too easy nor too difficult. The results revealed a positive effect on pseudoword reading and text reading fluency, but no effect was found for word decoding or reading motivation.

### **3.6.5. MindPlay Virtual Reading Coach**

Schneider [142] evaluated an online reading program called *MindPlay Virtual Reading Coach* (MVRC) [143]. At the beginning, through MindPlay Universal Screener, it assesses the reading skills of children. Then it creates an individualized syllabus for each student, containing direct, systematic, and explicit instructions on phonemic awareness, phonics, vocabulary, comprehension, fluency, and grammar, aligned with common core standards [3]. In addition to interactive lessons, it also provides pre-recorded videos from reading specialists and speech pathologists. Large effect sizes have been observed for the overall intervention, reading fluency, and non-word spelling, whereas a moderate effect size was found for real-word spelling. However, this study did not find any significant effect size for non-word and real-word reading.

### **3.6.6. Waterford Early Reading Program**

Cassady and Smith [144] investigated the effect of the *Waterford Early Reading Program* (WERP) on first graders' reading skills. It is an adaptive computer program that integrates the class-based assessments, instructional activities, and instructional materials for a systematic approach. It covers all the key reading components of phonological awareness, phonics, vocabulary, comprehension, and fluency. It has three levels, with each level containing instructional activities for a full year, as well as take-home activities. The intervention results showed that children involved in the intervention group outperformed those involved in the comparison group. It was also noticed that the intervention was more beneficial for children with low initial reading level.

## **3.7. Additional Research**

Although no conference publication met the inclusion criteria, it is important to mention that they were a valuable source of information in designing technology-based reading

interventions. Most of the conference papers are mainly focused on the developed technological approaches. They describe the innovative methods to remediate reading difficulties, but in general, they did not include experimentation to validate their proposed approaches. However, for completeness, selected conference papers are included in this analytical study.

Tzouveli et al. [145] presented an adaptive assistive reading tool called *AGENT-DYSL*. It used speech and image recognition to detect both the error types and the individual's affective state. Additionally, based on the user's created profile and the re-evaluation sessions, it provides assistive functions such as changing font attributes, highlighting, and pre-emptive reading.

Lin et al. [146] explored using a Mobile Augmented Reality (MAR) application called *Aurasma*, which interposes virtual objects on the real-world environment via a camera. It scans the words on flashcards and shows the related educational materials such as videos, animation, or data.

Daud and Abas [147] described a mobile app called *Dyslexia Baca* developed based on the ADDIE model, having five phases: Analysis, Design, Development, Implementation, and Evaluation. It is intended to teach Malay letter recognition to children with dyslexia by emphasizing the identification of confusing letters, such as letters /p/ and /b/.

Rello et al. [148] presented a game application called *Dyseggia* that used the analysis of dyslexics' errors to design their exercises. It consisted of five phonics activities for Spanish: insertion, omission, substitution, derivation, and separation. Also, three different levels of difficulty were created such that when the level of difficulty is increased, less frequent, and longer words with more complex morphology were used [149].

Bittencourt et al. [150] described the workflow of developing a mobile application targeting 6-9 years old dyslexics. The design process was started by observing the speech therapists' sessions with dyslexic children. Based on the observed activities, two digital games called *Corrida* and *Memória* were developed. These two games address syllable awareness and working memory, respectively. They have also outlined mobile accessibility principles for dyslexia in terms of text, layout, and navigation.

**Table 2.** Study Characteristics

| Publication   | Intervention Type | Program Name                                       | Technology Type | Design Type | Treatment Group  | Control Group    | Target Age | Treatment Group Size | Control Group Size | Duration of Intervention | Frequency (Weekly) | Intervention (hour) | Sessions (min) | Follow Up | Language | Country        |
|---------------|-------------------|--|-----------------|-------------|------------------|------------------|------------|----------------------|--------------------|--------------------------|--------------------|---------------------|----------------|-----------|----------|----------------|
| Horne2017     | Comp              | Comprehension Booster                              | Computer        | TC          | Poor Readers     | Poor Readers     | 6 to 11    | 19                   | 19                 | 6 weeks                  | 1 to 2             | 3 to 6              | 30             | No        | English  | United Kingdom |
| Macarus2006   | Phonics           | Lexia  | Computer        | TC          | Normal & At Risk | Normal & At Risk | Grade 1    | 83                   | 84                 | 6 months                 | 2 to 4             | 27                  | 20 to 30       | No        | English  | United States  |
| Sainio2011    | Phonics           | GraphoGame   | Computer        | MT          | At Risk          | At Risk          | Grade 1    | 25                   | 25, 116            | 28 weeks                 | 4                  | 50+                 | 45             | Yes       | Finnish  | Finland        |
| Wise2000      | Comp, Phonics     | Accurate Reading in Context, Phonological Analysis | Computer        | MT          | Poor Readers     | -                | Grade 2-5  | 91, 109              | -                  | 6 months                 | -                  | 27 to 29            | 30             | Yes       | English  | United States  |
| Kim2011       | Multi             | READ 180 Enterprise                                | Computer        | TC          | Poor Readers     | Poor Readers     | Grade 4-6  | 155                  | 157                | 23 weeks                 | 4                  | 50+                 | 60             | No        | English  | United States  |
| Chambers2011  | Multi             | Alpha's Alley                                      | computer        | TC          | Poor Readers     | Poor Readers     | Grade 1-2  | 372                  | 274                | -                        | 4 to 5             | -                   | 45             | No        | English  | United States  |
| Kim2010       | Multi             | READ 180   | Computer        | TC          | Poor Readers     | Poor Readers     | Grade 4-6  | 133                  | 131                | 23 weeks                 | 4                  | 50+                 | 60             | No        | English  | United States  |
| Savage2009    | Phonics           | ABRACADABRA  | Computer        | MT          | Normal           | Normal           | Grade 1    | 43, 44               | 57                 | 12 weeks                 | 4                  | 13                  | 20             | Yes       | English  | Canada         |
| Sung2008      | Comp              | CASTLE   | Computer        | TC          | Normal           | Normal           | Grade 6    | 65                   | 65                 | 11 weeks                 | 2                  | 18                  | 50             | No        | Chinese  | Taiwan         |
| Torgesen2010  | Multi-Comp, PA    | Read, Write, and Type (RWT), LIPS                  | Computer        | MT          | At Risk          | At Risk          | Grade 1    | 36, 36               | 40                 | School Year              | 4                  | 50+                 | 50             | Yes       | English  | United States  |
| Nicolson2000  | Phonics           | RTA  | Computer        | MT          | Poor Readers     | Poor Readers     | 6 & 8      | 74                   | 95, 103            | 10 weeks                 | 2                  | 10                  | 30             | No        | English  | United Kingdom |
| Cohen2015     | PA                | Fast Forward                                       | Computer        | MT          | SLI              | SLI              | 6 to 10    | 23                   | 27, 27             | 6 weeks                  | 5                  | 45                  | 90             | Yes       | English  | United Kingdom |
| Pokorny2004   | PA                | Fast Forward, Eurobes, LIPS                        | Computer        | MT          | Poor Readers     | -                | 7.5 to 9   | 20, 16, 18           | -                  | 20 days                  | -                  | 50+                 | 180            | No        | English  | United States  |
| Megan2004     | PA                | PLAY-ON  | Computer        | MT          | RD               | RD               | 8 to 12    | 7                    | 7                  | 5 weeks                  | 4                  | 10                  | 30             | Yes       | French   | France         |
| Gustafson2011 | Comp, PA, Multi   | Omegais, COMPHOT, Combined                         | Computer        | MT          | RD               | RD & Normal      | Grade 2    | 25, 25, 25           | 25, 30             | 5 to 9 weeks             | -                  | 7                   | 15 to 25       | Yes       | Swedish  | Sweden         |
| Kyle2013      | Phonics           | GraphoGame   | Computer        | MT          | Poor Readers     | Poor Readers     | Grade 2    | 11, 10               | 10                 | 12 weeks                 | 5                  | 11                  | 10 to 15       | Yes       | English  | United Kingdom |
| Shannon2015   | Fluency           | Accelerated Reader                                 | Computer        | TC          | Normal           | Normal           | Grade 1-4  | 189                  | 155                | 24 weeks                 | 5                  | 50+                 | 30 to 45       | No        | English  | United States  |
| Nunery2006    | Fluency           | Accelerated Reader                                 | Computer        | TC          | At Risk          | At Risk          | Grade 3-6  | 537                  | 441                | School Year              | 5                  | -                   | 30 to 60       | No        | English  | United States  |
| Chambers2008  | Multi             | Alpha's Alley                                      | Computer        | TC          | At Risk          | At Risk          | Grade 1    | 224                  | 188                | -                        | 5                  | -                   | 20             | No        | English  | United States  |
| VandeVen2017  | Multi             | Letter Prince                                      | Smartphone      | MT          | LD               | LD               | Mean 8.8   | -                    | -                  | -                        | -                  | 2                   | 15             | Yes       | Dutch    | Netherlands    |
| Schneider2016 | Multi             | MindPlay Virtual Reading Coach                     | Computer        | TC          | At Risk          | At Risk          | Grade 2    | 89                   | 81                 | School Year              | 4                  | 44                  | 30             | No        | English  | United States  |
| Messer2017    | Phonics           | Traintext  | Computer        | TC          | Poor Readers     | Poor Readers     | Mean 7.6   | 45                   | 33                 | 16 months                | 2 to 3             | 33                  | 10 to 15       | No        | English  | United Kingdom |
| Rosa2017      | Phonics           | GraphoGame   | computer        | TC          | At Risk          | At Risk          | Grade 1    | 44                   | 43                 | 3 months                 | 5                  | 6                   | 30             | No        | Spanish  | Chile          |

Continued on next page

Table 2. Continued

| Publication             | Intervention Type | Program Name                      | Technology Type | Design Type | Treatment Group                 | Control Group                   | Target Age | Treatment Group Size | Control Group Size | Duration of Intervention | Frequency (Weekly) | Intervention (hour) | Sessions (min) | Follow Up | Language | Country       |
|-------------------------|-------------------|-----------------------------------|-----------------|-------------|---------------------------------|---------------------------------|------------|----------------------|--------------------|--------------------------|--------------------|---------------------|----------------|-----------|----------|---------------|
| Patel2011               | Fluency           | ReadN'Karaoke <sup>e</sup>        | Computer        | ST          | Normal                          | -                               | 6 to 9     | 10                   | -                  | -                        | -                  | 1                   | 60             | No        | English  | United States |
| Gustafson2007           | PA, Phonics       | COMPHOT, DOT                      | Computer        | MT          | RD                              | RD & Normal                     | Grade 2-3  | 41, 39               | 20, 34             | -                        | -                  | 10                  | 18             | Yes       | Swedish  | Sweden        |
| Johnson2003             | Fluency           | Accelerated Reader                | Computer        | ST          | At Risk                         | -                               | Grade 3-5  | 755                  | -                  | School Year              | -                  | -                   | -              | No        | English  | United States |
| Lopez-escribano2016     | Fluency           | Reading Acceleration Program      | Computer        | ST          | RD                              | -                               | Grade 4-5  | 12                   | -                  | 4 weeks                  | 5                  | 10                  | 30             | No        | Spanish  | Spain         |
| Stellingsma2015         | Fluency           | Reading Acceleration Program      | Computer        | TC          | RD & Normal                     | RD & Normal                     | Grade 4    | 15, 22               | 9, 13              | 3 weeks                  | 3                  | 3                   | 20             | No        | Dutch    | Netherlands   |
| Nagler2015              | Fluency           | Reading Acceleration Program      | Computer        | TC          | Normal                          | Normal                          | Grade 3    | 10                   | 12                 | 3 weeks                  | 3                  | 4.5                 | 30             | No        | German   | Germany       |
| Page2011                | Fluency           | Reading Acceleration Program      | Computer        | TC          | RD                              | RD                              | Grade 6    | 9                    | 11                 | 12 weeks                 | 6                  | 5                   | 5 to 6         | No        | English  | United States |
| Madden2017              | Multi             | Tutoring with Alphie (TVA)        | Computer        | MT          | Poor Readers                    | Poor Readers                    | Grade 1-3  | 127-242              | 309, 236           | School Year              | 5                  | 7.5+                | 30             | No        | English  | United States |
| Steenbeck-Planting2012  | Fluency           | Computerized Flashcard Training   | Computer        | MT          | Poor Readers                    | -                               | Mean 8.8   | 20, 18, 20, 21       | -                  | -                        | 1 to 2             | 3.3                 | 20             | No        | Dutch    | Netherlands   |
| Eccles2013              | Phonics           | Classsyno, Opposition Phonologues | Computer        | MT          | Poor Readers                    | Poor Readers                    | Grade 1-2  | 9, 9 - 9, 9          | 9                  | 5 weeks                  | 4                  | 10                  | 30             | Yes       | French   | France        |
| Potocki2013             | Comp              | LoCoTex                           | Computer        | TC          | Poor Comprehenders <sup>s</sup> | Poor Comprehenders <sup>s</sup> | Grade 2    | 15                   | 15                 | 5 weeks                  | 4                  | 10                  | 30             | Yes       | French   | France        |
| Patel2014               | Fluency           | ReadN'Karaoke <sup>e</sup>        | Computer        | ST          | Normal                          | -                               | 7 to 8     | 8                    | -                  | 3 weeks                  | 1                  | 4                   | 60 to 90       | No        | English  | United States |
| Steenbeck-Planting2013a | Fluency           | Computerized Flashcard Training   | Computer        | MT          | Poor Readers                    | Poor Readers                    | Grade 2-6  | 20, 20               | 20                 | -                        | 1 to 2             | 3.6                 | 20             | No        | Dutch    | Netherlands   |
| Nieder2014              | Fluency           | Reading Acceleration Program      | Computer        | TC          | Slow Readers                    | Slow Readers                    | Grade 4    | 7                    | 7                  | -                        | -                  | 9                   | 60             | No        | English  | United States |
| Xin2001                 | Vocab             | The Great Quake of '89, Waterford | Videodisc       | TC          | LD                              | LD                              | Grade 4-6  | 40                   | 36                 | 6 weeks                  | 3                  | 9                   | 30             | Yes       | English  | United States |
| Cassady 2005            | Multi             | Early Reading Program (WERP)      | Computer        | TC          | Normal                          | Normal                          | Grade 1    | 46                   | 47                 | School Year              | 5                  | 30                  | 20             | No        | English  | United States |
| Ponce2012               | Comp              | e-PELS                            | Computer        | TC          | Normal                          | Normal                          | Grade 4    | 939                  | 102                | School Semester          | -                  | 45                  | 90             | No        | Spanish  | Chile         |
| Moses2017               | Phonics           | 8 Great Word Patterns             | Tablet          | TC          | Normal                          | Normal                          | Grade 4    | 14                   | 15                 | 10 weeks                 | 5                  | 10                  | 10 to 15       | No        | English  | United States |
| Savage2010              | Phonics           | ABRACADA BKA                      | Computer        | TC          | Normal                          | Normal                          | Grade 1    | 15, 16, 18           | 11                 | 8 weeks                  | -                  | 16                  | -              | No        | English  | Canada        |

Note: PA= Phonological Awareness; Vocab=Vocabulary; Comp= Reading Comprehension; Multi= Multi-Component; CTC=Treatment vs Comparison; MT= Multi-Treatment groups; ST= Single Treatment group; RD= Reading Disabled; LD=Learning Disabled; STP= Speech and Language Impairment.

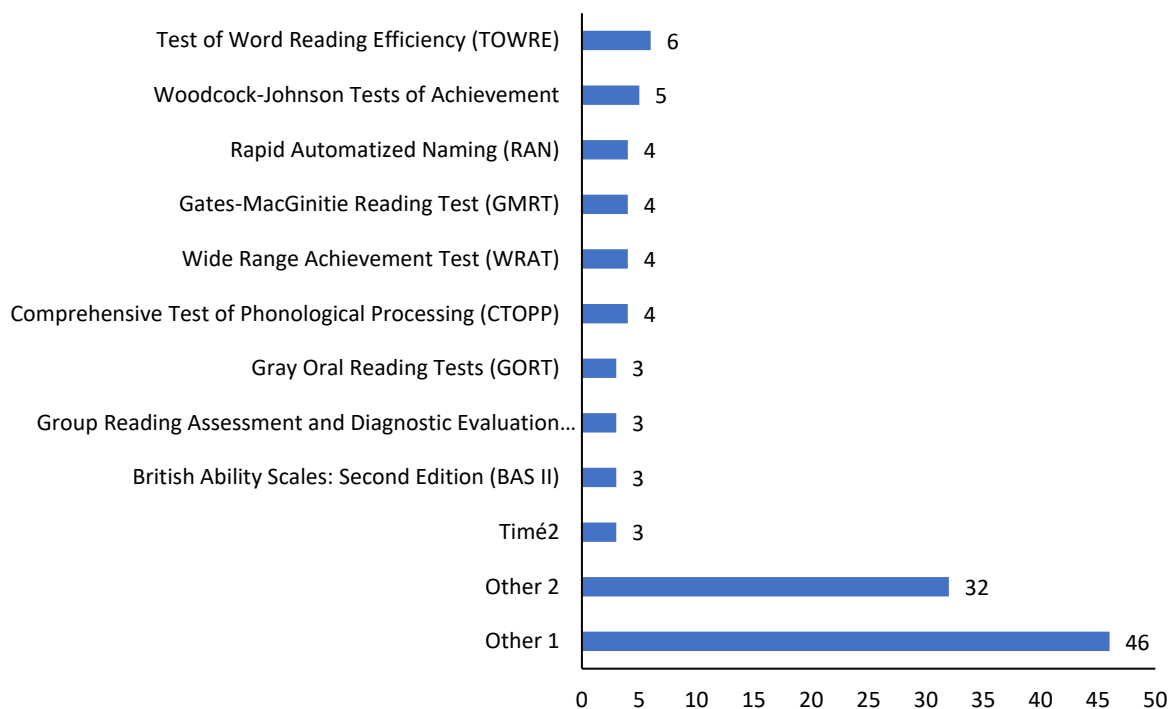
## 4. Analysis of the Reviewed Research

This chapter has identified and reviewed studies on technology-based and technology-assisted reading interventions for elementary grades. The review resulted in 42 articles and 32 programs. The content, the instructional mechanisms, and intervention outcome of each reading program were outlined. The important details and characteristics of these studies are presented in Table 2. Different aspects of reading intervention studies are analyzed in this section.

### 4.1. Measures

Since there were many differences in study designs and too many different variables were involved in these studies, it was considered that conducting a meta-analysis would not bring sufficiently reliable results. Apart from the author-devised measures, 72 different measures were found used 117 times over the studies. Figure 3 shows the most used measures. These frequently used measures are: Test of Word Reading Efficiency (TOWRE) [151], Woodcock-Johnson Tests of Achievement [152], [153], Rapid Automatized Naming [154], [155], Gates-MacGinitie Reading Test (GMRT) [156], [157], Wide Range Achievement Test (WRAT) [158], Comprehensive Test of Phonological Processing (CTOPP) [159], Gray Oral Reading Tests (GORT) [160], [161], Group Reading Assessment and Diagnostic Evaluation (GRADE) [162], British Ability Scales: Second Edition (BAS II) [163], and Timé2 [164]. Category Other 1 in this figure represents the number of measures used only once in this review and Other 2 represents the number of measures used twice.

Fifty percent of the studies used only the standardized measures, twelve percent only used author-devised measures, and thirty-eight percent used both. The analysis of measures based on the intervention types showed that multi-component studies had the highest rates (70%) in using only standardized measures. This result is understandable, as multi-component studies target several reading skills that can be evaluated by general reading tests, and author-devised measures were not required. Fluency and phonological awareness (58% and 50%) are the next intervention types using standardized measures more frequently. It might be due to the powerful existing measures for these two skills, or the generality of these two skills, which can be evaluated as a whole. However, a smaller number of phonics studies (33%) used only standardized measures. This might be due to the difference in phonics interventions since different skills were targeted, from phoneme-grapheme matching to using bigger chunks such as morphemes, syllables, and whole words. Comprehension (16%) used a lower amount of purely standardized tests. Most of the comprehension studies tended to teach certain comprehension strategies to the children. Some authors needed to devise tests to measure how good the learners have acquired these strategies. Only one vocabulary intervention is used in the review, but generally, vocabulary interventions use mostly author-devised tests [165]. It is due to the sheer amount of individual vocabulary knowledge compared to the limited amount of vocabulary taught in an intervention. It is very difficult to impact the general vocabulary knowledge measured by standardized tests. It is why researchers try to devise measures for limiting the tests to the taught words.



**Figure 3.** Frequency of the measures throughout the studies

## 4.2. Interventions

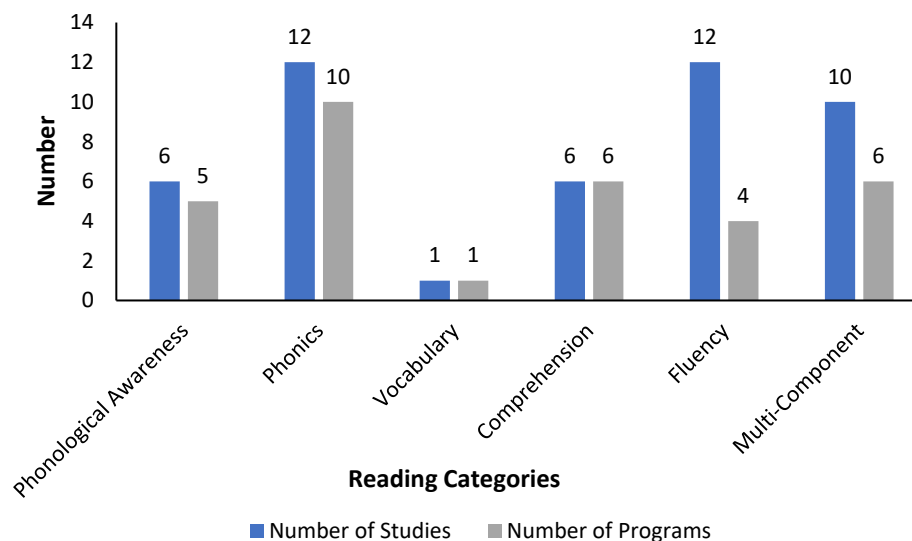
As illustrated in Figure 4, the studies that used phonics, fluency, or multi-component hold bigger shares. It was expected that few studies used phonological awareness intervention because this approach is most effective in pre-elementary education, and when children start elementary grades, it becomes less effective [32], [166]. Thus, a significant number of interventions focused on phonics by teaching phonological-orthographical matching.

Despite the high number of studies focusing on fluency, they have used only four different programs. This lack of variety may show the difficulty of designing a fluency-based intervention. When these four programs are further examined, it becomes more apparent that there is a lack of a well-rounded approach to fluency interventions. Flashcard training works only on word recognition automaticity. ReadN'Karaoke focuses solely on prosody. RAP concentrates on reading rate, and finally, AR is a reading practice monitoring system that encourages independent reading without providing any reading instruction. Therefore, none of the reviewed programs provides a holistic approach for improving fluency, considering its complex and multifaceted nature. The use of technology to improve the automaticity and speed of reading can be useful. However, in typical school conditions, there is not enough practice for struggling readers to reach mastery levels in decoding [167].

Despite the high variability of intervention programs in phonics, comprehension, and phonological awareness, they are similar. For example, phonics interventions mainly target phonological-orthographic matching, and comprehension interventions are specifically instructing certain comprehension strategies. However, as mentioned previously, fluency intervention programs are very different; this explains the existing uncertainty about the effectiveness of the current fluency intervention approaches and the increasing need for better

solutions. Innovative approaches can automatically adapt themselves to the learner's capacity; they modulate the difficulty of challenges in the gamified context for allowing the acquisition of automaticity. Gamified approaches are specifically useful for fostering fluency; they can motivate the learners, increase the adherence and engagement for long-time training leading to automaticity acquisition.

Surprisingly, only one article on vocabulary intervention that met the inclusion criteria of this study was found. Despite knowing the prominent role of vocabulary knowledge on reading failure and its clear link with reading comprehension [78], [79], [168], it is surprising to notice the limited number of technology-based vocabulary interventions for early readers. With the proven effect of multimedia on vocabulary learning [169], vocabulary is the easiest reading category to be taught through technology. There is a big number of studies on vocabulary intervention for second language learning. More studies should exploit technology to bridge the first language vocabulary gap among children [80]. Most of the vocabularies that we learn are indeed through incidental acquisition [84]. However, it is worthwhile looking for more effective methods of teaching vocabularies explicitly. The acquisition of a large number of vocabularies is not possible during school time [6]. Hence, extensive in-home interventions could be an appropriate solution to meet this objective. In addition, technologies such as computers could be exploited to enhance the rate of incidental acquisition of vocabularies.



**Figure 4.** Distribution of reviewed studies and programs based on intervention type

### 4.3. Duration of Interventions

From the analysis of studies that provided the total hour of their interventions, multi-component studies had the longest interventions; this is reasonable because they are focused on multiple reading skills. The phonological awareness comes in the second-order, with an average of exceeding 28 hours. In the third position, phonics and comprehension with more than 18 hours, and fluency had an average intervention duration of 12 hours. Unexpectedly, fluency interventions had the shortest durations while acquiring this skill requires a long



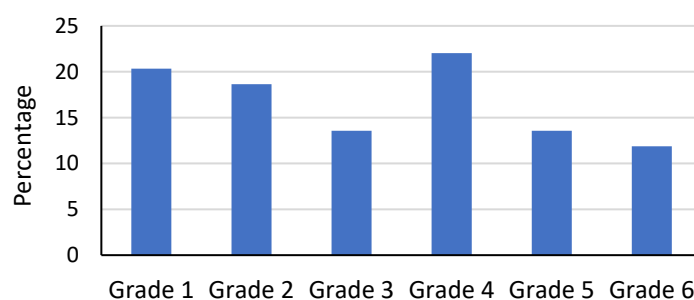
training time. Fluency requires automaticity and mastery, which needs extensive training, that is not provided in these interventions.

#### 4.4. Technologies

Of the 32 reading programs discussed in this chapter, 29 of them were computer-based. Smartphones and tablets were used only once. Considering the estimations that there are more than two billion smartphone users and more than one billion tablet users worldwide in 2017 [170], it is surprising that so few studies are dedicated to evaluating the impact of smartphone and tablet's literacy apps on reading skills. Given the abundant number of available literacy applications in smartphones and tablets, it would be beneficial to study their effectiveness and see which instructional elements impact more. Moreover, too few papers studied an in-home intervention. It might be because tracking the fidelity of the intervention is harder and less reliable. However, given that nowadays, most people are connected to the internet, all the use and performance data can be recorded and transferred easily and automatically. Software programs can be created adaptive enough to propose personalized learning sessions to the children, in the absence of their teacher, which can provide an opportunity for learners who do not have sufficient access to qualified tutors.

#### 4.5. Grade Levels

In the studied articles, as it is shown in Figure 5, the first, second, and fourth graders received more attention (each with near 20 percent). This result was predictable. For the first and second grade, there is a widespread belief that earlier interventions are more effective, and this is backed by many scientific studies [171], [172]. It is also known that fourth grade is the period that the gap between struggling and normal readers will become more distinct, as it is traditionally called the 'fourth-grade slump' [173]. It is argued that the reason behind this gap is that in fourth grade, children are no longer learning to read, and instead, they should read to learn other materials. The distribution of interventions over the grades shows that phonological awareness and phonics interventions were used mostly in the early grades. The comprehension interventions were more evenly distributed between the grades. On the other hand, fluency interventions were used mostly in upper elementary grades. However, if the intervention processes were gamified using intuitive interaction techniques that improve motivation and adherence, some fluency interventions could be performed in early grades.



**Figure 5.** The percentage share of each grade among the reviewed studies

#### **4.6. Game-Based and Gamified Interventions**

Game-based intervention is an approach where a game such as computer-based narrative games is used to enhance learning. However, gamification is “the use of game elements in non-game contexts” [174]. The gamification concept has been shown to increase attendance, motivation, and engagement in learning, which are invaluable in developing the reading skills [175], [176]. Nevertheless, less than one-fourth of the reviewed studies used gamified or game-based programs, which is less than expected, considering the target groups' young age. Most of the gamified or game-based approaches were used in phonological awareness and phonics interventions. The gamification of comprehension and fluency programs is not a common practice; this is due to the nature of reading, which demands high focus and full presence. However, to make fluency and comprehension training more appealing, innovative approaches can include game elements such as challenges, rewards, and visible cues of progress. It motivates the learner to persevere, which is crucial for reaching a high level of mastery.

Moreover, the impact of the extrinsic motivations (provided by game elements) on the intrinsic motivation of reading should be investigated. However, most studies only examined the effect of interventions on reading skills, and the role of motivation to read was overlooked. As it is known, extensive independent reading plays a prominent role in the development of reading skills [177]. Therefore, increasing the motivation to engage in reading activities can be crucial to future success in reading, especially for struggling readers who find reading a hard and laborious task.

It is important to note that the present review of technology-based interventions was focused on all types of applications using ICT systems, including gamified and game-based approaches. However, not all the reviewed works using gamification or games have provided a sufficiently detailed description of their reading programs. Hence, it was impossible to distinguish whether they used games or the gamification concept in these reviewed research works. For this reason, additional investigations are necessary to specify the appropriate design of gamified/game-based interventions and determine their effect on acquiring reading skills.

#### **4.7. Languages**

Sixty-four percent of the studies were conducted for English speaking children, and about 10 percent targeted Dutch-speaking children. The other languages, such as French and Spanish, represent less than 10 percent each. The studies were classified into two categories of orthographies. The first category includes studies with deeper orthographies (English, French and Chinese), and the second category included studies of shallower orthographies (Finnish, Spanish, German, Swedish, and Dutch). The deeper orthographies focused more on phonics interventions (28%). However, studies in shallow orthographies focused more on fluency interventions (36%). It has been suggested that in deep orthographies such as English, the most serious reading difficulty is the decoding accuracy, while in transparent languages such as Spanish, the biggest problem of poor readers is reading rate [124], [178]. However, it seems improbable that addressing decoding accuracy of English speaking children would solve all of the reading difficulties, and still, the slow reading rate should be addressed.

#### **4.8. Study Limitations**

The study of the published works had some limitations. First, due to the differences in methodologies and excess of variables involved in reviewed papers, it was decided to exclude the meta-analysis procedure. However, this study can be used as a precursor to a meta-analysis. Also, in their titles, many studies do not mention that their interventions are carried out using a certain technology, making it difficult to identify them. Therefore, a more thorough search for these types of studies should result in more papers. Furthermore, in future studies, it will be useful to present a more in-depth description of the programs' characteristics, such as interface design, adaptiveness, affordance, and the game elements. However, unfortunately, many studies did not provide sufficient information and enough details about their intervention programs.

### **5. Conclusion**

Due to the utmost importance of reading acquisition and the severe consequences of reading failure, many researchers have been motivated to develop innovative approaches to create more effective reading interventions. This chapter presented a comprehensive overview of the technology-based or technology-assisted reading intervention studies for elementary grades. The purpose was to provide exhaustive information about the current approaches applied to each reading component and suggest insights by analyzing the reviewed studies from diverse aspects. Since similar studies were published before 2000, the period from 2000 to 2017 was chosen for this analytical review. The reading programs were classified into phonological awareness, phonics, vocabulary, comprehension, fluency, and multi-component. A general introduction to each of these categories was presented. The reading programs were summarized by focusing on their content, instructional processes, key features, and the outcome of the interventions. Finally, the chapter presents the analysis of the identified pertinent aspects of the reviewed reading intervention studies.

Forty-two studies have been reviewed, resulting in 32 reading programs. Surprisingly, only one study investigated a vocabulary intervention; this indicates that using technology to enhance first language vocabulary acquisition is highly overlooked. As mentioned previously, it is challenging to influence general vocabulary knowledge; this suggests that extensive home-based intervention could be a promising solution. Additionally, the use of non-computer technologies, such as tablets and smartphones, is less than what was expected. Furthermore, the average intervention time for fluency was shorter than the duration of other intervention types. This does not satisfy the required extensive training time for achieving a high fluency level. Besides, current fluency approaches lack a holistic view that can target different fluency components at various levels. Moreover, compared to other types of interventions, fluency and comprehension programs benefit less from game-based and gamified approaches. Additionally, fluency programs were less used in early elementary grades.

This study's findings suggested that for languages with opaque orthographies such as English, the speed aspect of reading was neglected. Despite the emphasis on teaching phonics for these languages to address precision, dyslexia is still a problem of both accuracy and speed; these aspects should be addressed together in future studies. Hence, it is recommended to study

the efficacy of reading programs that accentuate both the decoding speed and decoding accuracy. Letter recognition automaticity could be trained at first, then word recognition automaticity could be practiced, and later, fluency approaches on the phrase, paragraph, and longer texts could be worked. Furthermore, instead of making a simple integral digital copy of the existing pedagogical methods used in schools, it is recommended that designers take benefit from the recent advances of information and communication technologies to design innovative methods not available in normal schooling conditions.

Intelligent self-adaptable systems should be developed to assess the student's skills. These systems should provide individualized interaction modalities and adapt the instruction's content based on the real-time capture of the user performance. Gamified multisensory interactive modalities such as spoken speech and tactile interfaces could be integrated to address different human sensory channels and enhance the learning process. Since it is already an established knowledge that emotion can drive motivation, which is crucial for developing reading skills, emotional interaction could also be included to enhance the reading programs' effectiveness.

Speech recognition was not used at its maximum potential for enhancing the reading acquisition. If advanced integration of speech recognition is made with being sensitive to smaller parts of the language so that it detects not only the pronunciation errors but also the prosody attributes and even the emotional states of the speaker, it would be promising for the creation of intelligent assistive reading systems. Moreover, the technologies of virtual and augmented realities can be used to teach vocabulary. In these technologies, the context relative to each word can be built in a meaningful and realistic way to enhance its retention.

## 6. References

- [1] L. Blomert and D. Froyen, "Multi-sensory learning and learning to read," *Int. J. Psychophysiol.*, vol. 77, no. 3, pp. 195–204, 2010.
- [2] S. Valencia, "A portfolio approach to classroom reading assessment: The whys, whats, and hows.," *Read. Teach.*, vol. 43, no. 4, pp. 338–340, 1990.
- [3] National Governors Association, "Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects," 2010.
- [4] J. R. Toste and S. Ciullo, "Reading and Writing Instruction in the Upper Elementary Grades," *Interv. Sch. Clin.*, vol. 52, no. 5, pp. 259–261, 2017.
- [5] National Assessment of Educational Progress, "The Nation's Report Card: 2015 Mathematics and Reading Assessments," Washington, DC, 2015.
- [6] National Reading Panel, "Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction," 2000.
- [7] S. Vaughn, S. Linan-thompson, K. Kouzekanani, D. P. Bryant, S. Dickson, and S. A. Blozis, "Reading Instruction Grouping for Students with Reading Difficulties," *Remedial Spec. Educ.*, vol. 24, no. 5, pp. 301–315, 2003.
- [8] H. Lyytinen, J. Erskine, M. Aro, and U. Richardson, "Reading and Reading Disorders," *Blackwell Handb. Lang. Dev.*, pp. 454–474, 2008.
- [9] K. Scior, "Public awareness, attitudes and beliefs regarding intellectual disability: A systematic review," *Res. Dev. Disabil.*, vol. 32, no. 6, pp. 2164–2182, 2011.
- [10] M. J. Kennedy and D. D. Deshler, "Literacy instruction, technology, and students with learning disabilities: Research we have, research we need," *Learn. Disabil. Q.*, vol. 33, no. 4, pp. 289–298, 2010.
- [11] D. B. Malouf, "The Effect of Instructional Computer Games on Continuing Student Motivation," *J. Spec. Educ.*, vol. 21, no. 4, pp. 27–38, 1988.

- [12] M. Papastergiou, "Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation," *Comput. Educ.*, vol. 52, no. 1, pp. 1–12, 2009.
- [13] R. E. Mayer and R. Moreno, "Nine Ways to Reduce Cognitive Load in Multimedia Learning," *Educ. Psychol.*, vol. 38, no. 1, pp. 43–52, 2010.
- [14] T. C. Williams and H. Zahed, "Computer-based training versus traditional lecture: Effect on learning and retention," *J. Bus. Psychol.*, vol. 11, no. 2, pp. 297–310, 1996.
- [15] K. E. Ricci, E. Salas, and J. A. Cannon-bowers, "Do Computer-Based Games Facilitate Knowledge Acquisition and Retention?," *Mil. Psychol.*, vol. 8, no. 4, pp. 295–307, 1996.
- [16] R. Andreev, V. Terzieva, and P. Kademova-Katzarova, "An approach to development of personalized e-learning environment for dyslexic pupils' acquisition of reading competence," *Proc. Int. Conf. Comput. Syst. Technol. Work. PhD Students Comput. - CompSysTech '09*, p. 1, 2009.
- [17] T. Athanaselis, S. Bakamidis, I. Dologlou, E. N. Argyriou, and A. Symvonis, "Making assistive reading tools user friendly: A new platform for Greek dyslexic students empowered by automatic speech recognition," *Multimed. Tools Appl.*, vol. 68, no. 3, pp. 681–699, 2014.
- [18] A. Corbett, "Cognitive Computer Tutors: Solving the Two-Sigma Problem," in *User Modeling 2001*, 2001, pp. 137–147.
- [19] C. A. MacArthur, R. P. Ferretti, C. M. Okolo, and A. R. Cavalier, "Technology applications for students with literacy problems: A critical review," *Elem. Sch. J.*, vol. 101, no. 3, pp. 273–301, 2001.
- [20] H. Blok, R. Oostdam, M. E. Otter, and M. Overmaat, "Computer-Assisted Instruction in Support of Beginning Reading Instruction: A Review," *Rev. Educ. Res.*, vol. 72, no. 1, pp. 101–130, Jan. 2002.
- [21] A. C. K. Cheung and R. E. Slavin, "The Effectiveness of Education Technology for Enhancing Reading Achievement: A Meta-Analysis," *Cent. Res. reform Educ.*, 2011.
- [22] A. Grant, E. Wood, A. Gottardo, M. A. Evans, L. Phillips, and R. Savage, "Assessing the Content and Quality of Commercially Available Reading Software Programs: Do They Have the Fundamental Structures to Promote the Development of Early Reading Skills in Children?," *NHSA Dialog*, vol. 15, no. 4, pp. 319–342, 2012.
- [23] L. Edwards Santoro and M. J. Bishop, "Selecting software with caution: An empirical evaluation of popular beginning reading software for children with early literacy difficulties," *Comput. Sch.*, vol. 27, no. December 2014, pp. 99–120, 2010.
- [24] L. Cidrim and F. Madeiro, "Information and Communication Technology (ICT) applied to dyslexia: literature review," *Rev. CEFAC*, vol. 19, no. 1, pp. 99–108, Feb. 2017.
- [25] D. J. Chard and S. Dickson, "Phonological awareness: Instructional and assessment guidelines," *Interv. Sch. Clin.*, vol. 34, pp. 261–270, 1999.
- [26] R. Treiman and A. Zukowski, "Levels of phonological awareness," in *Phonological processes in literacy: A tribute to Isabelle Y.*, 1991, pp. 67–83.
- [27] B. F. Pennington and D. L. Lefly, "Early Reading Development in Children at Family Risk for Dyslexia," *Child Dev.*, vol. 72, no. 3, pp. 816–833, 2012.
- [28] H. S. Scarborough, "Antecedents to reading disability: Preschool language development and literacy experiences of children from dyslexic families," *Read. Writ.*, vol. 3, no. 3–4, pp. 219–233, 1991.
- [29] F. Ramus, "Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction?," *Curr. Opin. Neurobiol.*, vol. 13, no. 2, pp. 212–218, 2003.
- [30] M. J. Snowling, "From Language to Reading and Dyslexia," *Dyslexia*, vol. 7, no. 1, pp. 37–46, 2001.
- [31] M. Melby-Lervåg, S.-A. Lyster, and C. Hulme, "Phonological skills and their role in learning to read: A meta-analytic review," *Psychol. Bull.*, vol. 138, no. 2, pp. 322–352, 2012.
- [32] L. C. Ehri, S. R. Nunes, D. M. Willows, B. V. Schuster, Z. Yaghoub-Zadeh, and T. Shanahan, "Phonemic Awareness Instruction Helps Children Learn to Read: Evidence From the National Reading Panel's Meta-Analysis," *Read. Res. Q.*, vol. 36, no. 3, pp. 250–287, Jul. 2001.
- [33] J. K. Torgesen, R. K. Wagner, C. A. Rashotte, J. Herron, and P. Lindamood, "Computer-assisted

- instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches,” *Ann. Dyslexia*, vol. 60, no. 1, pp. 40–56, 2010.
- [34] P. C. Lindamood, P. Lindamood, and S. Truch, “The Lindamood phoneme sequencing program for reading, spelling and speech: The LIPS Program.” Pro-Ed, 1998.
- [35] J. L. Pokorni, C. K. Worthington, and P. J. Jamison, “Phonological Awareness Intervention: Comparison of Fast ForWord, Earobics, and LiPS,” *J. Educ. Res.*, vol. 97, no. 3, pp. 147–158, 2004.
- [36] Scientific Learning Corporation, “Fast ForWord.” Scientific Learning Corporation, Oakland, CA, 1997.
- [37] W. Cohen, A. O. Hare, J. Boyle, and E. McCartney, “Intervention through acoustically modified speech (Fast ForWord ) in severe mixed receptive–expressive language impairment: Outcomes from a randomized controlled trial,” *J. Speech, Lang. Hear. Res.*, vol. 48, no. June 2005, pp. 715–729, 2015.
- [38] G. K. Strong, C. J. Torgerson, D. Torgerson, and C. Hulme, “A systematic meta-analytic review of evidence for the effectiveness of the ‘Fast ForWord’ language intervention program,” *J. Child Psychol. Psychiatry Allied Discip.*, vol. 52, no. 3, pp. 224–235, 2011.
- [39] I. Cognitive Concepts, “Earobics auditory development and phonics program step 2.” Evanston, IL: Author., 1998.
- [40] A. Magnan, J. Ecalle, E. Veuillet, and L. Collet, “The effects of an audio-visual training program in dyslexic children,” *Dyslexia*, vol. 10, no. 2, pp. 131–140, 2004.
- [41] L. Danon-Boileau and D. Barbier, “Play-On: Un logiciel d’entraînement à la lecture.” Paris: Audivi-Média, 2000.
- [42] S. Gustafson, L. Fälth, I. Svensson, T. Tjus, and M. Heimann, “Effects of three interventions on the reading skills of children with reading disabilities in grade 2.,” *J. Learn. Disabil.*, vol. 44, no. 2, pp. 123–35, 2011.
- [43] J. Ferreira, S. Gustafson, and J. Rönnerberg, “COMPHOT: Computerized phonological training.” Linköping: Institutionen för beteendevetenskap/Institutet för handikappvetenskap, Linköpings universitet, 2003.
- [44] L. Fälth, S. Gustafson, T. Tjus, M. Heimann, and I. Svensson, “Computer-assisted interventions targeting reading skills of children with reading disabilities - A longitudinal study,” *Dyslexia*, vol. 19, no. 1, pp. 37–53, 2013.
- [45] S. Gustafson, J. Ferreira, and J. Rönnerberg, “Phonological or orthographic training for children with phonological or orthographic decoding deficits,” *Dyslexia*, vol. 13, no. 3, pp. 211–229, Aug. 2007.
- [46] S. A. Stahl, “Saying the ‘p’ Word: Nine Guidelines for Exemplary Phonics Instruction,” *Read. Teach.*, vol. 45, no. 8, pp. 618–625, 1992.
- [47] R. L. Peterson and B. F. Pennington, “Developmental Dyslexia,” *Annu. Rev. Clin. Psychol.*, vol. 11, no. 1, pp. 283–307, Mar. 2015.
- [48] L. C. Ehri, S. R. Nunes, S. A. Stahl, and D. M. Willows, “Systematic Phonics Instruction Helps Students Learn to Read: Evidence from the National Reading Panel’s Meta-Analysis,” *Rev. Educ. Res.*, vol. 71, no. 3, pp. 393–447, 2001.
- [49] D. Wyse and U. Goswami, “Synthetic phonics and the teaching of reading,” *Br. Educ. Res. J.*, vol. 34, no. 6, pp. 691–710, 2008.
- [50] C. J. Torgerson, G. Brooks, and J. Hall, *A Systematic Review of the Research Literature on the Use of Phonics in the Teaching of Reading and Spelling*. 2006.
- [51] S. Vaughn, D. J. Chard, D. P. Bryant, M. Coleman, B. Tyler, and S. Linan-thompson, “Fluency and Comprehension Interventions for Third-Grade Students,” *Remedial Spec. Educ.*, vol. 21, no. 6, pp. 325–335, 1999.
- [52] H. A. E. Mesmer and P. L. Griffith, “Everybody’s Selling It-But Just What Is Explicit, Systematic Phonics Instruction?,” *Read. Teach.*, vol. 59, no. 4, pp. 366–376, Dec. 2005.
- [53] P. Macaruso, P. E. Hook, and R. McCabe, “The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students,” *J. Res. Read.*, vol. 29, no. 2, pp. 162–172, 2006.
- [54] Lexia Learning Systems, “Phonics Based Reading and Strategies for Older Students.” Lexia

- Learning Systems, Inc., Lincoln, MA, 2001.
- [55] N. L. Saine, M. K. Lerkkanen, T. Ahonen, A. Tolvanen, and H. Lyytinen, "Computer-Assisted Remedial Reading Intervention for School Beginners at Risk for Reading Disability," *Child Dev.*, vol. 82, no. 3, pp. 1013–1028, 2011.
- [56] S. Hintikka, M. Aro, and H. Lyytinen, "Computerized training of the correspondences between phonological and orthographic units," *Writ. Lang. Lit.*, vol. 8, no. 2, pp. 79–102, 2005.
- [57] H. Lyytinen, M. Ronimus, A. Alanko, A.-M. Poikkeus, and M. Taanila, "Early identification of dyslexia and the use of computer game-based practice to support reading acquisition," *Nord. Psychol.*, vol. 59, no. 2, pp. 109–126, 2007.
- [58] H. Lyytinen, J. Erskine, J. Kujala, E. Ojanen, and U. Richardson, "In search of a science-based application: A learning tool for reading acquisition," *Scand. J. Psychol.*, vol. 50, no. 6, pp. 668–675, 2009.
- [59] E. Ojanen *et al.*, "GraphoGame-A catalyst for multi-level promotion of literacy in diverse contexts," *Front. Psychol.*, vol. 6, no. 671, pp. 1–13, Jun. 2015.
- [60] F. Kyle, J. Kujala, U. Richardson, H. Lyytinen, and U. Goswami, "Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG Rime and GG Phoneme," *Read. Res. Q.*, vol. 48, no. 1, pp. 61–76, 2013.
- [61] R. Rosas, J.-P. Escobar, M.-P. Ramírez, A. Meneses, and A. Guajardo, "Impact of a computer-based intervention in Chilean children at risk of manifesting reading difficulties / Impacto de una intervención basada en ordenador en niños chilenos con riesgo de manifestar dificultades lectoras," *Infanc. y Aprendiziz.*, vol. 40, no. 1, pp. 158–188, 2017.
- [62] B. W. Wise, J. Ring, and R. K. Olson, "Individual Differences in Gains from Computer-Assisted Remedial Reading," *J. Exp. Child Psychol.*, vol. 77, no. 3, pp. 197–235, 2000.
- [63] R. S. Savage, P. Abrami, G. Hipps, and L. Deault, "A randomized controlled trial study of the ABRACADABRA reading intervention program in grade 1.," *J. Educ. Psychol.*, vol. 101, no. 3, pp. 590–604, 2009.
- [64] R. S. Savage, O. Erten, P. Abrami, G. Hipps, E. Comaskey, and D. van Lierop, "ABRACADABRA in the hands of teachers: The effectiveness of a web-based literacy intervention in grade 1 language arts programs," *Comput. Educ.*, vol. 55, no. 2, pp. 911–922, 2010.
- [65] P. Abrami, E. Borohkovski, and L. Lysenko, "The Effects of ABRACADABRA on Reading Outcomes: A Meta-Analysis of Applied Field Research," *J. Interact. Learn. Res.*, vol. 26, no. 4, pp. 337–367, 2015.
- [66] R. Nicolson, A. Fawcett, and M. Nicolson, "Evaluation of a computer-based reading intervention in infant and junior schools," *J. Res. Read.*, vol. 23, no. 2, pp. 194–209, 2000.
- [67] D. Messer and G. Nash, "An evaluation of the effectiveness of a computer-assisted reading intervention," *J. Res. Read.*, pp. 1–19, 2017.
- [68] Easyread, "Trainertext." DM Education, 29 Beaumont Street, Oxford, OX1 2NP, United Kingdom, 2014.
- [69] S. Gustafson, J. Ferreira, and J. Rönnerberg, "DOT: Datorbaserad ortografisk träning." Linköping: Institutionen för beteendevetenskap/Institutet för handikappvetenskap, Linköpings universitet, 2003.
- [70] J. Ecalle, N. Kleinsz, and A. Magnan, "Computer-assisted learning in young poor readers: The effect of grapho-syllabic training on the development of word reading and reading comprehension," *Comput. Human Behav.*, vol. 29, no. 4, pp. 1368–1376, 2013.
- [71] J. Ecalle, A. Magnan, and D. Jabouley, "Chassymo: un logiciel d'aide au traitement syllabique. [Chassymo: A software for grapho-syllabic processing]." Adeprio Diffusion, Châteauroux, 2010.
- [72] P. Revy, "Oppositions Phonologiques [Software Phonological Oppositions]." Gerip, Lyon, 2005.
- [73] G. P. Moser, T. G. Morrison, and B. Wilcox, "Supporting Fourth-Grade Students' Word Identification Using Application Software," *Read. Psychol.*, vol. 38, no. 4, pp. 349–368, 2017.
- [74] G. P. Moser, "8 Great Word Patterns." Bugbrained, LLC, Orem, UT, 2012.
- [75] M. L. Kamil and E. H. Hiebert, "Teaching and Learning Vocabulary: Perspective and Persistent

- Issues,” in *Teaching and learning vocabulary: Bringing Research to Practice*, 2005, pp. 1–23.
- [76] S. A. Stahl, “Four problems with teaching word meanings,” in *Teaching and learning vocabulary: Bringing research to practice*, 2005, pp. 95–114.
- [77] A. E. Cunningham and K. E. Stanovich, “Early reading acquisition and its relation to reading experience and ability 10 years later,” *Dev. Psychol.*, vol. 33, no. 6, pp. 934–945, 1997.
- [78] H. S. Scarborough, S. B. Neuman, and D. Dickinson, “Connecting early language and literacy to later reading (dis) abilities: Evidence, theory, and practice,” in *Approaching difficulties in literacy development: Assessment, pedagogy and programmes*, 2001, pp. 97–110.
- [79] M. Sénéchal and G. Ouellette, “The misunderstood giant: On the predictive role of early vocabulary to future reading,” in *Handbook of early literacy research*, vol. 2, 2006, pp. 173–182.
- [80] B. Hart and T. R. Risley, “The early catastrophe: The 30 million word gap by age 3,” *Am. Educ.*, vol. 27, no. 1, pp. 1–6, 2003.
- [81] E. Hoff, “The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech,” *Child Dev.*, vol. 74, no. 5, pp. 1368–1378, 2003.
- [82] B. Hart and T. R. Risley, *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing, 1995.
- [83] A. Biemiller and N. Slonim, “Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition,” *J. Educ. Psychol.*, vol. 93, no. 3, pp. 498–520, 2001.
- [84] W. Nagy, “Why vocabulary instruction needs to be long-term and comprehensive,” in *Teaching and learning vocabulary: Bringing research to practice*, 2005, pp. 27–44.
- [85] S. B. Neuman and J. Dwyer, “Missing in Action: Vocabulary Instruction in Pre-K,” *Read. Teach.*, vol. 62, no. 5, pp. 384–392, 2009.
- [86] S. J. Kuder, “Vocabulary Instruction for Secondary Students With Reading Disabilities: An Updated Research Review,” *Learn. Disabil. Q.*, vol. 40, no. 3, pp. 155–164, 2017.
- [87] A. E. Cunningham, “Vocabulary growth through independent reading and reading aloud to children,” in *Teaching and learning vocabulary: Bringing research to practice*, 2005, pp. 45–68.
- [88] J. F. Xin and H. Rieth, “Video-Assisted Vocabulary Instruction for Elementary School Students with Learning Disabilities,” *Inf. Technol. Child. Educ. Annu.*, vol. 1, pp. 87–103, 2001.
- [89] T. L. Harris and R. E. Hodges, *The Literacy Dictionary: The Vocabulary of Reading and Writing*. Order Department, International Reading Association, 800 Barksdale Road, P.O. Box 8139, Newark, DE 19714-8139 (Book No. 138: \$25 members, \$35 nonmembers).: International Reading Association, 1995.
- [90] K. Nation, “Children’s reading comprehension difficulties,” *Sci. Read. A Handb.*, pp. 248–265, 2005.
- [91] P. Kendeou, P. Van Den Broek, A. Helder, and J. Karlsson, “A cognitive view of reading comprehension: Implications for reading difficulties,” *Learn. Disabil. Res. Pract.*, vol. 29, no. 1, pp. 10–16, 2014.
- [92] K. Cain and J. Oakhill, “Profiles of children with specific reading comprehension difficulties,” *Br. J. Educ. Psychol.*, vol. 76, no. 4, pp. 683–696, Dec. 2006.
- [93] R. Gersten, L. S. Fuchs, J. P. Williams, and S. Baker, “Teaching Reading Comprehension Strategies to Students With Learning Disabilities: A Review of Research,” *Rev. Educ. Res.*, vol. 71, no. 2, pp. 279–320, 2001.
- [94] M. G. McKeown, I. L. Beck, and R. G. K. Blake, “Rethinking Reading Comprehension,” *Read. Res. Q.*, vol. 44, no. 3, pp. 218–253, 2009.
- [95] J. K. Horne, “Reading Comprehension: A Computerized Intervention with Primary-age Poor Readers,” *Dyslexia*, vol. 23, no. 2, pp. 119–140, May 2017.
- [96] Lucid Research, “Comprehension Booster.” Lucid Research Limited, 3 Spencer Street, Beverley, East Yorkshire. HU17 9EG. U.K., 2010.
- [97] Y. T. Sung, K. E. Chang, and J. S. Huang, “Improving children’s reading comprehension and use of strategies through computer-based strategy training,” *Comput. Human Behav.*, vol. 24,



- no. 4, pp. 1552–1571, 2008.
- [98] R. E. Mayer, “Learning Strategies for Making Sense out of Expository Text : The SOI Model for Guiding Three Cognitive Processes in Knowledge Construction,” vol. 8, no. 4, pp. 357–371, 1996.
  - [99] M. Heimann, M. Lundälv, T. Tjus, and K. Nelson, “Omega–Interactive sentences, a multimedia software for language exploration and play.” Göteborg: Topic Data & Språkbehandling HB, 2004.
  - [100] A. Potocki, J. Ecalte, and A. Magnan, “Effects of computer-assisted comprehension training in less skilled comprehenders in second grade: A one-year follow-up study,” *Comput. Educ.*, vol. 63, pp. 131–140, 2013.
  - [101] H. R. Ponce, M. J. López, and R. E. Mayer, “Instructional effectiveness of a computer-supported program for teaching reading comprehension strategies,” *Comput. Educ.*, vol. 59, no. 4, pp. 1170–1183, 2012.
  - [102] S. L. Klauda and J. T. Guthrie, “Relationships of three components of reading fluency to reading comprehension,” *J. Educ. Psychol.*, vol. 100, no. 2, pp. 310–321, 2008.
  - [103] R. F. Hudson, P. C. Pullen, H. B. Lane, and J. K. Torgesen, “The Complex Nature of Reading Fluency: A Multidimensional View,” *Read. Writ. Q.*, vol. 25, no. 1, pp. 4–32, 2009.
  - [104] M. R. Kuhn, P. J. Schwanenflugel, and E. B. Meisinger, “Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency,” *Read. Res. Q.*, vol. 45, no. 2, pp. 230–251, 2010.
  - [105] M. Wolf and T. Katzir-Cohen, “Reading Fluency and Its Intervention,” *Sci. Stud. Read.*, vol. 5, no. 3, pp. 211–239, 2001.
  - [106] T. Rasinski, “Reading fluency instruction: Moving beyond accuracy, automaticity and prosody,” *Read. Teach.*, vol. 59, no. 7, pp. 704–706, 2006.
  - [107] D. LaBerge and S. J. Samuels, “Toward a theory of automatic information processing in reading,” *Cogn. Psychol.*, vol. 6, no. 2, pp. 293–323, 1974.
  - [108] J. Erikson, “Prosody: The problem of expression in fluency,” *Annu. Meet. Int. Read. Assoc. Orlando, FL*, 2003.
  - [109] N. Calet, N. Gutiérrez-Palma, and S. Defior, “Effects of fluency training on reading competence in primary school children: The role of prosody,” *Learn. Instr.*, 2016.
  - [110] R. Patel and C. McNab, “Displaying prosodic text to enhance expressive oral reading,” *Speech Commun.*, vol. 53, no. 3, pp. 431–441, 2011.
  - [111] S. J. Samuels, “The Method of Repeated Readings,” *Read. Teach.*, vol. 32, no. 4, pp. 403–408, 1979.
  - [112] R. G. Heckelman, “A neurological-impress method of remedial-reading instruction,” *Acad. Ther.*, vol. 4, no. 4, pp. 277–282, 1969.
  - [113] F. Greene, “Radio reading,” *Read. Compr. four Linguist. levels*, pp. 104–107, 1979.
  - [114] K. Topping, “Paired Reading: A Powerful Technique for Parent Use,” *Read. Teach.*, vol. 40, no. 7, pp. 608–614, 1987.
  - [115] L. C. Hunt, “The Effect of Self-Selection, Interest, and Motivation upon Independent, Instructional, and Frustration Levels,” *Read. Teach.*, vol. 24, no. 2, pp. 146–151, 1970.
  - [116] Renaissance Learning, “Accelerated Reader.” Renaissance Learning, P.O. Box 8036, Wisconsin Rapids, WI 54495; (800) 338-4204., 1986.
  - [117] E. a. Stevens, M. a. Walker, and S. Vaughn, “The effects of reading fluency interventions on the reading fluency and reading comprehension performance of elementary students with learning disabilities: A synthesis of the research from 2001 to 2014,” *J. Learn. Disabil.*, pp. 1–15, 2016.
  - [118] M. K. Kim, D. P. Bryant, B. R. Bryant, and Y. Park, “A synthesis of interventions for improving oral reading fluency of elementary students with learning disabilities,” *Prev. Sch. Fail. Altern. Educ. Child. Youth*, vol. 61, no. 2, pp. 116–125, 2017.
  - [119] J. Lee and S. Y. Yoon, “The Effects of Repeated Reading on Reading Fluency for Students With Reading Disabilities,” *J. Learn. Disabil.*, vol. 50, no. 2, pp. 213–224, Mar. 2017.
  - [120] J. A. Nunnery, S. M. Ross, and A. McDonald, “A Randomized Experimental Evaluation of the Impact of Accelerated Reader/Reading Renaissance Implementation on Reading Achievement in Grades 3 to 6,” *J. Educ. Students Placed Risk*, vol. 11, no. 1, pp. 1–18, 2006.

- 
- [121] L. C. Shannon, M. K. Styers, S. B. Wilkerson, and E. Peery, "Computer-Assisted Learning in Elementary Reading: A Randomized Control Trial," *Comput. Sch.*, vol. 32, no. 1, pp. 20–34, 2015.
  - [122] R. A. Johnson and C. A. Howard, "The effects of the accelerated reader program on the reading comprehension of pupils in grades three, four and five," *Read. Matrix*, vol. 3, no. 3, pp. 30–31, 2003.
  - [123] C. Clark and A. E. Cunningham, "Reading Enjoyment , Behaviour and Attitudes in Pupils who use Accelerated Reader," 2016.
  - [124] C. López-escribano, "Training Reading Fluency and Comprehension of Spanish Children with Dyslexia," in *Reading Fluency*, Springer International Publishing, 2016, pp. 141–161.
  - [125] Z. Breznitz and M. Nevat, "The reading acceleration program (RAP)." The Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Israel, 2004.
  - [126] D. D. Paige, "Testing the Acceleration Hypothesis: Fluency Outcomes Utilizing Still-Versus Accelerated-Text in Sixth-Grade Students With Reading Disabilities," *Lit. Res. Instr.*, vol. 50, no. 4, pp. 294–312, 2011.
  - [127] P. Snellings, A. van der Leij, P. F. de Jong, and H. Blok, "Enhancing the reading fluency and comprehension of children with reading disabilities in an orthographically transparent language.," *J. Learn. Disabil.*, vol. 42, no. 4, pp. 291–305, 2015.
  - [128] T. Nagler *et al.*, "Text-fading based training leads to transfer effects on children's sentence reading fluency," *Front. Psychol.*, vol. 6, pp. 1–8, 2015.
  - [129] J. Niedo, Y. Lee, Z. Breznitz, and V. W. Berninger, "Computerized Silent Reading Rate and Strategy Instruction for Fourth Graders at Risk in Silent Reading Rate," *Learn. Disabil. Q.*, vol. 37, no. 2, pp. 100–110, 2014.
  - [130] E. G. Steenbeek-Planting, W. H. J. van Bon, and R. Schreuder, "Improving word reading speed: Individual differences interact with a training focus on successes or failures," *Read. Writ.*, vol. 25, no. 9, pp. 2061–2089, 2012.
  - [131] K. Van Den Bosch, W. H. J. van Bon, and R. Schreuder, "Poor readers' decoding skills: Effects of training with limited exposure duration," *Read. Res. Q.*, vol. 30, no. 1, pp. 110–125, 1995.
  - [132] E. G. Steenbeek-Planting, W. H. J. Van Bon, and R. Schreuder, "Improving the reading of bisyllabic words that involve context-sensitive spelling rules: Focus on successes or on failures?," *Read. Writ.*, vol. 26, no. 9, pp. 1437–1458, 2013.
  - [133] R. Patel, H. Kember, and S. Natale, "Feasibility of augmenting text with visual prosodic cues to enhance oral reading," *Speech Commun.*, vol. 65, pp. 109–118, 2014.
  - [134] J. S. Kim, J. F. Samson, R. Fitzgerald, and A. Hartry, "A randomized experiment of a mixed-methods literacy intervention for struggling readers in grades 4-6: Effects on word reading efficiency, reading comprehension and vocabulary, and oral reading fluency," *Read. Writ.*, vol. 23, no. 9, pp. 1109–1129, 2010.
  - [135] J. S. Kim, L. Capotosto, A. Hartry, and R. Fitzgerald, "Can a Mixed-Method Literacy Intervention Improve the Reading Achievement of Low-Performing Elementary School Students in an After-School Program?: Results From a Randomized Controlled Trial of READ 180 Enterprise," *Educ. Eval. Policy Anal.*, vol. 33, no. 2, pp. 183–201, 2011.
  - [136] B. Chambers, R. E. Slavin, N. A. Madden, P. Abrami, M. K. Logan, and R. Gifford, "Small-Group, Computer-Assisted Tutoring to Improve Reading Outcomes for Struggling First and Second Graders," *Elem. Sch. J.*, vol. 111, no. 4, pp. 625–640, Jun. 2011.
  - [137] B. Danis, S. Rainville, M. Therrien, B. Tucker, and P. Abrami, "Alphie's Alley Early Literacy Tutoring Software," in *ED-MEDIA 2005--World Conference on Educational Multimedia, Hypermedia & Telecommunications*, 2005, pp. 3829–3834.
  - [138] B. Chambers *et al.*, "Computer-Assisted Tutoring in Success for All: Reading Outcomes for First Graders," *J. Res. Educ. Eff.*, vol. 1, no. 2, pp. 120–137, Apr. 2008.
  - [139] J. Herron, "Read, write, & type." Freemont, CA: The Learning Company, 1995.
  - [140] M. van de Ven, L. de Leeuw, M. van Weerdenburg, and E. G. Steenbeek-Planting, "Early reading intervention by means of a multicomponent reading game," *J. Comput. Assist. Learn.*, vol. 33, no. 4, pp. 320–333, 2017.
  - [141] E. G. Steenbeek-Planting, M. Boot, J. C. de Boer, M. Van de Ven, N. M. Swart, and D. van der

- Hout, "Evidence-based psycholinguistic principles to remediate reading problems applied in the playful app Letterprins: A perspective of quality of healthcare on learning to read.," in *Games for Health*, Wiesbaden: Springer Fachmedien Wiesbaden, 2013, pp. 281–291.
- [142] D. Schneider, A. Chambers, N. Mather, R. Bauschatz, M. Bauer, and L. Doan, "The Effects of an ICT-Based Reading Intervention on Students' Achievement in Grade Two," *Read. Psychol.*, vol. 37, no. 5, pp. 793–831, 2016.
- [143] MindPlay, "MindPlay Virtual Reading Coach." Mindplay: Research-Based Educational Software for Reading Instruction, 4400 E. Broadway Blvd., Suite 400 Tucson, AZ 85711, 2012.
- [144] J. C. Cassady and L. L. Smith, "The Impact of a Structured Integrated Learning System on First-Grade Students' Reading Gains," *Read. Writ. Q.*, vol. 21, no. 4, pp. 361–376, 2005.
- [145] P. Tzouveli, A. Schmidt, M. Schneider, A. Symvonis, and S. Kollias, "Adaptive reading assistance for the inclusion of students with dyslexia: The AGENT-DYSL approach," in *The 8th IEEE International Conference on Advanced Learning Technologies, ICALT 2008*, 2008, pp. 167–171.
- [146] C. Y. Lin, W. J. Yu, W. J. Chen, C. W. Huang, and C. C. Lin, "The Effect of Literacy Learning via Mobile Augmented Reality for the Students with ADHD and Reading Disabilities," in *International Conference on Universal Access in Human-Computer Interaction*, 2016, pp. 103–111.
- [147] S. M. Daud and H. Abas, "'Dyslexia baca' mobile app - The learning ecosystem for dyslexic children," in *Proceedings - 2013 International Conference on Advanced Computer Science Applications and Technologies, ACSAT 2013*, 2014, pp. 412–416.
- [148] L. Rello, C. Bayarri, and A. Gorriz, "What is wrong with this word? dysegxixia: a game for children with dyslexia," in *14th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2012*, 2012, pp. 219–220.
- [149] L. Rello, C. Bayarri, and A. Gòrriz, "Dyslexia exercises on my tablet are more fun," in *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility - W4A '13*, 2013, p. 20.
- [150] T. Bittencourt, J. Savino, H. Fernandes, and L. H. Boueri Rebello, "Mobile Device Development and Its Contribution to the Treatment of Young Dyslexic Brazilian Children," in *Proceedings of the AHFE 2016 International Conference on Design for Inclusion*, 2016, pp. 339–349.
- [151] J. K. Torgesen, C. A. Rashotte, and R. K. Wagner, "TOWRE: Test of word reading efficiency." Pro-ed, Austin, TX, 1999.
- [152] R. W. Woodcock, K. S. McGrew, and N. Mather, "Woodcock-Johnson tests of achievement." Riverside Publishing, Itasca, IL, 2001.
- [153] F. A. Schrank, N. Mather, and K. S. McGrew, "Woodcock-Johnson IV tests of achievement." Riverside, Rolling Meadows, IL, 2014.
- [154] M. B. Denckla and R. Rudel, "Rapid 'Automatized' Naming of Pictured Objects, Colors, Letters and Numbers by Normal Children," *Cortex*, vol. 10, no. 2, pp. 186–202, 1974.
- [155] M. Wolf and M. B. Denckla, "RAN/RAS: Rapid automatized naming and rapid alternating stimulus tests." Pro-ed, Austin, TX, 2005.
- [156] W. H. Macginitie, R. K. Macginitie, R. B. Cooter, J. Cooter, and S. Curry, "Assessment: Gates-Macginitie Reading Tests, Third Edition," *Read. Teach.*, vol. 43, no. 3, pp. 256–258, 1989.
- [157] W. H. MacGinitie, R. K. MacGinitie, K. Maria, L. G. Dreyer, and K. E. Hughes, "Gates-MacGinitie Reading Test (GMRT) Fourth Edition." Riverside Publishing, Itasca, IL, 2000.
- [158] S. Jastak and G. Wilkinson, "Wide range achievement test-revised." Jastak Associates, Wilmington, DE, 1984.
- [159] R. Wagner, J. Torgesen, and C. Rashotte, "Comprehensive test of phonological processes (CTOPP)." Pro-Ed, Austin, TX, 1999.
- [160] J. L. Wiederholt and B. R. Bryant, "Gray Oral Reading Tests: GORT-3." Pro-Ed, Austin, TX, 1992.
- [161] J. L. Wiederholt and B. R. Bryant, "Gray oral reading test-(GORT-4)." Pro-Ed, Austin, TX, 2001.
- [162] K. T. Williams, "Group Reading Assessment and Diagnostic Evaluation: Teacher's scoring & interpretive manual." American Guidance Service, Circle Pines, MN, 2001.

- [163] C. D. Elliott, "BAS II: British Ability Scales Second Edition." NFER-Nelson, London, 1996.
- [164] J. Écalé, "Timé2: test d'identification de mots écrits de 6 à 8 ans." EAP, Éditions et applications psychologiques., 2004.
- [165] National Reading Panel, "Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction," *NIH Publ. No. 00-4769*, 2000.
- [166] A. G. Bus and M. H. van IJzendoorn, "Phonological awareness and early reading: A meta-analysis of experimental training studies," *J. Educ. Psychol.*, vol. 91, no. 3, pp. 403–414, 1999.
- [167] F. R. Vellutino, J. M. Fletcher, M. J. Snowling, and D. M. Scanlon, "Specific reading disability (dyslexia): What have we learned in the past four decades?," *J. Child Psychol. Psychiatry Allied Discip.*, vol. 45, no. 1, pp. 2–40, 2004.
- [168] G. P. Ouellette, "What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension," *J. Educ. Psychol.*, vol. 98, no. 3, pp. 554–566, 2006.
- [169] L. B. Abraham, "Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis," *Comput. Assist. Lang. Learn.*, vol. 21, no. 3, pp. 199–226, 2008.
- [170] Statista, "Number of smartphone users worldwide from 2014 to 2020 (in billions)." [Online]. Available: <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>. [Accessed: 10-Oct-2017].
- [171] Y. Park, E. A. Chaparro, J. Preciado, and K. D. Cummings, "Is Earlier Better? Mastery of Reading Fluency in Early Schooling," *Early Educ. Dev.*, vol. 26, no. 8, pp. 1187–1209, 2015.
- [172] M. W. Lovett, J. C. Frijters, M. Wolf, K. A. Steinbach, R. A. Sevcik, and R. D. Morris, "Early intervention for children at risk for reading disabilities: The impact of grade at intervention and individual differences on intervention outcomes," *J. Educ. Psychol.*, 2017.
- [173] J. S. Chall and V. A. Jacobs, "The classic study on poor children's fourth-grade slump," *Am. Educ.*, vol. 27, no. 1, pp. 14–15, 2003.
- [174] S. Deterding, M. Sicart, L. Nacke, K. O'Hara, and D. Dixon, "Gamification. using game-design elements in non-gaming contexts," *Proc. 2011 Annu. Conf. Ext. Abstr. Hum. factors Comput. Syst. - CHI EA '11*, p. 2425, 2011.
- [175] D. Gooch, A. Vasalou, L. Benton, and R. Khaled, "Using Gamification to Motivate Students with Dyslexia," *Proc. 2016 CHI Conf. Hum. Factors Comput. Syst. - CHI '16*, pp. 969–980, 2016.
- [176] G. Y. Hong and M. Masood, "Effects of gamification on lower secondary school students' motivation and engagement," *Int. J. Soc. Educ. Econ. Manag. Eng.*, vol. 8, no. 12, pp. 3750–3757, 2014.
- [177] T. Bell, "Extensive reading: speed and comprehension," *Read. Matrix*, vol. 1, no. 1, 2001.
- [178] C. López-Escribano, P. Sánchez-Hípola, J. Suro Sánchez, and F. Leal Carretero, "Comparative Analysis of Rapid Automatized Naming Studies in Spanish and Reading Acquisition and Reading Difficulties," *Univ. Psychol.*, vol. 13, no. 2, pp. 757–769, 2014.



## Contents

|  |           |
|--|-----------|
| <b>CHAPTER III Scientific Approach.....</b>              | <b>50</b> |
| <b>1. Introduction .....</b>                             | <b>50</b> |
| <b>2. Objective.....</b>                                 | <b>51</b> |
| <b>3. Problem Statement .....</b>                        | <b>51</b> |
| <b>4. Scientific Ground.....</b>                         | <b>52</b> |
| 4.1. Automaticity of Grapheme-Phoneme Associations ..... | 52        |
| 4.2. Extensive Vocabulary Instruction .....              | 53        |
| 4.3. Gamification.....                                   | 54        |
| 4.4. Home-Based Training .....                           | 55        |
| 4.5. Cross-Platform Design .....                         | 55        |
| 4.6. Intelligent Adaptive Instruction .....              | 55        |
| 4.7. Reading Tasks .....                                 | 56        |
| <b>5. System Architecture .....</b>                      | <b>57</b> |
| <b>6. The Novelty of the Approach.....</b>               | <b>58</b> |
| <b>7. Conclusion .....</b>                               | <b>59</b> |
| <b>8. References.....</b>                                | <b>60</b> |

## CHAPTER III

### Scientific Approach

#### 1. Introduction

This chapter presents our approach for the remediation of dyslexia and the scientific ground for each proposed solution. Slow and effortful reading is an important characteristic of dyslexia. A universal and long-term symptom of dyslexia is poor Rapid Automatized Naming (RAN) [1]. It is suggested that RAN is a better predictor of word reading fluency than Phonological Awareness (PA) [2], [3]. Studies attribute RAN to the speed of access to the phonological representations [4]. The automaticity of access to the phonological representation of written forms is a hallmark of skilled reading and is missing in dyslexics [5]. Throughout the literature, countless studies could be found that have pointed to the importance of automaticity in reading. However, Chapter 2 showed that working on automaticity is neglected in early grades and is neglected even more in languages with opaque orthographies. Opaque (or deep) orthographies such as English and French are the writing systems with inconsistent letter-sound correspondences. It contrasts with transparent (or shallow) orthographies such as Italian and Spanish with consistent letter-sound associations.

A language has an opaque orthography mainly due to the lack of modern reform in its writing system. Language is a dynamic phenomenon and cannot be confined precisely in the prison of a writing system. Over time, pronunciations change in every language. Writing systems, however, are less dynamic, especially with the invention of printing and standardization of orthography, there is always a resistance to change. For example, since the 15th century, when William Caxton introduced the printing press to England, the English language went on major sound changes such as the Great Vowel Shift and silent consonants. However, due to the general inertia of writing systems, English orthography did not follow these sound changes. Besides, during the renaissance, English was an insatiable language borrowing and inventing thousands of technical and scientific words from Latin and Greek roots that did not always comply with the English language's writing conventions. Furthermore, some scholars actively meddled in the spelling of the language to reflect the etymology rather than pronunciation. An example is the spelling change of the word "dette" to "debt" or "doute" to "doubt" to reflect their Latin etymology, even though the letter "b" was no longer pronounced in the English language. These reasons altogether contributed to the inconsistencies of the English orthography.

Opaque orthographies can cause a significant delay in the acquisition of reading [7]. Therefore, in this thesis, a set of facilitating and enhancing approaches are proposed to reduce the reading acquisition time of an opaque orthography. In comparison to transparent writing systems, in opaque orthographies, there are far more grapheme-phoneme associations to master. Reaching the automaticity level in the conversion of grapheme-phoneme associations is one of the fundamental steps in becoming a skilled reader. Hence, intensive and systematic practice is required to acquire the automaticity of grapheme-phoneme conversion in an opaque

orthography. Another feature of opaque orthographies is that fluent reading depends considerably on vocabulary knowledge because the readers regularly face the words that cannot be decoded by simple letter-sound conversion. They should guess the word based on their vocabulary knowledge or the context [8]. Therefore, vocabulary intervention has more importance in the instruction of opaque orthographies. However, for being effective, a vocabulary intervention requires to be sufficiently extensive [9].

In any extensive training program, adherence to the program is fundamental. The Gamification concept is proposed to increase motivation, engagement, and adherence [10], [11]. Moreover, normal schools do not possess enough resources to offer extensive training programs. Home-based training proposed in this study brings several advantages, such as enhanced accessibility and increased individualization. However, a home-based approach requires intelligence to adapt to the ability level of the learner. For achieving an acceptable level of Adaptiveness, a combination of optimization model and artificial neural network is proposed. The optimization model maximizes the value of the training session while not surpassing the maximum difficulty level of the session. The artificial neural network and linear regression models predict the content's difficulty level through a set of lexical variables such as word length, frequency, and grapheme-phoneme consistency.

Section 2 reiterates the objective of this study. Section 3 outlines the problems and disadvantages of current reading programs identified in the analytical review of the published research presented in Chapter 2. Section 4 explains each of the facets of the proposed approach for mitigating the current problems. Section 5 proposes the general architecture of the system. Section 6 presents the novelty of the proposed approach, and finally, Section 7 concludes this chapter by summarizing the proposed approach.

## **2. Objective**

The objective of this study is to facilitate and enhance the reading acquisition of opaque orthographies among young dyslexic children. In terms of facilitating, it refers to home-based learning and unlimited access to the designed pedagogical system, as well as the gamification to increase motivation and adherence. In terms of enhancement, it refers to offering a systematic and individualized practice based on the scientific principles of automaticity acquisition. Finally, opaque orthographies (such as English and French) refer to the writing systems with inconsistent letter-sound correspondences.

## **3. Problem Statement**

The results of the analytical review of the published research revealed several issues listed below.

- I. Working on the automaticity aspect of reading is generally neglected in elementary schools, and it is specifically neglected in earlier grades.
- II. There is a clear lack of a holistic approach to fluency interventions, and each of the current interventions addresses a single aspect of fluency.



- III. Fluency approaches were short, which does not correspond to the extensive practice time required for acquiring it. This lack of sufficient training may be attributed to the limited time available in schools, and a home-based approach is a potential solution.
- IV. Vocabulary interventions were overlooked. It is disadvantageous for opaque orthographies, in which decoding depends heavily on making analogies based on the current vocabulary knowledge [8]. School-based interventions do not allow the extensive time required for influencing general vocabulary knowledge. Therefore, home-based interventions can be a viable solution.
- V. Fluency interventions were not motivating. Gamification could allow learners practicing for longer durations, which is vital for fostering fluency.
- VI. In opaque orthographies, the studies mainly focused on accuracy issues through phonics approaches. However, dyslexia is a problem of both accuracy and speed, and none of them should be neglected.
- VII. Non-computer technologies, such as tablets and smartphones, were completely overlooked.
- VIII. There was a lack of home-based approaches among the studied reading programs.

## **4. Scientific Ground**

In this section, the scientific foundation of the approach for the remediation of dyslexia in opaque orthographies is presented. It is constructed specifically to overcome the challenges of opaque orthographies while addressing the issues raised in the Problem Statement Section.

### **4.1. Automaticity of Grapheme-Phoneme Associations**

Reading is an automatic skill. A skilled reader does not process words in a conscious and effortful letter-by-letter decoding. Hence, the mere knowledge of letters does not lead to efficient word recognition. It means that orthographic-phonological associations should become automatic. According to the automatic orthographic-phonological binding theory of dyslexia, the automaticity of letter-speech sound is the key to the success or failure of reading development [12], [13]. However, this theory was not accurate for English language that has an opaque orthography [14]. This may be explained by another theory called grain size theory [8], stating that decoding in different languages relies on different unit sizes of the language. It means that for a language with opaque orthography such as English, the automaticity should be achieved not only at the letter-sound level but also at the grapheme-phoneme level.

Moreover, the automaticity theory of dyslexia proposed by R.I. Nicolson and A.J. Fawcett also attributes the cause of this condition to the difficulties in procedural learning and,

consequently, the deficits in automaticity of reading [15]–[17]. Even if these theories are not valid, and the deficit in automaticity acquisition is not the real cause of dyslexia, this does not exclude the fact that reading is an automatic process and it should be trained as a skill that some of its sub-parts should reach automatic levels. Grapheme-phoneme association is the most basic process that should become automatic in readers. However, in opaque orthographies, the number of these associations is much higher than the number of letters in the language. All these associations should be practiced and overlearned to the mastery level. Phonics approaches that teach the associations have been used more in countries with opaque orthographies because there are many more associations to acquire than transparent orthographies. Knowledge of the grapheme-phoneme associations is the first step in reading instruction. However, the next step of acquiring the automaticity in grapheme-phoneme associations is more crucial.

Acquiring grapheme-phoneme automaticity in an opaque orthography like English, which has more than 300 common associations, is a time-consuming process. This is why English children are far behind the children of countries with transparent orthographies [7]. However, the analytical review chapter pointed out that automaticity is neglected, especially in earlier grades, particularly in languages with opaque orthographies with more associations required to become automatic. Even if the instructors decide to follow the automaticity acquisition approach, what are the principles for training the automaticity acquisition?

Automaticity is a sub-part of skill, and hence, automaticity acquisition is a sub-process of skill acquisition [18]. Despite the importance of automaticity, the studies focusing on modeling the effective training approaches for automaticity acquisition are scarce. Instead, most of the studies have targeted the broad aspects of skill acquisition. Therefore, Chapter 4 is dedicated to identifying and modelling the effective training principles of automaticity acquisition. The training mechanisms in the reading tasks should be designed based on that model to achieve automaticity in grapheme-phoneme conversion.

## **4.2. Extensive Vocabulary Instruction**

In opaque orthographies, vocabulary knowledge plays a big role in reading, as in numerous cases in those languages letter by letter decoding will not lead to the correct pronunciation. The readers should have a sufficient level of vocabulary knowledge to guess the right pronunciation directly or by making an analogy from the related words that they already know [30]. However, it has been shown that most of the traditional vocabulary interventions were not satisfactorily effective in having an impact on general vocabulary knowledge [31]. Vocabulary interventions should be sufficiently extensive for being effective [32], which is often not viable in normal schooling with limited resources.

In highly opaque orthographies, the reliance on vocabulary is high, putting two categories of students at a real disadvantage. The first category is the students whose parents do not speak the language at home or speak the language as a second language. The reduced exposure to the language will drastically reduce the size of the vocabulary knowledge [19]. The second category is dyslexic students who have a delay in reading skills, and because of their difficulty, they feel demotivated for engaging in reading activities. The reduced exposure to written material decreases the growth of vocabulary knowledge [20], [21]. These disadvantages in vocabulary knowledge widen the gap between students.

Bridging the vocabulary gap has been the subject of many studies [22], [23]. However, the results of the analytical review showed that technology-based vocabulary interventions are completely overlooked. Nevertheless, approaches that target the remediation of dyslexia in an opaque orthography should include extensive vocabulary instruction. It is suggested that any vocabulary intervention should be sufficiently extensive for affecting general vocabulary knowledge [9]. For this reason, extensive vocabulary instruction is proposed in this research.

### 4.3. Gamification

As stated in previous sections, reading instruction, particularly in opaque orthographies, should be extensive. In an extensive training program, adherence is crucial. Learning to read can be hard and laborious, especially when the child is underachieving and feeling the pressure of lagging behind. Having difficulty in learning to read can lead to demotivation and discouragement for engaging in reading activities, and this can further widen the gap between achieving and underachieving readers [9], [24], [25]. For poor readers, reading activities, by themselves, do not induce enough motivation to encourage persistence and adherence, and therefore, instructors need to target both reading skills and motivation [26]. Hence, the gamification concept and serious game design are proposed as mechanisms to increase motivation, engagement, and persistence in the process of reading acquisition.

Gamification is defined as "using game design elements in non-gaming context" [10]. It is shown that gamification can make the school lessons more interesting and motivating for the students [27]. It can positively affect the learner's attitude and behavior towards learning tasks [28]. In addition, it can be used to improve the learners' engagement [29], [30], as well as their autonomy and creativity [31]. Furthermore, despite increasing cognitive load, it can lead to higher achievement among students [32].

With the artistic aspects of graphic design and storytelling, video game design has always been partly in the realm of art. In addition, with the abundance of game mechanics, and the arbitrary mode of choosing them, game design has kept a considerable distance from the rigorous scientific methods. However, in serious game design, some researchers have tried to frame the process of game design based on scientific theories. An example is the LM-GM model [33] in which the authors extracted learning mechanics and game mechanics from the scientific literature and linked the two sets to each other through a classification based on Bloom's higher-order skills [34], [35]. However, this model is too general; it does not differentiate the distinct requirements of each category of learning, including automaticity acquisition, which is a focus of this study. In addition, it did not consider motivational theories such as self-efficacy [36], [37] and self-determination [38], [39]. In this thesis, it is aimed to develop effective gamification models. This requires conducting experimental studies for evaluating the effectiveness of different game design scenarios. In Chapter 5, two gamification models are proposed, and their effectiveness is evaluated through randomized controlled trials. The first gamification model is based on the motivational theories of self-determination and self-efficacy. The second gamification model is specifically designed for automaticity acquisition, which is the focus of this research.

#### **4.4. Home-Based Training**

Some studies have pointed out that for treating dyslexia, the instruction duration should be between 80 and 100 hours, while their healthy peers need 30 to 60 hours [40], [41]. The extra amount of needed instruction is often not provided in schools. Technology-based in-home approaches can potentially mitigate this lack of sufficient instructions at school. These approaches offer several advantages to traditional instruction. It can be accessed at any time and in any location. Therefore, individual training time can be increased, and also, the instructor's time can be freed up for doing other tasks [42]. In addition, the COVID-19 pandemic, which led to unprecedented global lockdowns, has highlighted the value of home-based educational tools [43].

In this study, a home-based approach is proposed to provide sufficient training time required to develop automaticity in reading and the extensive vocabulary instruction needed for influencing the generalized vocabulary knowledge.

#### **4.5. Cross-Platform Design**

Almost all of the current reading programs studied in the previous chapter were computer-based. The use of ubiquitous technologies, such as smartphones and tablets, was completely overlooked. However, the use of computers is in decline. Tablets and smartphones are becoming computationally powerful and increasingly user-friendly, and currently, they are substituting traditional computers in performing common digital tasks. Therefore, in this research, it is proposed to develop the system compatible with major platforms, including Microsoft Windows, Apple's Mac, iOS, and Android. Unity is a game design engine that offers this chance to develop cross-platform games. However, all other technologies incorporated into the system should follow cross-platform design too. This includes technologies such as databases, speech recognition, text-to-speech. Finally, for each platform, the games should be able to offer appropriate interaction styles.

#### **4.6. Intelligent Adaptive Instruction**

It is suggested that reading interventions should be systematic [44]. Technology-based in-home remedial programs can offer undeniable advantages. However, a crucial aspect of these programs is their ability to systematically adapt the learning material to the users' capabilities. If the materials are too difficult, the user will get frustrated or anxious, and if the learning materials are too easy, the user will feel bored [45]. In addition to the adaptive difficulty level, the program should also provide the content with maximum educational value. These intelligent characteristics are essential for any in-home interventional programs. Traditionally, it is the role of the instructor to choose the appropriate content. However, in an in-home approach, the intelligent component of the game plays the role of the instructor.

This study proposes using an intelligent component based on the knapsack optimization model, predictive models such as neural network and regression, as well as threshold adaptive mechanisms. The optimization model has the role of maximizing the value of the content while respecting the user's ability level. The artificial neural network has the role of predicting the difficulty level of the content based on the type of the task. Threshold adaptive mechanisms

increase or decrease the content's educational value based on the user's interaction with the contents. In addition, adaptive mechanisms can increase or decrease the maximum difficulty level of the sessions based on the user's overall performance. Furthermore, game mechanics such as the speed of the game should be adaptive as well. Therefore, it is proposed to use similar threshold adaptive mechanisms for the game mechanics that affect the games' difficulty level.

Depending on each session's length, the grade level of the user, and his/her past performances, the optimal educational content should be chosen. Traditionally, in the reading education, the first words are chosen from the most common and simplest words, and then gradually, the less common and more difficult words are introduced. However, this process is not followed faithfully and accurately because it requires linguistic knowledge and deep computations. The system proposed in this study aims at solving this issue by choosing the most optimal content. For example, if ten words should be selected for instruction from a list of 100 words, there are  $1.73E+13$  possibilities. Finding the optimal combination that maximizes the educational value and does not surpass the student's abilities is an optimization problem that is too complex for any instructor to solve. Even computers cannot solve this kind of problem without the help of meta-heuristic approaches. The Knapsack optimization problem, which is already used in education [46], has the appropriate elements required for modeling the reading tasks in this research. In this study, it is proposed to use a meta-heuristic approach called Genetic Algorithm (GA) to solve the knapsack problem.

The knapsack model has parameters such as value and weight. If these parameters are not accurately set, the optimal solution will not be precise. Word frequency in the language could be a reasonably good mark of each word's value. Knowing a word with a higher frequency is more valuable for the reader because he/she will face the word more often. However, the case for word difficulty is not as simplistic. Many lexical variables are involved in determining word difficulty. The relation between these lexical variables is not clear. Therefore, in this thesis, it was proposed to use predictive models such as artificial neural networks and linear regression to predict word difficulty. Current lexical databases containing many lexical variables can be coupled with lexical mega-studies in behavioral psychology testing children or adults' performance on a considerable amount of words. When these two sources are coupled together, their relationship can be analyzed to build predictive models. If the behavioral studies have been conducted with children and the data for each grade is separate, then the predictive model can predict the word difficulty based on the student's grade level. The resulting model can also estimate the difficulty level of the sentences or passages by calculating the difficulty level of each word present in the text and averaging the total. However, the difficulty level of words varies depending on the type of the task. For example, a word may be difficult in writing but easy in reading. This creates a requirement to have multiple word difficulty models. In this thesis, it is proposed to create four word-difficulty models of auditory and visual word recognition, word decoding, and spelling. These four lexical skills cover most of early reading and writing tasks.

#### **4.7. Reading Tasks**

Four different reading tasks are proposed for mitigating the issues raised in the problem statement section and addressing the specific needs of opaque orthographies highlighted in the

Scientific Ground section. The first two tasks are called Accelerated Word Decoding (AWD) and Accelerated Word-Sound Recognition (AWSR). These two reading tasks act as the opposite of each other. In the first task (AWD), the word is presented whereupon the user should decode it and read it aloud. Speech recognition (SR) technology is used to decide whether the pronunciation was accurate or not. However, in the second task (AWSR), the user hears the pronunciation of a word through text to speech (TTS) technology, and then he/she should recognize the pronounced word among four possible choices. Depending on the users' level, the three distractors can be either random words for novices or orthographically similar words for more advanced users.

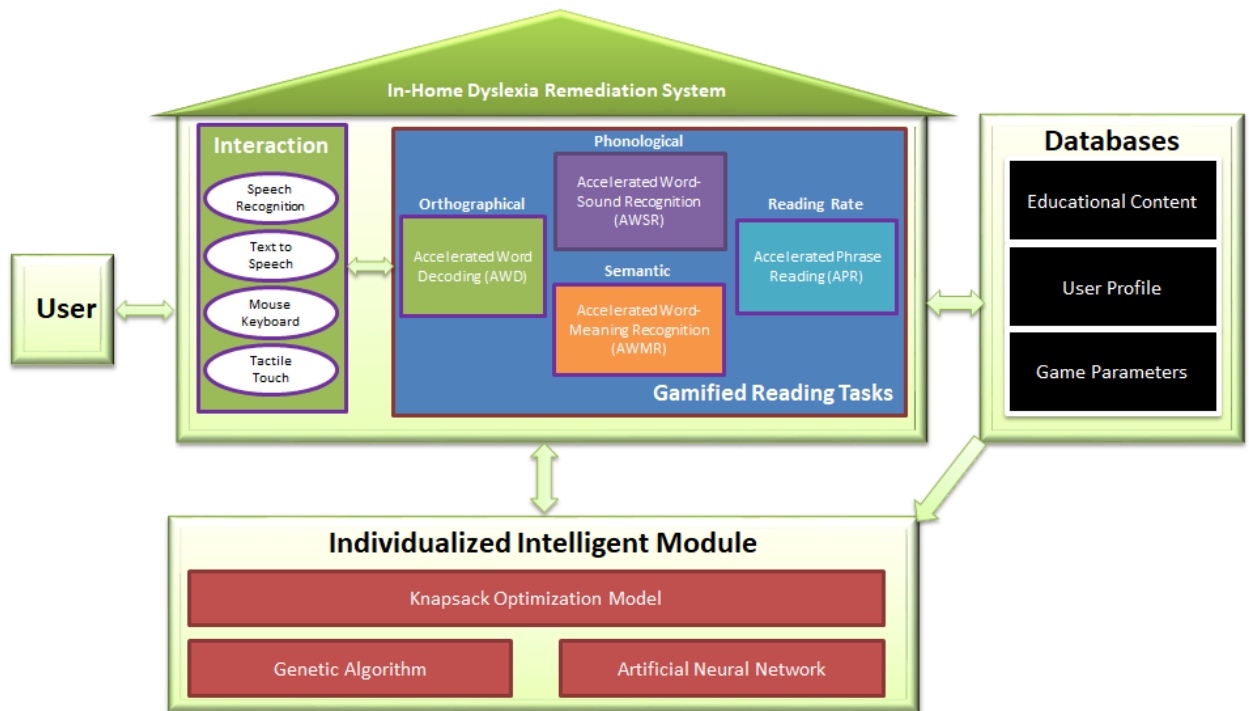
The third fluency task is called Accelerated Phrase Reading (APR), which focuses on increasing the reading rate. This task is based on a text-fading method developed by Breznitz et al. [47]. It presents a phrase to the user, and the text starts to disappear letter by letter from the beginning. Once the text is completely disappeared or the user signaled that he/she has finished reading, a four-choice comprehension question will be presented to the user. Depending on the answer, the text-fading speed will be increased or decreased. This keeps the users on the edge of their reading rate and forces them to read the phrases faster.

Finally, the reading task in the vocabulary module that is called Accelerated Word-Meaning Recognition (AWMR) is based on cloze questions. It represents a phrase with one word missing, and the user has to select the appropriate word and ignore the distractors. This task is primarily designed to improve the user's vocabulary. However, it can increase automaticity in word decoding and sentence reading.

## 5. System Architecture

Based on the Scientific Ground section of this chapter and in order to mitigate the issues raised in the Problem Statement section, a gamified interactive approach for the remediation of dyslexia in opaque orthographies is proposed. It primarily focuses on fostering the automaticity of decoding at grapheme-phoneme, and phoneme-grapheme level, which is the biggest challenge in the instruction of opaque orthographies. It also lays the ground for a gamified in-home vocabulary instruction, which is specifically important in acquiring opaque orthographies. Finally, to propose a more holistic approach to fluency, besides the reading tasks proposed for grapheme-phoneme and phoneme-grapheme automaticity, another gamified fluency task targets the reading rate.

The diagram below (Figure 1) presents the overall design of the proposed system. The core of the system is the four gamified reading tasks explained previously. However, these gamified tasks should be connected to the educational databases to get their reading content. They should also be connected to the user profile databases in the cloud to enable home-based training. A game parameter database that is also important for home-based training is proposed so that the instructor and the developer could modify and adjust the parameters of the games for students from the cloud. The intelligent component is necessary for tailoring the content of the games for each learner. Finally, multimodal interaction is proposed that includes speech recognition, speech synthesis, tactile touch, mouse, and keyboard.



**Figure 1.** The main architecture of the system

## 6. The Novelty of the Approach

The novelty of the proposed scientific approach is outlined below by highlighting each aspect.

- I. In contrast to the existing solutions, in which fluency approaches were not gamified, gamification is at the core of this research. This facilitates adherence to the reading program, which is crucial for reaching automaticity.
- II. This study includes an intelligent adaptive agent that tailors the training tasks to the user's level of ability while maximizing each training session's learning value. For this, an optimization model based on the knapsack problem is proposed.
- III. The proposed approach exploits natural user interfaces, such as speech recognition, to enhance interaction quality.
- IV. In comparison to the existing solutions, a more holistic approach to fluency has been proposed by focusing on increasing the automaticity of grapheme-phoneme and phoneme-grapheme associations, as well as offering the possibility to work on the students' reading rate.
- V. Too few studies in the literature addressed vocabulary acquisition (for the first language) utilizing technology. It is known that short and limited vocabulary interventions are not effective [48], which may justify the lack of interest in vocabulary

interventions. For overcoming this problem, it is proposed to have a home-based intervention that allows for extensive vocabulary learning. Besides, other than solving the vocabulary intervention problem, a home-based intervention can address the general problem of inadequate instruction, which is an important contributor to the existence of many poor readers [40].

- VI. The current mathematical models that select the reading content for the students are rudimentary. For example, one way is to randomly present the words to the user and then omit the correct answers from the pool of the possible words, and the other method is to keep the correct answers in the pool and omit the mistakes [49], [50]. Nevertheless, these approaches are basic and do not consider the difficulty level of the words. Moreover, the value of all the words in the pool is always considered the same. However, in the proposed model of this thesis, both the value and difficulty level of the words were considered.
- VII. It is proposed to include the scientific training principles of automaticity acquisition in designing the reading tasks that target the automaticity of grapheme-phoneme and phoneme-grapheme associations.
- VIII. The game design process takes into account the motivational theories of self-efficacy and self-determination, as well as the temporal specificity of automaticity acquisition.
- IX. Estimating word difficulty is a complex task that depends on several lexical variables such as length, frequency, and consistency. However, accurate estimation of difficulty is needed for offering tailored instructional content to the learners. For achieving this objective, it is proposed to use artificial neural networks and linear regression to predict the word difficulty for the four lexical skills of auditory and visual word recognition, word decoding, and spelling. For each reading material and each reading task, the word difficulty models can calculate the weight parameter required in the knapsack model.

## 7. Conclusion

This chapter outlined the main elements of the proposed approach in the remediation of dyslexia in an opaque orthography. In comparison to transparent orthographies, the acquisition of reading in an opaque orthography requires more time. There are more grapheme-phoneme associations in opaque orthographies, which makes the acquisition process of reading more time-consuming. Since reading is an automatic process, these grapheme-phoneme associations have to become automatic for readers. To reach an automatic level in grapheme-phoneme conversion, the training design should be based on automaticity acquisition principles. However, the training models found in the literature were mostly addressing the broader aspects of skill acquisition, and the studies on the training principles of automaticity acquisition are sparse. For this reason, the next chapter was dedicated to the study of the training design for automaticity acquisition.



Another feature of opaque orthographies is that fluent reading depends considerably on the reader's vocabulary knowledge. It is because, in a multitude of cases, letter-sound decoding or even grapheme-phoneme decoding do not lead to the correct pronunciation. Therefore, vocabulary knowledge is essential in guessing the correct pronunciation of the words. However, affecting general vocabulary knowledge requires extensive training. For making extensive training possible, it was suggested to exploit the gamification concept to increase the learners' adherence. The home-based design was proposed to increase the time and access to the training program. However, home-based approaches require intelligence and adaptiveness to the learners' abilities to fill the missing role of instructor. To overcome this challenge and adapt the sessions' content to the learners' ability level, a combination of optimization modeling, threshold adaptive mechanisms, and predictive models of artificial neural network and linear regression was proposed.

## 8. References

- [1] S. Araújo and L. Fáisca, "A Meta-Analytic Review of Naming-Speed Deficits in Developmental Dyslexia," *Scientific Studies of Reading*, vol. 23, no. 5, pp. 349–368, Sep. 2019, doi: 10.1080/10888438.2019.1572758.
- [2] K. Landerl *et al.*, "Phonological Awareness and Rapid Automatized Naming as Longitudinal Predictors of Reading in Five Alphabetic Orthographies with Varying Degrees of Consistency," *Scientific Studies of Reading*, vol. 23, no. 3, pp. 220–234, May 2019, doi: 10.1080/10888438.2018.1510936.
- [3] D. Powell and L. Atkinson, "Unravelling the links between RAN, phonological awareness and reading," *Journal of Educational Psychology*, 2020.
- [4] G. K. Georgiou and R. Parrila, "What mechanism underlies the rapid automatized naming–reading relation?," *Journal of Experimental Child Psychology*, vol. 194, p. 104840, Jun. 2020, doi: 10.1016/j.jecp.2020.104840.
- [5] S. J. Joo, K. Tavabi, and J. D. Yeatman, "Automaticity in the reading circuitry," *Neuroscience*, preprint, Nov. 2019. doi: 10.1101/829937.
- [6] H. Jamshidifarsani, S. Garbaya, T. Lim, P. Blazevic, and J. M. Ritchie, "Technology-based reading intervention programs for elementary grades: An analytical review," *Computers & Education*, vol. 128, pp. 427–451, Jan. 2019, doi: 10.1016/j.compedu.2018.10.003.
- [7] P. H. K. Seymour, M. Aro, J. M. Erskine, and collaboration with COST Action A8 network, "Foundation literacy acquisition in European orthographies," *British Journal of Psychology*, vol. 94, no. 2, pp. 143–174, May 2003, doi: 10.1348/000712603321661859.
- [8] J. C. Ziegler and U. Goswami, "Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory.," *Psychological bulletin*, vol. 131, no. 1, p. 3, 2005.
- [9] W. Nagy, "Why vocabulary instruction needs to be long-term and comprehensive," in *Teaching and learning vocabulary: Bringing research to practice*, Routledge, 2005, pp. 27–44.
- [10] S. Deterding, M. Sicart, L. Nacke, K. O'Hara, and D. Dixon, "Gamification. using game-design elements in non-gaming contexts," in *CHI'11 extended abstracts on human factors in computing systems*, 2011, pp. 2425–2428.
- [11] J. Majuri, J. Koivisto, and J. Hamari, "Gamification of education and learning: A review of empirical literature," 2018.
- [12] L. Blomert and D. Froyen, "Multi-sensory learning and learning to read," *International journal of psychophysiology*, vol. 77, no. 3, pp. 195–204, 2010.
- [13] L. Blomert, "The neural signature of orthographic–phonological binding in successful and failing reading development," *Neuroimage*, vol. 57, no. 3, pp. 695–703, 2011.

- [14] H. M. Nash *et al.*, "Are the literacy difficulties that characterize developmental dyslexia associated with a failure to integrate letters and speech sounds?," *Developmental Science*, vol. 20, no. 4, p. e12423, Jul. 2017, doi: 10.1111/desc.12423.
- [15] R. I. Nicolson and A. J. Fawcett, "Automaticity: A new framework for dyslexia research?," *Cognition*, vol. 35, no. 2, pp. 159–182, May 1990, doi: 10.1016/0010-0277(90)90013-A.
- [16] R. I. Nicolson, A. J. Fawcett, R. L. Brookes, and J. Needle, "Procedural learning and dyslexia," *Dyslexia*, vol. 16, no. 3, pp. 194–212, Jul. 2010, doi: 10.1002/dys.408.
- [17] R. I. Nicolson and A. J. Fawcett, "Development of Dyslexia: The Delayed Neural Commitment Framework," *Frontiers in Behavioral Neuroscience*, vol. 13, p. 112, May 2019, doi: 10.3389/fnbeh.2019.00112.
- [18] G. D. Logan, "Skill and automaticity: Relations, implications, and future directions.," *Canadian Journal of Psychology/Revue canadienne de psychologie*, vol. 39, no. 2, p. 367, 1985.
- [19] B. Hart and T. R. Risley, "The early catastrophe: The 30 million word gap by age 3," *American educator*, vol. 27, no. 1, pp. 4–9, 2003.
- [20] R. F. West and K. E. Stanovich, "The Incidental Acquisition of Information from Reading," *Psychol Sci*, vol. 2, no. 5, pp. 325–330, Sep. 1991, doi: 10.1111/j.1467-9280.1991.tb00160.x.
- [21] K. Vidal, "A Comparison of the Effects of Reading and Listening on Incidental Vocabulary Acquisition: Incidental Vocabulary Acquisition," *Language Learning*, vol. 61, no. 1, pp. 219–258, Mar. 2011, doi: 10.1111/j.1467-9922.2010.00593.x.
- [22] T. Christ and X. C. Wang, "Bridging the vocabulary gap," *Young Children*, vol. 85, 2010.
- [23] C. R. Greenwood *et al.*, "Conceptualizing a public health prevention intervention for bridging the 30 million word gap," *Clinical Child and Family Psychology Review*, vol. 20, no. 1, pp. 3–24, 2017.
- [24] A. Biemiller and N. Slonim, "Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition.," *Journal of educational psychology*, vol. 93, no. 3, p. 498, 2001.
- [25] B. Hart and T. R. Risley, *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing, 1995.
- [26] P. L. Morgan and D. Fuchs, "Is there a bidirectional relationship between children's reading skills and reading motivation?," *Exceptional children*, vol. 73, no. 2, pp. 165–183, 2007.
- [27] T. Jagušt, I. Boticki, V. Mornar, and H.-J. So, "Gamified Digital Math Lessons for Lower Primary School Students," in *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2017, pp. 691–694.
- [28] S. Sandusky, "Gamification in Education," *ASBBS American Society of Business and Behavioral Sciences*, vol. 21, no. 1, pp. 32–39, 2014.
- [29] G. Barata, S. Gama, J. Jorge, and D. Gonçalves, "Engaging engineering students with gamification," in *2013 5th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*, 2013, pp. 1–8.
- [30] G. Barata, S. Gama, J. Jorge, and D. Gonçalves, "Improving participation and learning with gamification," in *Proceedings of the First International Conference on gameful design, research, and applications*, 2013, pp. 10–17.
- [31] G. Barata, S. Gama, M. J. Fonseca, and D. Gonçalves, "Improving student creativity with gamification and virtual worlds," in *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, 2013, pp. 95–98.
- [32] Z. Turan, Z. Avinc, K. Kara, and Y. Goktas, "Gamification and education: Achievements, cognitive loads, and views of students," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 11, no. 07, pp. 64–69, 2016.
- [33] S. Arnab *et al.*, "Mapping learning and game mechanics for serious games analysis: Mapping learning and game mechanics," *British Journal of Educational Technology*, vol. 46, no. 2, pp. 391–411, Mar. 2015, doi: 10.1111/bjet.12113.
- [34] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, "Taxonomy of educational objectives: the classification of educational goals. Handbook I: cognitive domain. New York: David McKay Company," *Inc.(7th Edition 1972)*, 1956.

- [35] L. W. Anderson *et al.*, "A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, abridged edition," *White Plains, NY: Longman*, 2001.
- [36] A. Bandura, "Self-Efficacy," in *The Corsini Encyclopedia of Psychology*, I. B. Weiner and W. E. Craighead, Eds. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2010, pp. 1–3.
- [37] A. Bandura, "Self-efficacy," in *Encyclopedia of human behavior*, V. S. Ramachaudran (Ed.), vol. 4, New York: Academic Press, 1994, pp. 71–81.
- [38] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.," *American psychologist*, vol. 55, no. 1, p. 68, 2000.
- [39] E. L. Deci and R. M. Ryan, "The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior," *Psychological inquiry*, vol. 11, no. 4, pp. 227–268, 2000.
- [40] F. R. Vellutino, J. M. Fletcher, M. J. Snowling, and D. M. Scanlon, "Specific reading disability (dyslexia): What have we learned in the past four decades?," *Journal of child psychology and psychiatry*, vol. 45, no. 1, pp. 2–40, 2004.
- [41] H. Lyytinen, J. Erskine, M. Aro, and U. Richardson, "Reading and reading disorders.," 2007.
- [42] J. Ecalte and A. Magnan, *L'apprentissage de la lecture et ses difficultés-2e éd.* Dunod, 2015.
- [43] W. Ali, "Online and Remote Learning in Higher Education Institutes: A Necessity in light of COVID-19 Pandemic," *Higher Education*, vol. 10, no. 3, 2020.
- [44] L. C. Ehri, S. R. Nunes, S. A. Stahl, and D. M. Willows, "Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis," *Review of educational research*, vol. 71, no. 3, pp. 393–447, 2001.
- [45] M. Csikszentmihalyi, *Finding flow*. Hachette Audio, 2017.
- [46] N. Zandi Atashbar and F. Rahimi, "Optimization of Educational Systems Using Knapsack Problem," *International Journal of Machine Learning and Computing*, vol. 2, no. 5, p. 552, 2012.
- [47] Z. Breznitz and M. Nevat, *The reading acceleration program (RAP)*. Israel: University of Haifa (The Edmond J. Safra Brain Research Center for the Study of Learning Disabilities), 2004.
- [48] N. R. Panel (US) *et al.*, *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development, National ..., 2000.
- [49] E. G. Steenbeek-Planting, W. H. J. van Bon, and R. Schreuder, "Improving word reading speed: individual differences interact with a training focus on successes or failures," *Read Writ*, vol. 25, no. 9, pp. 2061–2089, Oct. 2012, doi: 10.1007/s11145-011-9342-7.
- [50] E. G. Steenbeek-Planting, M. Boot, J. C. de Boer, M. Van de Ven, N. M. Swart, and D. van der Hout, "Evidence-based psycholinguistic principles to remediate reading problems applied in the playful app Letterprins: A perspective of quality of healthcare on learning to read.," in *Games for Health*, Springer, 2013, pp. 281–291.

## Contents

|  |           |
|--|-----------|
| <b>CHAPTER IV Training Model for Automaticity Acquisition.....</b>       | <b>64</b> |
| <b>1. Introduction .....</b>   | <b>64</b> |
| <b>2. Automaticity .....</b>   | <b>65</b> |
| <b>3. Key Attributes of Automaticity Acquisition .....</b>               | <b>66</b> |
| 3.1. Consistency .....   | 67        |
| 3.2. Contiguity.....   | 68        |
| 3.3. Repetition .....  | 68        |
| 3.4. Practice Format .....   | 69        |
| 3.5. Fidelity .....  | 71        |
| 3.6. Motivation .....  | 72        |
| 3.7. Time Pressure.....  | 74        |
| 3.8. Interference.....   | 75        |
| <b>4. Automaticity Stages.....</b>                                       | <b>76</b> |
| 4.1. Fast Learning.....  | 76        |
| 4.2. Slow Learning .....   | 77        |
| 4.3. Automatization .....  | 77        |
| <b>5. Proposed Model for Automaticity Acquisition Training.....</b>      | <b>77</b> |
| 5.1. Step 1 - Task Analytics .....                                       | 78        |
| 5.2. Step 2 - The Stage-Based Descriptive and Prescriptive Analysis..... | 79        |
| <b>6. Conclusion .....</b>   | <b>81</b> |
| <b>7. References.....</b>  | <b>81</b> |

## CHAPTER IV

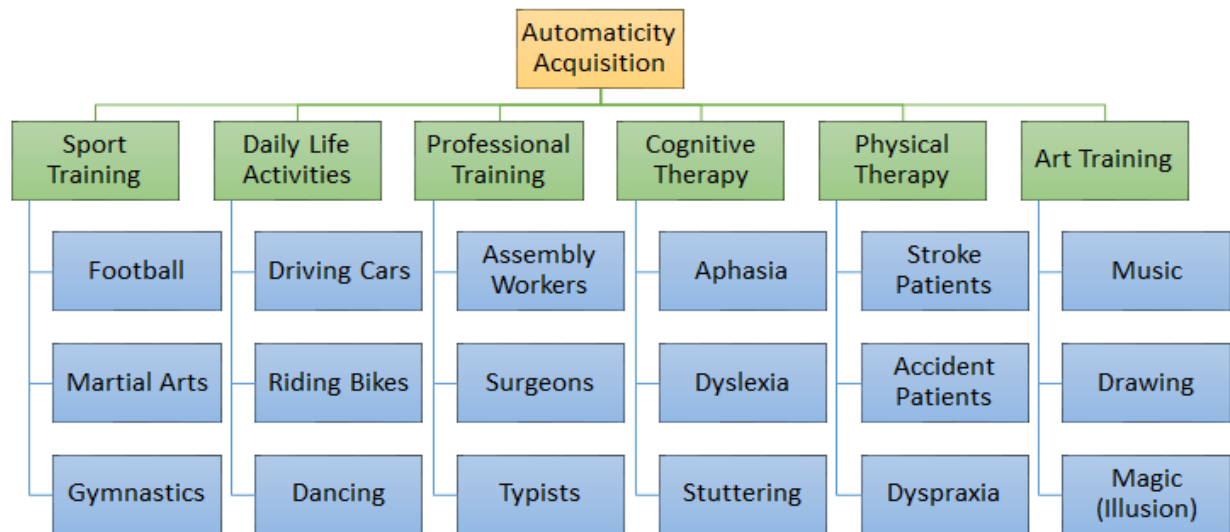
### Training Model for Automaticity Acquisition

#### 1. Introduction

Reading is a skill that contains automatic processes. Grapheme-phoneme conversion is one of the sub-processes of reading that is automatic in skilled readers. Training for automaticity (such as riding a bike) is different from learning declarative knowledge (such as learning facts). However, the studies focusing on modeling effective training approaches for automaticity acquisition have been scarce. Instead, most of the studies have targeted the broader aspects of skill acquisition. Since one of the main aims of this thesis is training the grapheme-phoneme conversion automaticity, this chapter is dedicated to identifying and modeling the principal training attributes of automaticity acquisition in general.

Automatic behaviors are omnipresent in our daily lives, from breathing, walking, and language, to complex sports activities. Performing an activity at an automatic level can have crucial benefits such as consuming less time and energy, as well as freeing up the cognitive load that can be engaged in parallel activities. This chapter provides a comprehensive account of the effective training attributes for automaticity acquisition. A total of eight key attributes and seven sub-attributes were identified, and for each, the main scientific findings were outlined. Based on this extensive review of the current literature, a two-step model was devised to guide the training designers in building effective training programs. The first step of the model, named Task Analytics, deals with task-specific design decisions. The second step provides a descriptive and prescriptive approach for each of the three phases of automaticity acquisition. The descriptive approach provides the psychological and performance-related characteristics of each phase. The prescriptive approach recommends specific training decisions for each phase of automaticity acquisition. The proposed automaticity-acquisition model can be used for designing training programs aiming at fostering automaticity in various skills. There are many potential application domains of the automaticity training model such as professional, sport, musical, and medical training. It covers a wide range of applications illustrated in Figure 1.

The related works of this research can be categorized into studies that investigate the principles of automaticity acquisition [1], [2], skill acquisition [3], [4], and general training [5]–[7]. Section 2 presents a brief introduction to automaticity. Section 3 provides an extensive literature review by identifying the key training attributes of automaticity acquisition. Section 4 describes the three phases of automaticity acquisition. Section 5 presents the two-step model of automaticity acquisition. The outcome of this research is summarized and discussed in section 6, and a conclusion highlighting new research directions is presented in Section 7.



**Figure 1.** Examples of applications of automaticity training

## 2. Automaticity

It is already an established knowledge that a vast part of our daily life behaviors is automatic, meaning that it is not mediated by our conscious choices [8]. In 1994, J. Bargh [9] defined four characteristics for automatic behavior:

- The lack of awareness of the lower mental processes of the behavior
- The lack of intention in the initiation of the behavior
- The high efficiency in performing and the minimal cognitive load
- The lack of control once the behavior is initiated

However, these characteristics were challenged, and the debate persisted on both the definition of automaticity and its features [9]–[12]. Nevertheless, this lack of consensus did not prevent the concept of automaticity from being widely used in various fields such as linguistics [13], behavioral and social psychology [14]–[17], sports [18], and learning [19], [20]. The importance of automaticity cannot be overstated. There are powerful evolutionary benefits for automatizing routine behaviors. This includes consuming less energy, performing more rapidly, and often more accurately, and freeing up the cognitive load to perform other skills in parallel [21].

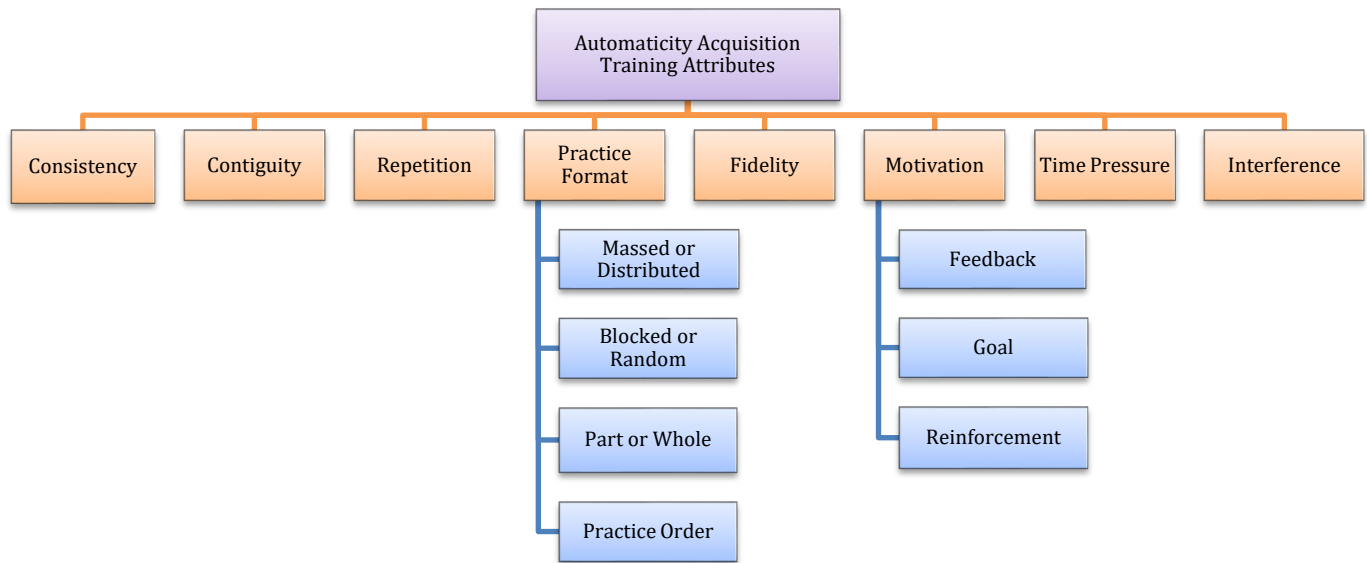
There are three types of automatic behaviors: preconscious, postconscious, and goal-dependent automaticity [8]. Preconscious automaticity initiates automatically (in the presence of environmental cues) without the awareness of the individual. Postconscious automaticity needs priming to get initiated, but still, the user is unaware of the effect of priming. Finally, goal-dependent automaticity is only initiated in the presence of the intention of achieving a goal, but the rest of the process will be automatic [8]. Most of the research on automaticity in the field of skill acquisition has focused on the third type of automaticity (goal-dependent or goal-directed) [22]. It is important to mention that automaticity and skill are not the same concepts. Automaticity is a sub-component of skill, and hence, automaticity acquisition is a

sub-process of skill acquisition. Additional clarification and detailed information on the difference between automaticity and skill were presented by Logan (1985) [11].

There are various theories of automaticity. They can be divided into process-based theories, attention-based theories, algorithm efficiency theories, and memory-based theories [23]. In addition, there are various models of automatic skill acquisition. According to these models, the acquisition of automaticity goes through different phases. Fitts and Posner proposed a model consisting of three stages of cognitive, associative, and autonomous [24]. Anderson proposed declarative, procedural, and automatic stages [25], [26]. More recently, based on neuroimaging studies, Doyon and Ungerleider proposed five stages of fast learning, slow learning, consolidation, automatization, and retention [27]–[29]. However, as described in their studies, the consolidation and retention stages are not active training phases.

### **3. Key Attributes of Automaticity Acquisition**

This section aims at identifying the key training attributes that affect the process of automaticity acquisition. The research on automaticity acquisition is rather sparse, and most of the literature has targeted broader aspects of skill acquisition. There have been some attempts to identify key attributes in skill acquisition [3]. However, the objective of this chapter is to extend the established knowledge about this issue and provide a more comprehensive overview of the most important attributes affecting the process of automaticity acquisition. Eight main attributes, as well as seven sub-attributes, were identified in this section. Each of them is briefly introduced, and their most important scientific findings are outlined. It should be noticed that there is an overlap between the attributes that affect automaticity and skill acquisition in general. This is because automaticity is a sub-process of skill. For further information, refer to Logan (1985) [11]. The eight identified attributes (Figure 2) were selected to cover the three main principles of training: acquisition, retention, and transfer [6].



**Figure 2.** Key training attributes and sub-attributes of automaticity acquisition

### 3.1. Consistency

For training automaticity, the first step is to determine which tasks can be automatized. According to many studies, the consistency of mapping between stimuli and response is the key element [10], [30]–[33]. These studies divided practices into two forms: Consistent Mapping (CM) and Varied Mapping (VM). In CM, a particular stimulus always receives the same response. However, in VM, the response to a stimulus varies during the practice. The manifestation of consistency concept can vary in different tasks. It is explored in-depth in the visual search paradigm (also known as feature extraction task). In this task, the person will be presented with a set of target(s) and distractors to be discerned. Researchers have indicated that when the targets and distractors are always different during the training, and there is no overlap between them, the chance of developing automaticity is higher in comparison to the case of overlapping targets and distractors (having previous targets as distractors or previous distractors as targets) [1], [10], [30].

Some researchers considered CM as a necessary precondition for reaching automaticity [34]. However, others challenged the necessity of lower-order CM in automatization [35], [36]. Furthermore, it has been argued that not all types of CM would lead to automatization [37]. In a more complex task, a set of stimuli can exist. For complex cases, consistency can be defined for multiple components or at multiple levels. However, the impact of consistency is not the same for every component or at each level [36], [37]. Therefore, it is important to identify the component(s) or level(s) for which the consistency should be respected. There are many examples of studies that incorporated consistency at higher levels than mere individual stimulus-responses [33], [38]–[41]. Hence, maybe the definition presented by Rogers et al. depicts a better picture for consistency. They defined consistency as "The learning and application of invariant rules, invariant components of processing, or invariant sequences of



information processing components that are essential to successful task performance" [1], [33], [40]. It is worth noting that the contextual elements of a training program that are not stable in the real task should be varied during training to decontextualize the skill. However, in the absence of consistency in lower/higher-order components of the task, it has been shown that creating temporary consistencies by using consistent contextual cues can have positive effects [1], [42], [43]. For example, an inconsistent set of learning material can be broken down into consistent subsets that can be trained separately.

### **3.2. Contiguity**

Consistency is paramount for the creation of automatic associations between stimulus-responses. However, consistency is not the only factor in forming automatic associations. Proximity or contiguity of the stimulus and response in terms of space and time is important and substantially affects learning [44]. According to the law of contiguity, two items form association when they are temporally and spatially contiguous to each other [45]. However, in some tasks, the stimuli-responses, rules, sequences, or any other regularities are distanced (in time or space) far apart. In these cases, the trainees cannot form the associations, or if the associations are formed, the task does not reach automatic levels. Hence, it may be better to compress the stimulus and response of the consistent components in the time and space to make them more salient to the trainee.

Compressing the stimulus-response of a task can be useful in both cases of implicit and explicit learning. In the case of implicit learning, the person implicitly learns the regularities during the execution of a task. However, this is only possible if the regularities are detectable. When the stimulus and response are largely disconnected from each other, they become less detectable and could not be learned implicitly (for example, see [46], [47]). Compressing the distance between the stimulus and response can make them more salient for detection.

Similar to the case of implicit learning, the contiguity of stimulus and response plays a key role in automatizing a task requiring explicit learning. The positive effect of contiguity on the strength of memory associations has been known for a long time [48], and according to memory-based theories of automaticity, the strength of memory traces is key for reaching automaticity [49].

### **3.3. Repetition**

The important role of repetition for reaching automaticity is evident. In addition to consistency and contiguity, stimulus-responses should happen frequently enough to promote automaticity. In the case of a task with infrequent consistent components, one solution is to practice it in isolation for increasing the number of trials. Another solution is to alter the task by reducing the time intervals between each incidence. This might reduce the training's fidelity, but it is a beneficial technique leading to automaticity [50].

There have been some attempts to quantify the process of automaticity acquisition as a function of repetitions counts. Thorndike noted that skill acquisition follows a non-linear pattern, fast initially, but slowing down by practice [51]. The power law of practice was first proposed by Snoddy [52] and later highlighted by Fitts and Posner [24], Newell and

Rosenbloom [53], and Anderson [25]. This law states that the decrease in response time (a performance measure) by repetitions is a decelerating process best depicted by a power function. Logan [34] explained the reasons behind the power law of practice through the instance theory of automaticity that he developed. However, Heathcote et al. argued that the power function better fits the averaged data of subjects and/or tasks, and an exponential function better matches the individual learning process [54].

Furthermore, as the theory of deliberate practice [55], [56] suggests, repetitions should focus on improving specific weaknesses. This requires the trainees to go out of their comfort zones. Working on weaknesses also requires conscious and focused monitoring of the performance and avoiding mindless repetitions. Having a qualified coach or instructor can help spot the weaknesses and teach effective strategies to overcome them. The theory of deliberate practice also suggests an environmental and monotonic relationship between performance and deliberate practice. However, some studies have challenged these views and suggested that deliberate practice is not the sole factor in explaining the level of expert performance [57]–[59]. Nevertheless, a carefully designed evidence-based practice can mitigate the asymptotic effect of practice, regardless of its name (e.g., deliberate, optimal, purposeful).

### **3.4. Practice Format**

The optimal format and structure of the training depend on the type of skill. Many decisions should be made to organize the training plan. Four of these practice formats are outlined in the following subsections.

#### **3.4.1. Massed or Distributed Practice**

When the practice sessions or repetitions are in immediate successions (zero interval), the practice is called "massed". By contrast, when the sessions or repetitions are separated by time (greater than zero intervals), the practice is called "distributed" [60]. The research in the effects of distributed and massed practice has a history of well over a century [61], and nowadays, it is well established that distributed practice will lead to better retention [62]–[64]. This effect is called the "distributed-practice effect". Despite the long history of research in this domain, still, there are important open questions. The first question is the optimal duration of intervals between the practice sessions for the best retention. A comprehensive study carried out by Cepeda et al. [65] suggested that the optimal interval depends on how long the person needs the information to be retained. When the retention test is in the near future, the optimal interval is short. However, when the information should be retained for a long period, the longer intervals are more effective. The second open debate is on the fixed versus variable intervals (expanding intervals). The research on this problem has yielded ambiguous results [66], [67].

When it comes to automaticity, overlearning gains more importance. The value of overlearning has been questioned by Rohrer et al. [64], suggesting that more distributed adequate learning may be a better use of time in improving retention in comparison to overlearning. However, studies such as Servant et al. [68] suggested that the process necessary for automatization (switch from working memory to long term memory) is slower in spaced practice than massed practice.

Furthermore, in a review study published in 2015, Rohrer suggested that instructions should be distributed over long periods [69]. However, in the same article, the data from foreign language learning did not follow the same conclusion. Rohrer pointed to the issues in the design of the reviewed studies as a possible explanation. However, language is a skill that has automatic processes, and it is different from other memorization tasks reviewed in Rohrer's article. The result of Rohrer's study can be reformulated as: when the training is all about memorizing declarative knowledge, a distributed practice is more effective, and when the training requires automaticity acquisition (procedural knowledge), a massed practice is more effective.

When automatization is the goal, a massed practice should be used until the task is fully automatized. However, when the task is already automatic, and the goal is only to retain the skill, the distributed practice format can be used to maximize retention. Furthermore, it is important to note that retention is not the only key factor in deciding the massed versus distributed practice. Factors such as the user's motivation, the complexity, and the energy cost of the task are important to consider. A user with high motivation facing a simple task that requires low energy can withstand massed practice, but for the reverse situation, a distributed practice is more suitable [70].

### **3.4.2. Blocked or Random**

Blocked practice refers to the condition where the training trials remain relatively constant (low-variability). It contrasts with the Random practice in which the training trials can vary significantly (high-variability) [71]. In blocked practice, the cognitive disruption is low because the focus is on performing a single task. However, cognitive disruption will go higher in random practice because the attention should be switched frequently to focus on different types of tasks [72]. According to the Contextual Interference hypothesis [73], blocked practice (low contextual interference) will lead to better acquisition, while random practice (high contextual interference) leads to better retention. This implies that for automaticity acquisition, it is better to start with blocked practice, and in later stages, a random practice can be used to retain the skill. Furthermore, it has been shown that random practice provides a higher transfer effect [74]–[76].

The contextual interference effect can happen in two ways. The first way is by reconstructing the mental representation of the task each time during the random practice, and consequently, by strengthening the mental representation (Reconstruction Theory [77], [78]). The second way is by acquiring the comparative information available during random practice, and consequently, storing a more elaborated version of the task (Elaboration Theory [79]). Research supports the claim that random practice will lead to better retention at a delayed test; however, it appears that results are more robust in laboratory conditions than in real-world situations [80]. In addition, it has been suggested that the contextual interference effect is more valid for adults than children [80]–[83], and for children, blocked practice is more effective. Finally, retention is not the only deciding factor. If the users are low in self-efficacy, it may be more appropriate to let them first build their sense of self-efficacy by achieving mastery in the blocked practice format.

### 3.4.3. Part or Whole

Another decision in practice format is whether to train the task in parts or as a whole. According to Naylor and Briggs [84], the optimal decision depends on the task's levels of complexity and organization. The complexity of a task can be determined by the required cognitive load and the number of its subparts. An example of a task with high complexity is dancing. Organization is the extent to which the parts of a task rely on each other. When a task can be easily decomposed into discrete parts, it has low organization. An example of a task with a low organization is soccer, in which the parts (passing, shooting, dribbling) can be practiced separately. This contrasts with the golf swing, which has a high organization and cannot be easily broken down into subparts, and hence, it should be practiced as a whole. It has been suggested that the tasks with high complexity and low organization could be practiced as parts, and the tasks with low complexity and high organization should be practiced as a whole [84]–[86].

Another notable view belongs to Schmidt and Wrisberg [87]. They classify tasks into the three categories of discrete, serial, and continuous. Discrete refers to the tasks that are short and have a clear beginning and ending (e.g., throwing a dart). Serial refers to the tasks comprised of multiple discrete tasks, which should be performed in a certain sequence (e.g., brushing teeth). Continuous refers to the tasks that do not have a clear beginning and ending (e.g., running). Schmidt argues that the whole practice is more suited to discrete tasks, as well as serial tasks with high organization, and the part practice is more appropriate for serial tasks with low organization.

### 3.4.4. Practice Order

The order of the practice can have a considerable effect on training results [88]. There are three types of practice order: blocked order, serial order, and random order [70]. Blocked order is when the same training trial is presented multiple times, and once it is mastered, the next trial is introduced (e.g., AAABBBCCC). Serial order is when the training trials are presented in a sequence (e.g., ABCABCABC). Random order refers to the situation where trials can be presented in a random and unpredictable order (e.g., BACCBAACB). Similar to the discussion about Blocked or Random practice, the blocked order is superior in increasing the early performance. However, the serial and random order is superior in long term retention and slower in early acquisition [89]. This is due to the contextual interference effect [73]. Finally, starting with a blocked practice and then systematically increasing the contextual interference may be an effective way for both performance and retention [90].

## 3.5. Fidelity

Automaticity is task-specific, which means that deviations from the task context or the task itself can hinder the skill transfer [68]. Thorndike's law of identical elements [91] suggests that the more training elements are similar to the actual task elements, the higher the transfer probability. However, some researchers regard this view as too simplistic [92], [93]. They suggested that the relationship between the level of fidelity and the transfer is not linear, and higher levels of fidelity might be beneficial at a diminished rate [94]. In addition, it is argued

that a lower level of fidelity might suit novices better (for avoiding cognitive load), and higher levels may be more appropriate for advanced learners [95]–[97].

Furthermore, different aspects of fidelity do not have the same added value. This fact led researchers to propose a selective fidelity approach [92], in which the designer should identify the most pertinent aspects of fidelity to the training program. Some of these aspects are physical fidelity, visual-audio fidelity, psychological-cognitive fidelity, and functional fidelity [98]. It is important to note that there is a difference between the actual level of fidelity and the perceived level of fidelity. It has been observed that sometimes as the level of fidelity goes up, the perceived level of fidelity goes down; this phenomenon is called "Uncanny Valley" [99], [100].

There is a tradeoff between the rapid acquisition of the task and the transfer of the task to real-life situations [101]. In terms of design for automaticity, higher levels of fidelity are expected [97]. In addition, contextual cues are important in acquiring automatic skills [102]–[104]. Moreover, it has been shown that in the absence of general stimulus-response consistency, creating temporary consistencies by using contextual cues can be beneficial in the development of automaticity [1], [42], [43]. Therefore, in training design, it is important to identify the necessary stimuli-responses and important contextual cues. Then, the level of fidelity should be determined by the experience level of the user. The level and aspects of fidelity can be designed to dynamically change as the user progresses [105]. By following these principles, rather than merely attempting to create a photorealistic training program, there will be room for adding playful graphics and other ludic elements.

### **3.6. Motivation**

In order to acquire automaticity, a high amount of repetitions is required. When the task is no longer perceived as a novel activity, the practice can become boring. This increases the risk of dropout. In the following subsections, three main external factors that affect motivation (feedback, goal setting, and reinforcement) are discussed.

#### **3.6.1. Feedback**

Feedback helps automaticity acquisition by providing corrective information. However, it can also help by enhancing motivation. Generally, feedbacks can be categorized as intrinsic and extrinsic. The intrinsic feedback refers to the sensory-perceptual information that is available to the body while performing a task. An example is the proprioceptive feedback of the body. It gives information about the self-movement and the position of the body. The extrinsic (augmented) feedback refers to a non-intrinsic type of feedback originating from an external source such as a technology or a trainer [106].

Extrinsic feedback can be divided into two categories of concurrent and terminal feedback. Concurrent feedback refers to the feedback that the person receives during the task. Concurrent feedback leads to the Knowledge of Performance (KP), which provides information on the quality of the actions taken to complete the task. On the other hand, terminal feedback is the overall information that the person receives when a task is completed, which leads to the Knowledge of Results (KR) [107]. Each type of feedback has its advantages and drawbacks in

a given specific condition [108]–[110], as well as having an interplay between them [111]. In general, concurrent feedback can better suit novices or complex tasks, while terminal feedback is more appropriate for non-novices or simple tasks [107], [112].

Another aspect of a feedback system is its intensity or the frequency of providing feedback. When the task is complex, more frequent feedback can be helpful, particularly at the beginning [113], [114]. However, at the later stages of acquisition, a high frequency of feedback can slow the retention of skills [115]–[117] and impede automaticity [118]. Therefore, at the start, full feedback or thin bandwidth KR feedback (high frequency) can be provided to correct the errors at the initial stages of skill acquisition. Then, wide Bandwidth-KR feedback (low frequency) can be provided at the automatization phase [117]. Concerning automaticity, it is important to choose the feedbacks that make the consistent components of a task more salient to the user [1], [119].

### **3.6.2. Goal**

Based on the premise set by Ryan [120] that conscious goals affect actions, E. Locke and G. Latham developed the Goal Setting Theory [121], [122]. The original model was proposed by Locke in 1968 [123], and since then, it has been the subject of continuous improvement. The first principle of the goal-setting theory suggests that the difficulty level of a goal has a direct relationship with the level of performance. This, however, assumes that the person has the sufficient ability to reach the goal, otherwise the performance may drop [124]. The second basic principle of the goal-setting theory suggests that a specific goal leads to greater achievement than a general goal such as "do your best". In addition, the specificity of a goal reduces the variability of the performance [125].

Apart from these two principles of specificity and difficulty, a number of moderators are involved in setting an optimal goal. Goal commitment is one of the most important moderators. Goal commitment is affected by attributes such as expectancy (e.g., outcome expectancy, self-efficacy) and goal attractiveness (valence, instrumentality) [126]. In other terms, goal commitment is high when achieving the goal is perceived as possible and important [127]. Allowing people to participate in goal setting may increase the perceived importance of a goal. However, assigned goals provided with a legitimate rationale were equally effective in raising the performance [122], [128].

Another important moderating attribute in the goal-setting theory is feedback. It has been suggested that feedbacks are more effective when they can be used to set goals [121]. Similarly, goals are more effective when the progress towards them can be measured by relevant feedbacks [129]. Hence, the combination of goals and feedbacks is more effective than any of them alone [130]–[132].

Another important moderator of goal setting theory is task complexity. It has been shown that when task complexity is high, the relation between goal-setting (difficulty and specificity) and performance is weaker [133]. Setting specific goals can be detrimental at the starting phase of automaticity acquisition [134]. In addition, some studies suggested that for complex and novel tasks, setting learning (process) goals are more effective than performance (outcome) goals [135], [136]. Bandura and Schunk [137] suggested that setting proximal goals is more

effective than distal goals or no goals when self-efficacy is low for a task. Furthermore, for complex tasks, distal learning goals, coupled with proximal learning goals, appear to be more effective than only distal goals or "do your best" goals [136], [138]. Finally, closer deadlines result in higher performance up to a certain point, and after that, they reduce the performance, and this reduction is more pronounced for complex tasks [139]–[141].

### **3.6.3. Reinforcement**

According to B. F. Skinner [142], [143], reinforcements are external stimuli intended to coerce individuals to specific responses. By applying rewards and punishments, the behavior of the individuals is modeled towards the desired action. Reinforcements can be scheduled in two main types, non-intermittent and intermittent [144]. Non-intermittent schedules include the two categories of continuous reinforcements and extinctions. In continuous reinforcements, all responses are reinforced, and in extinction, none of the responses is reinforced.

Intermittent schedules include fixed-ratio, variable-ratio, fixed-interval, and variable-interval. In fixed-ratio, reinforcements are given after a fixed amount of responses. In variable-ratio, reinforcements are given after a random amount of responses. In fixed-interval, reinforcements are given on the first response after certain intervals of time. Finally, in variable-interval, reinforcements are given on the first response after random intervals of time. Out of all the presented schedules, the variable-ratio schedule has shown to be the most effective in encouraging individuals to follow the desired behavior [143], [145]–[148].

Reinforcements can effectively maintain the desired behavior. However, when the reinforcements are no longer provided, and the intrinsic motivation is not present, individuals may stop the behavior [149]. Even worse, external incentives may change the intrinsic motivational locus to an extrinsic motivational locus [150]. For example, when money is provided for engaging in an activity, intrinsic motivation can drop. On the other hand, verbal reinforcement and positive feedback can increase intrinsic motivation [151], [152]. Nevertheless, the psychological meaning of a given reward changes from individual to individual based on their intrinsic needs, values, and goals [149], [153].

Concerning automaticity acquisition, it is better to start with continuous reinforcement to create the necessary associations. Then, the reinforcements should gradually become intermittent to be able to sustain the behavior [154]. Csikszentmihalyi recommends external reinforcements for the starting phase for increasing the commitment to the tasks, but at the same time, intrinsic motivation should be cultivated [155]. Once intrinsic motivation is achieved, the reliance on external reinforcements can be reduced [149].

### **3.7. Time Pressure**

The Instance theory of automaticity states that at any stage of acquisition, there is a race between the algorithm-based performance and the memory-based performance [34]. The algorithm-based performance is performed in working memory, and depending on the nature of the task, it can come in multiple steps [49], [68]. However, the memory-based performance comes as a single-step direct-access retrieval from long-term memory [34], [49]. During the earlier phases of automaticity acquisition, the algorithm-based performance wins the race

because the memory traces are still weak and need further strengthening. However, towards the advanced stages of skill acquisition, a transition from multi-step algorithm-based performance to single-step direct-access memory-based performance occurs. Once this happens, the task is known to be automatic [34].

According to the Instance theory, automatization is this transition from algorithm-based to memory-based performance. Therefore, if one can accelerate this process, automaticity will be reached faster. Making the automatization process faster has important benefits, such as reducing the training costs. Time Pressure technique can push the person towards the transition from algorithm-based (controlled processing) to memory-based (automatic processing) performance.

Furthermore, automaticity is a continuum [12], [156]. Therefore, the performance can get faster and more precise by further practice. In addition, it becomes more resistant to the effects of alcohol, fatigue, stress, and vigilance decrement [157]–[160]. During the advanced stages, Time Pressure can make memory retrieval faster, more precise, and more resistant to the mentioned stressors. It has been shown that simulations that use Above Real-Time Training (ARTT) can better emulate the tasks that have a demanding nature in reality, such as military training or high-performance sports [161], [162]. ARTT leads to a better transfer and retention while reducing the training time [163], [164].

However, not all the skills require Time Pressure to become automatic. Most of the automatic skills acquired during life (e.g., brushing the teeth) are not learned under time pressure. However, using Time Pressure can accelerate the automatization process. It can also push the person towards performance levels that normally are not reached without pressure. Nevertheless, time pressure should only be introduced after the initial learning is accomplished.

### 3.8. Interference

In the context of learning, the term interference has been used to describe different phenomena, which concern both acquisition and retention processes. In memory research, the influence of interference has been known for a long time [165]–[167], and it is also a prominent theory explaining the phenomenon of forgetting [168]. According to the interference theory of forgetting, memories weaken because the new memories interfere with them and impede their retrieval [169].

Two of the important interferences are retroactive and proactive interferences. Retroactive interference refers to the case where new memories interfere or hinder the retrieval of older memories. In contrast, proactive interference refers to the problem in forming new memories caused by old memories. The similarity of learning materials and learning environments are important in the degree of interference [168], [170]. In addition, studies suggested that stimulus similarity, when coupled with response variation, causes the most interference [171].

Another type of interference happens when the tasks are learned in close succession. Tasks' similarity is not a defining factor for this type of interference, known as generalized or non-specific interference [172]. However, the temporal proximity of the two learning tasks is key. The studies have shown that this type of interference happens only when the interval between the tasks are short. The growth of the interval time reduces the interference effect [173]–[176].



This interference also comes in two types: retrograde and anterograde interference. When tasks of A and B are learned in close succession, retrograde interference refers to the negative effect that learning of task B has on retrieving the memory of task A. Anterograde interference refers to the diminished capacity of learning task A after learning task B [177].

At the starting phase of automaticity acquisition, the task is novel, and the retrograde and anterograde interferences are high. At the advanced phase of acquisition, the task has been mastered, and these interferences are low. Conversely, at the beginning of automaticity acquisition, proactive and retroactive interference are low due to the weak memory traces. However, in the advanced phases, these two types of interference are high.

Therefore, at the initial phase of automaticity acquisition, either inter-session breaks should be longer for stabilizing the learned skill, or the length of the session should be extended until overlearning is reached. However, at the advanced phase of automaticity acquisition, the inter-session breaks can be shorter.

Similar to inter-session breaks, intra-session breaks (interruptions) can have an interfering effect on automaticity acquisition. The disruptive effects of interruptions on a skill will diminish as the person progresses towards automaticity. Therefore, in the automaticity acquisition, the interruptions should be avoided at earlier stages when the task is complex, and the working memory load is high. However, when the task becomes automatized and the working memory load is low, interruptions become less detrimental. Even some studies argue that interruptions at the automatic stage can be beneficial [3]. This beneficial effect could be due to the contextual interference effect.

## 4. Automaticity Stages

As mentioned previously, a few models of automaticity exist in the literature. In general, the process of automaticity acquisition is divided into multiple phases. The model proposed by Fitts and Posner goes through the three stages of cognitive, associative, and autonomous [24]. Anderson proposed declarative, procedural, and automatic stages [25], [26]. In this chapter, the model developed by Doyon and Ungerleider [27]–[29] is considered as the basis for creating the two-step model of automaticity acquisition. Based on the neuroimaging studies, Doyon and Ungerleider specified that automaticity acquisition evolves through the five phases of fast learning, slow learning, consolidation, automatization, and retention. However, consolidation and retention, as described in their studies, are not active training phases. Therefore, the three phases of **fast learning**, **slow learning**, and **automatization** were used for developing the training model. Each of these three phases is described by the identification of their characteristics in the following sections.

### 4.1. Fast Learning

Fast learning refers to the initial phase, during which the person understands the overall plan required to perform the skill. Instructions are needed, and the person may benefit from cues and hints. At this stage, the users create a so-called cognitive map in the form of declarative knowledge. It requires high cognitive processing to perform the skill. Hence, the cognitive load

is high at this stage, and any optimal training design should avoid unnecessary cognitive load. Improvement at this stage is fast (as specified by the power law of practice). However, the user commits more errors than the later stages, and corrective feedback is essential. At this phase, the risk of dropout is at its maximum due to the low self-efficacy. Hence, the reliance on extrinsic motivation is high.

#### **4.2. Slow Learning**

After the initial fast learning stage, when the user learns basic strategies to perform the skill, improvements slow down, and further gains become less often. Nevertheless, the improvement continues with practice, although rather slowly. In addition, supported by either the power-law or exponential law of practice, the rate by which further improvements slowdown is not linear. It is characterized by a gradually decelerating process that ultimately reaches an asymptote. Consequently, when the gains become insignificant, motivation becomes key in pushing users towards further practice. As a result, the risk of dropout is still high. Therefore, both intrinsic and extrinsic motivational elements are important. Particularly, the sense of progress should be made more salient, even though, at times, it will be only the illusion of progress.

Furthermore, at this stage, users need fewer instructions, hints, and cues. In addition, their declarative knowledge gets compiled and starts to become proceduralized [26]. However, the performance is still effortful, the cognitive load is still high, and the user is prompt to errors, although less than the fast learning phase.

#### **4.3. Automatization**

At this stage, the procedural knowledge of the skill is fully composed and integrated into a one-step direct-access memory [26], [34]. Hence, the trainee does not need to hold the algorithm-based model of the skill in his/her working memory. This frees up the cognitive load. Performance is mostly implicit and with minimal conscious monitoring. In addition, the performance is fast and effortless, and the user is less prone to errors. However, as stated previously, automaticity is a continuum, and its features (e.g., speed, cognitive load, errors) can get better by practice [156]. At this stage, the terminal or even summary feedback is more suitable than the concurrent feedback. Moreover, the trainee should be challenged frequently; otherwise, the task becomes too easy and boring. As explained previously, the time pressure technique can be beneficial in enhancing the level of automaticity.

### **5. Proposed Model for Automaticity Acquisition Training**

Based on the thorough review and analysis of the literature presented in the previous sections, a Two-Step Automaticity Training Model was designed to encapsulate its most relevant outcomes. This model offers a better understanding of automaticity acquisition and enables training designers to build an optimal training program. The first step of the model is Task Analytics, which refers to the specific decisions that training designers should make for each task. The second step of the model provides the distinct characteristics of the three phases of automaticity acquisition. These characteristics are organized in the two parts of Table 1. The

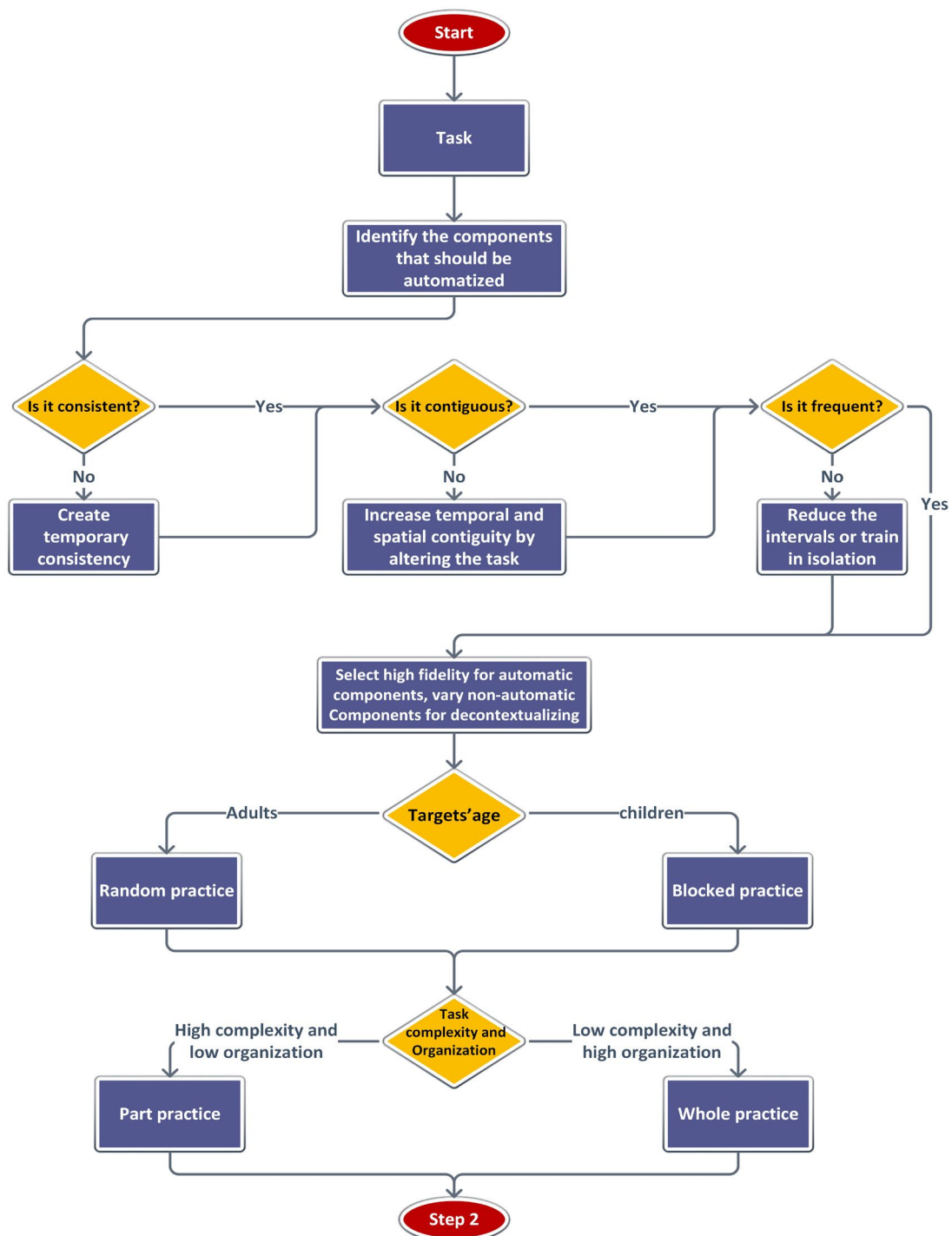
left side of the table is a descriptive outlook. It unfolds the psychological or performance related changes that happen during the phases of automaticity acquisition. The right side of the table is a prescriptive approach. It recommends various training design choices for each of the three phases of automaticity acquisition.

### **5.1. Step 1 - Task Analytics**

The objective of the Task Analytics step is to help training designers make task-specific training choices. This step of the model is illustrated in Figure 3. The first step is to identify the specific parts of the task that have to be automatized. The second step is to examine the stimulus-response consistency of the targeted components. If they are consistent, the task can be automatized. However, if they are not consistent, it is better to create temporary consistencies. An example is to classify the subparts into consistent categories and to train each category separately.

The third step is to determine whether the stimulus and response are sufficiently proximal or contiguous to each other. This aspect has particular importance in implicit learning where the individual has to learn consistent rules and regularities, such as sequences or patterns. This can be achieved by altering the task and bringing stimulus and response close to each other. The next step examines whether the stimulus-responses occur infrequently compared to the task's duration. The solution to this problem is either to reduce the intervals between their occurrences or to practice them in isolation. The next task-specific design choice is to decide which components should be trained under high fidelity and which should be varied to decontextualize the skill.

Moreover, as mentioned previously, blocked practice seems more appropriate for children and random practice better suits adults. Finally, as discussed in Section 3.4, if the task was high in complexity and low in organization, it is better to practice it in parts. However, when the task is low in complexity and high in organization, it should be practiced as a whole.



**Figure 3.** The diagram of Task Analytics

## 5.2. Step 2 - The Stage-Based Descriptive and Prescriptive Analysis

This step is divided into two parts representing a descriptive outlook and a prescriptive approach to automaticity acquisition. These two parts are organized into two sides, presented in Table 1. The left side of the table is a descriptive outlook. It lists psychological and performance-related characteristics for each stage of automaticity acquisition. Most of these characteristics are provided in the form of dichotomies along with a bidirectional arrow,

indicating the continuum that exists between the fast learning and automatization stages. This table helps to better understand the detailed characteristics of each phase of automaticity training. The right side of the table takes on a prescriptive approach. It is developed based on the inferences derived from the characteristics outlined in the descriptive table (on the left) and the comprehensive review of the literature presented in Section 3. The prescriptive table (on the right) recommends effective training elements for each stage of automaticity acquisition. Similar to the descriptive table, the prescriptive table has employed dichotomies to represent the continuum that exists within the stages of automaticity acquisition. The prescriptive table can be used in the design of effective automaticity training programs. It is worth mentioning that it is assumed that the targeted task for training is complex in this model.

**Table 1.** The stage-based descriptive (left) and prescriptive (right) analysis of automaticity acquisition

| Fast Learning                                | Slow Learning         | Automatization                              | Fast Learning                         | Slow Learning                       | Automatization                        |
|--|-----------------------|---|---------------------------------------|-------------------------------------|---------------------------------------|
| Declarative Knowledge                        | Knowledge Compilation | Procedural Knowledge                        | Low Fidelity                          | ←→                                  | High Fidelity                         |
| Controlled                                   | ←→                    | Automatic                                   | More Explicit Instruction             | ←→                                  | Less Explicit Instruction             |
| Serial Processing                            | ←→                    | Parallel Processing                         | Massed                                | ←→                                  | Distributed                           |
| High Cognitive Load                          | ←→                    | Low Cognitive Load                          | Blocked Practice                      | ←→                                  | Random Practice                       |
| Complex, Effortful, Slow                     | ←→                    | Simple, Effortless, Fast                    | Part                                  | ←→                                  | Whole                                 |
| High Error Rate                              | ←→                    | Low Error Rate                              | Blocked Order                         | Serial Order                        | Random Order                          |
| High Improvement Rate                        | ←→                    | Low Improvement Rate                        | Concurrent Feedback                   | Terminal Feedback                   | Summary Feedback                      |
| Low Self-Efficacy                            | ←→                    | High Self-Efficacy                          | High Frequency of Feedback            | ←→                                  | Low Frequency of Feedback             |
| High Retrograde and Anterograde Interference | ←→                    | Low Retrograde and Anterograde Interference | Proximal Goals                        | Proximal & Distal Goals             | Distal Goals                          |
| Low Retroactive and Proactive Interference   | ←→                    | High Retroactive and Proactive Interference | Learning Goals                        | Learning & Outcome Goals            | Outcome Goals                         |
| High Disruptive Effect of Interruptions      | ←→                    | Low Disruptive Effect of Interruptions      | No Time Pressure                      | ←→                                  | Time Pressure                         |
| High Risk of Dropout                         | ←→                    | Low Risk of Dropout                         | Long Inter-Session Breaks             | ←→                                  | Short Inter-Session Breaks            |
| High Reliance on Extrinsic Motivation        | ←→                    | Low Reliance on Extrinsic Motivation        | Low-Frequency of Intra-Session Breaks | ←→                                  | High-Frequency Intra-Session Breaks   |
|  |                       |   | Continuous Reinforcement              | Frequent Intermittent Reinforcement | Infrequent Intermittent Reinforcement |

## 6. Conclusion

Automaticity is a crucial part of our everyday lives. A small problem in the acquisition of automatic skills such as language can result in debilitating consequences. This chapter synthesized the current knowledge relating to the most important attributes required for effective and optimal training of automatic skills. The first part of the chapter provided an in-depth review of these attributes and their relationship with automaticity acquisition. Based on this review, a two-step model for the design of automaticity training programs was proposed. The first step, named Task Analytics, helps with the task-specific decisions that impact training design. The second step proposes descriptive and prescriptive models to the three phases of automaticity acquisition (fast learning, slow learning, and automatization). The descriptive model describes the psychological and performance-related characteristics of each phase of automaticity acquisition. The prescriptive model recommends the best choices of training design for each phase.

This model can be fully applied in creating training programs for reading acquisition and the remediation of dyslexia. Consistency is the most important parameter for automaticity acquisition. Therefore, in the design of reading programs, the consistency of grapheme-phoneme associations should be reinforced. Grouping the words with the same grapheme-phoneme association is key to promoting consistency. Word lists, with each list containing the words with the same grapheme-phoneme, can be used as the training material. If the instruction is based on texts, each text can be composed by frequently using the target grapheme-phoneme association. The consistent grapheme-phonemes should be practiced in a block at the initial phases until the conversion speed reaches an automatic level. Random practice can be adopted when this stage is reached. Then the time pressure mechanism can be applied to speed up the automaticity acquisition process. Furthermore, it was discussed that the contextual elements of a training program that are not stable in the real task should be varied during training to decontextualize the skill. However, the educational tools that display each grapheme with consistent colors contextualize the grapheme-phoneme conversion skill, and their effectiveness should be investigated.

Other than dyslexia, the application possibilities of the proposed model are countless. It can be used to train professional skills, sports, music, and education (e.g., mathematics), to name a few. A beneficial way to use this model is to design rehabilitation for physical or cognitive problems that result from a deficit in an automatic skill. For example, it can be used for designing the rehabilitation program of language-related problems such as aphasia. Furthermore, as currently there is no gamification or serious game design model for automaticity acquisition, the next chapter proposes and evaluates a stage-based gamification model based on the second step of the training model proposed in this chapter.

## 7. References

- [1] W. A. Rogers, T. J. Maurer, E. Salas, and A. D. Fisk, "Training design, cognitive theory, and automaticity: Principles and a methodology," in *Improving training effectiveness in work organizations*, 1997, pp. 19–45.

- [2] B. J. Holt and S. J. Rainey, "An overview of automaticity and implications for training the thinking process," ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES ALEXANDRIA VA, 2002.
- [3] J. Langan-Fox, K. Armstrong, N. Balvin, and J. Anglim, "Process in skill acquisition: Motivation, interruptions, memory, affective states, and metacognition," *Aust. Psychol.*, vol. 37, no. 2, pp. 104–117, 2002, doi: 10.1080/00050060210001706746.
- [4] A. M. Williams and N. J. Hodges, *Skill Acquisition in Sport: Research, Theory and Practice*. Routledge, 2004.
- [5] A. F. Healy, V. I. Schneider, and L. E. Bourne, "Empirically valid principles of training," in *Training cognition: Optimizing efficiency, durability, and generalizability*, Psychology Press, 2012, pp. 13–39.
- [6] A. F. Healy, J. A. Kole, and L. E. Bourne, "Training principles to advance expertise," *Front. Psychol.*, vol. 5, p. 131, 2014, doi: 10.3389/fpsyg.2014.00131.
- [7] F. Paas and T. van Gog, "Principles for Designing Effective and Efficient Training of Complex Cognitive Skills," *Rev. Hum. Factors Ergon.*, vol. 5, no. 1, pp. 166–194, Jun. 2009, doi: 10.1518/155723409X448053.
- [8] R. S. Wyer, Jr., *The Automaticity of Everyday Life: Advances in Social Cognition, Volume X*, 1st ed. Psychology Press, 2014.
- [9] J. A. Bargh, "The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition," *Handb. Soc. Cogn.*, vol. 1, pp. 1–40, 1994.
- [10] R. M. Shiffrin and W. Schneider, "Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory.," *Psychol. Rev.*, vol. 84, no. 2, pp. 127–190, 1977, doi: 10.1037/0033-295X.84.2.127.
- [11] G. D. Logan, "Skill and automaticity: Relations, implications, and future directions.," *Can. J. Psychol. Can. Psychol.*, vol. 39, no. 2, pp. 367–386, 1985.
- [12] A. Moors and J. De Houwer, "Automaticity: A Theoretical and Conceptual Analysis.," *Psychol. Bull.*, vol. 132, no. 2, pp. 297–326, 2006, doi: 10.1037/0033-2909.132.2.297.
- [13] N. Segalowitz and J. Hulstijn, "Automaticity in bilingualism and second language learning," *Handb. Biling. Psycholinguist. Approaches*, pp. 371–388, 2005.
- [14] C. G. McCusker and B. Gettings, "Automaticity of cognitive biases in addictive behaviours: Further evidence with gamblers," *Br. J. Clin. Psychol.*, vol. 36, no. 4, pp. 543–554, 1997.
- [15] B. Gardner, "Habit as automaticity, not frequency," *Eur. Health Psychol.*, vol. 14, no. 2, pp. 32–36, 2012.
- [16] J. A. Bargh, "Automaticity in social psychology.," in *Social psychology: Handbook of basic principles*, ET Higgins&AW Kruglanski (Eds.), New York, US: Guilford Press, 1996, pp. 169–183.
- [17] J. A. Bargh and E. L. Williams, "The automaticity of social life," *Curr. Dir. Psychol. Sci.*, vol. 15, no. 1, pp. 1–4, 2006.
- [18] R. Gray, "Movement automaticity in sport," in *Routledge Handbook of Sport Expertise*, Routledge, 2015, pp. 100–109.
- [19] B. S. Bloom, "Automaticity: 'The Hands and Feet of Genius.,'" *Educ. Leadersh.*, vol. 43, no. 5, pp. 70–77, 1986.
- [20] D. B. Crawford, "The third stage of learning math facts: Developing automaticity," *R Instr. Solut.*, pp. 1–40, 2003.
- [21] J. Shine and R. Shine, "Delegation to automaticity: the driving force for cognitive evolution?," *Front. Neurosci.*, vol. 8, p. 90, 2014, doi: 10.3389/fnins.2014.00090.
- [22] J. A. Bargh and T. L. Chartrand, "The unbearable automaticity of being," *Am. Psychol.*, vol. 54, no. 7, pp. 462–479, 1999, doi: 10.1037/0003-066X.54.7.462.
- [23] K. A. Rawson, "Defining and Investigating Automaticity in Reading Comprehension," in *Psychology of Learning and Motivation*, vol. 52, Elsevier, 2010, pp. 185–230.
- [24] P. M. Fitts and M. I. Posner, *Human performance*. Oxford, England: Brooks/Cole, 1967.
- [25] J. R. Anderson, "Acquisition of cognitive skill," *Psychol. Rev.*, vol. 89, no. 4, pp. 369–406, 1982, doi: 10.1037/0033-295X.89.4.369.

- [26] D. M. Neves and J. R. Anderson, "Knowledge compilation: Mechanisms for the automatization of cognitive skills," *Cogn. Ski. Their Acquis.*, pp. 86–102, 1981.
- [27] J. Doyon, L. G. Ungerleider, L. Squire, and D. Schacter, "Functional anatomy of motor skill learning," in *Neuropsychology of memory*, 3rd ed., New York: Guilford Press, 2002, pp. 225–238.
- [28] J. Doyon and H. Benali, "Reorganization and plasticity in the adult brain during learning of motor skills," *Curr. Opin. Neurobiol.*, vol. 15, no. 2, pp. 161–167, Apr. 2005, doi: 10.1016/j.conb.2005.03.004.
- [29] J. Doyon, V. Penhune, and L. G. Ungerleider, "Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning," *Neuropsychologia*, vol. 41, no. 3, pp. 252–262, Jan. 2003, doi: 10.1016/S0028-3932(02)00158-6.
- [30] W. Schneider and R. M. Shiffrin, "Controlled and automatic human information processing: I. Detection, search, and attention," *Psychol. Rev.*, vol. 84, no. 1, pp. 1–66, 1977, doi: 10.1037/0033-295X.84.1.1.
- [31] G. D. Logan, "On the use of a concurrent memory load to measure attention and automaticity," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 5, no. 2, pp. 189–207, 1979, doi: 10.1037/0096-1523.5.2.189.
- [32] W. Schneider and A. D. Fisk, "Degree of consistent training: Improvements in search performance and automatic process development," *Percept. Psychophys.*, vol. 31, no. 2, pp. 160–168, 1982, doi: 10.3758/BF03206216.
- [33] A. D. Fisk, N. A. Oransky, and P. R. Skedsvold, "Examination of the role of 'higher-order' consistency in skill development," *Hum. Factors*, vol. 30, no. 5, pp. 567–581, 1988.
- [34] G. D. Logan, "Toward an instance theory of automatization," *Psychol. Rev.*, vol. 95, no. 4, pp. 492–527, 1988, doi: 10.1037/0033-295X.95.4.492.
- [35] F. T. Durso, N. M. Cooke, T. J. Breen, and R. W. Schvaneveldt, "Is consistent mapping necessary for high-speed search?," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 13, no. 2, p. 223, 1987, doi: 10.1037/0278-7393.13.2.223.
- [36] J. Tzelgov, V. Yehene, L. Kotler, and A. Alon, "Automatic comparisons of artificial digits never compared: learning linear ordering relations," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 26, no. 1, pp. 103–120, 2000, doi: 10.1037/0278-7393.26.1.103.
- [37] J. Duncan, "Consistent and varied training in the theory of automatic and controlled information processing," *Cognition*, vol. 23, no. 3, pp. 279–284, Aug. 1986, doi: 10.1016/0010-0277(86)90036-3.
- [38] A. D. Fisk and M. Eboch, "An automatic/controlled processing theory application to training component map reading skills," *Appl. Ergon.*, vol. 20, no. 1, pp. 2–8, Mar. 1989, doi: 10.1016/0003-6870(89)90002-1.
- [39] G. L. Myers and A. D. Fisk, "Training Consistent Task Components: Application of Automatic and Controlled Processing Theory to Industrial Task Training," *Hum. Factors*, vol. 29, no. 3, pp. 255–268, Jun. 1987, doi: 10.1177/001872088702900302.
- [40] A. D. Fisk and S. J. Lloyd, "The Role of Stimulus-to-Rule Consistency in Learning Rapid Application of Spatial Rules," *Hum. Factors*, vol. 30, no. 1, pp. 35–49, Feb. 1988, doi: 10.1177/001872088803000104.
- [41] A. D. Fisk and J. K. Gallini, "Training Consistent Components of Tasks: Developing an Instructional System Based on Automatic/Controlled Processing Principles," *Hum. Factors*, vol. 31, no. 4, pp. 453–463, Aug. 1989, doi: 10.1177/001872088903100408.
- [42] A. D. Fisk and W. A. Rogers, "The role of situational context in the development of high-performance skills," *Hum. Factors*, vol. 30, no. 6, pp. 703–712, 1988, doi: 10.1177/001872088803000606.
- [43] W. A. Rogers, M. D. Lee, and A. D. Fisk, "Contextual Effects on General Learning, Feature Learning, and Attention Strengthening in Visual Search," *Hum. Factors*, vol. 37, no. 1, pp. 158–172, Mar. 1995, doi: 10.1518/001872095779049507.



- [44] P. Ginns, "Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects," *Learn. Instr.*, vol. 16, no. 6, pp. 511–525, Dec. 2006, doi: 10.1016/j.learninstruc.2006.10.001.
- [45] J. F. Miller, E. M. Lazarus, S. M. Polyn, and M. J. Kahana, "Spatial clustering during memory search.," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 39, no. 3, pp. 773–781, 2013, doi: 10.1037/a0029684.
- [46] A. Cleeremans and J. L. McClelland, "Learning the Structure of Event Sequences," *J. Exp. Psychol. Gen.*, vol. 120, no. 3, pp. 235–253, 1991.
- [47] M. F. St. John and D. R. Shanks, "Implicit learning from an information processing standpoint," in *How implicit is implicit learning?*, New York, NY, US: Oxford University Press, 1997, pp. 162–194.
- [48] H. C. Warren, "Mental association from Plato to Hume," *Psychol. Rev.*, vol. 23, no. 3, pp. 208–230, 1916, doi: <http://dx.doi.org/10.1037/h0073099>.
- [49] G. D. Logan, "Automaticity, Resources, and Memory: Theoretical Controversies and Practical Implications," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 30, no. 5, pp. 583–598, Oct. 1988, doi: 10.1177/001872088803000504.
- [50] M. Vidulich, Y.-Y. Yeh, and W. Schneider, "Time-Compressed Components for Air-Intercept Control Skills," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Sage CA: Los Angeles, 1983, vol. 27, pp. 161–164, doi: <https://doi.org/10.1177/154193128302700211>.
- [51] E. L. Thorndike, *The psychology of learning*, vol. 2. Teachers College, Columbia University, 1913.
- [52] G. S. Snoddy, "Learning and stability: a psychophysiological analysis of a case of motor learning with clinical applications.," *J. Appl. Psychol.*, vol. 10, no. 1, pp. 1–36, 1926, doi: <http://dx.doi.org/10.1037/h0075814>.
- [53] A. Newell and P. S. Rosenbloom, "Mechanisms of skill acquisition and the law of practice," in *Cognitive skills and their acquisition*, Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers, 1981, pp. 1–55.
- [54] A. Heathcote, S. Brown, and D. J. K. Mewhort, "The power law repealed: The case for an exponential law of practice," *Psychon. Bull. Rev.*, vol. 7, no. 2, pp. 185–207, Jun. 2000, doi: 10.3758/BF03212979.
- [55] K. A. Ericsson, "The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance," in *The Cambridge Handbook of Expertise and Expert Performance*, K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman, Eds. Cambridge: Cambridge University Press, 2006, pp. 683–704.
- [56] K. A. Ericsson, R. T. Krampe, and C. Tesch-Römer, "The role of deliberate practice in the acquisition of expert performance.," *Psychol. Rev.*, vol. 100, no. 3, p. 363, 1993.
- [57] D. Z. Hambrick, F. L. Oswald, E. M. Altmann, E. J. Meinz, F. Gobet, and G. Campitelli, "Deliberate practice: Is that all it takes to become an expert?," *Intelligence*, vol. 45, pp. 34–45, Jul. 2014, doi: 10.1016/j.intell.2013.04.001.
- [58] B. N. Macnamara, D. Z. Hambrick, and F. L. Oswald, "Deliberate Practice and Performance in Music, Games, Sports, Education, and Professions: A Meta-Analysis," *Psychol. Sci.*, vol. 25, no. 8, pp. 1608–1618, Aug. 2014, doi: 10.1177/0956797614535810.
- [59] B. N. Macnamara, D. Moreau, and D. Z. Hambrick, "The relationship between deliberate practice and performance in sports: A meta-analysis," *Perspect. Psychol. Sci.*, vol. 11, no. 3, pp. 333–350, 2016.
- [60] S. K. Carpenter, N. J. Cepeda, D. Rohrer, S. H. K. Kang, and H. Pashler, "Using Spacing to Enhance Diverse Forms of Learning: Review of Recent Research and Implications for Instruction," *Educ. Psychol. Rev.*, vol. 24, no. 3, pp. 369–378, Sep. 2012, doi: 10.1007/s10648-012-9205-z.
- [61] H. Ebbinghaus, *Memory: A Contribution to Experimental Psychology (Henry A. Ruger & Clara E. Bussenius)*. New York: Teachers College, Columbia University, 1885.

- [62] N. J. Cepeda, H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer, "Distributed practice in verbal recall tasks: A review and quantitative synthesis.," *Psychol. Bull.*, vol. 132, no. 3, pp. 354–380, 2006, doi: 10.1037/0033-2909.132.3.354.
- [63] N. J. Cepeda, N. Coburn, D. Rohrer, J. T. Wixted, M. C. Mozer, and H. Pashler, "Optimizing Distributed Practice: Theoretical Analysis and Practical Implications," *Exp. Psychol.*, vol. 56, no. 4, pp. 236–246, Jan. 2009, doi: 10.1027/1618-3169.56.4.236.
- [64] D. Rohrer and H. Pashler, "Increasing Retention Without Increasing Study Time," *Curr. Dir. Psychol. Sci.*, vol. 16, no. 4, pp. 183–186, Aug. 2007, doi: 10.1111/j.1467-8721.2007.00500.x.
- [65] N. J. Cepeda, E. Vul, D. Rohrer, J. T. Wixted, and H. Pashler, "Spacing Effects in Learning: A Temporal Ridgeline of Optimal Retention," *Psychol. Sci.*, vol. 19, no. 11, pp. 1095–1102, Nov. 2008, doi: 10.1111/j.1467-9280.2008.02209.x.
- [66] D. A. Balota, J. M. Duchek, and J. M. Logan, "Is Expanded Retrieval Practice a Superior Form of Spaced Retrieval? A Critical Review of the Extant Literature," in *The foundations of remembering: Essays in honor of Henry L. Roediger, III*, J. S. Nairne (Ed.), New York, NY, US: Psychology Press, 2007, pp. 83–105.
- [67] J. M. Logan and D. A. Balota, "Expanded vs. Equal Interval Spaced Retrieval Practice: Exploring Different Schedules of Spacing and Retention Interval in Younger and Older Adults," *Aging Neuropsychol. Cogn.*, vol. 15, no. 3, pp. 257–280, Apr. 2008, doi: 10.1080/13825580701322171.
- [68] M. Servant, P. Cassey, G. F. Woodman, and G. D. Logan, "Neural bases of automaticity.," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 44, no. 3, pp. 440–464, Mar. 2018, doi: 10.1037/xlm0000454.
- [69] D. Rohrer, "Student Instruction Should Be Distributed Over Long Time Periods," *Educ. Psychol. Rev.*, vol. 27, no. 4, pp. 635–643, Dec. 2015, doi: 10.1007/s10648-015-9332-4.
- [70] S. B. O'Sullivan and T. J. Schmitz, *Physical rehabilitation*. Philadelphia: F.A. Davis, 2007.
- [71] S. Merbah and T. Meulemans, "Learning a Motor Skill: Effects of Blocked Versus Random Practice a Review," *Psychol. Belg.*, vol. 51, no. 1, pp. 15–48, Feb. 2011, doi: 10.5334/pb-51-1-15.
- [72] L. A. Stambaugh, "When Repetition Isn't the Best Practice Strategy: Effects of Blocked and Random Practice Schedules," *J. Res. Music Educ.*, vol. 58, no. 4, pp. 368–383, Jan. 2011, doi: 10.1177/0022429410385945.
- [73] J. B. Shea and R. L. Morgan, "Contextual Interference Effects on the Acquisition, Retention, and Transfer of a Motor Skill," *J. Exp. Psychol. [Hum. Learn.]*, vol. 5, no. 2, pp. 179–187, 1979, doi: 10.1037/0278-7393.5.2.179.
- [74] S. Goode and R. A. Magill, "Contextual Interference Effects in Learning Three Badminton Serves," *Res. Q. Exerc. Sport*, vol. 57, no. 4, pp. 308–314, Dec. 1986, doi: 10.1080/02701367.1986.10608091.
- [75] C. A. Wrisberg and Z. Liu, "The Effect of Contextual Variety on the Practice, Retention, and Transfer of an Applied Motor Skill," *Res. Q. Exerc. Sport*, vol. 62, no. 4, pp. 406–412, Dec. 1991, doi: 10.1080/02701367.1991.10607541.
- [76] K. G. Hall, D. A. Domingues, and R. Cavazos, "Contextual Interference Effects with Skilled Baseball Players," *Percept. Mot. Skills*, vol. 78, no. 3, pp. 835–841, Jun. 1994, doi: 10.1177/003151259407800331.
- [77] T. D. Lee and R. A. Magill, "The locus of contextual interference in motor-skill acquisition.," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 9, no. 4, pp. 730–746, 1983, doi: 10.1037/0278-7393.9.4.730.
- [78] T. D. Lee and R. A. Magill, "Can forgetting facilitate skill acquisition?," *Differing Perspect. Mot. Learn. Mem. Control*, vol. 27, no. Advances in psychology, pp. 3–22, 1985, doi: 10.1016/S0166-4115(08)62528-5.
- [79] J. B. Shea and S. T. Zimny, "Context Effects in Memory and Learning Movement Information," in *Advances in Psychology*, vol. 12, Elsevier, 1983, pp. 345–366.
- [80] F. Brady, "Contextual Interference: A Meta-Analytic Study," *Percept. Mot. Skills*, vol. 99, no. 1, pp. 116–126, 2004, doi: 10.2466/pms.99.1.116-126.

- [81] M. Bertollo, M. Berchicci, A. Carraro, S. Comani, and C. Robazza, "Blocked and Random Practice Organization in the Learning of Rhythmic Dance Step Sequences," *Percept. Mot. Skills*, vol. 110, no. 1, pp. 77–84, Feb. 2010, doi: 10.2466/pms.110.1.77-84.
- [82] E. Maas and K. A. Farinella, "Random Versus Blocked Practice in Treatment for Childhood Apraxia of Speech," *J. Speech Lang. Hear. Res.*, vol. 55, no. 2, pp. 561–578, Apr. 2012, doi: 10.1044/1092-4388(2011/11-0120).
- [83] S. V. Duff and A. M. Gordon, "Learning of grasp control in children with hemiplegic cerebral palsy," *Dev. Med. Child Neurol.*, vol. 45, no. 11, pp. 746–757, Nov. 2003, doi: 10.1017/S0012162203001397.
- [84] J. C. Naylor and G. E. Briggs, "Effects of task complexity and task organization on the relative efficiency of part and whole training methods," *J. Exp. Psychol.*, vol. 65, no. 3, pp. 217–224, 1963, doi: 10.1037/h0041060.
- [85] R. A. Magill and D. I. Anderson, *Motor learning and control: Concepts and applications*, vol. 11. New York: McGraw-Hill, 2007.
- [86] F. E. Fontana, O. Furtado, O. Mazzardo, and J. D. Gallagher, "Whole and Part Practice: A Meta-Analysis," *Percept. Mot. Skills*, vol. 109, no. 2, pp. 517–530, Oct. 2009, doi: 10.2466/pms.109.2.517-530.
- [87] R. A. Schmidt and C. A. Wrisberg, *Motor learning and performance: A situation-based learning approach*. Human kinetics, 2008.
- [88] D. L. Wright, C. B. Black, M. A. Immink, S. Brueckner, and C. Magnuson, "Long-Term Motor Programming Improvements Occur Via Concatenation of Movement Sequences During Random But Not During Blocked Practice," *J. Mot. Behav.*, vol. 36, no. 1, pp. 39–50, May 2004, doi: 10.3200/JMBR.36.1.39-50.
- [89] T. E. Gabriele, C. R. Hall, and E. E. Buckolz, "Practice schedule effects on the acquisition and retention of a motor skill," *Hum. Mov. Sci.*, vol. 6, no. 1, pp. 1–16, Mar. 1987, doi: 10.1016/0167-9457(87)90019-4.
- [90] E. Saemi, J. M. Porter, A. G. Varzaneh, M. Zarghami, and P. Shafinia, "Practicing along the contextual interference continuum: A comparison of three practice schedules in an elementary physical education setting," *Kinesiology*, vol. 44, no. 2, pp. 191–198, 2012.
- [91] E. L. Thorndike, *Educational Psychology*. New York: Lemke and Buechner, 1903.
- [92] D. H. Andrews, L. A. Carroll, and H. H. Bell, "The Future of Selective Fidelity in Training Devices," Armstrong Lab Wright-Patterson AFB OH Human Resources Directorate, AL/HR-TR-1995-0195, Mar. 1996. Accessed: Jun. 12, 2019. [Online]. Available: <https://apps.dtic.mil/docs/citations/ADA316902>.
- [93] D. K. Detterman and R. J. Sternberg, *Transfer on trial: Intelligence, cognition, and instruction*. Ablex Publishing, 1993.
- [94] S. M. Alessi, "Fidelity in the design of instructional simulations," *J. Comput.-Based Instr.*, vol. 15, no. 2, pp. 40–47, 1988.
- [95] C. Noble, "The relationship between fidelity and learning in aviation training and assessment," *J. Air Transp.*, vol. 7, pp. 33–54, 2002.
- [96] G. S. Watson, "Simulator effects in a high fidelity driving simulator as a function of visuals and motion," in *Driving Simulation Conference, DSC*, 1995, vol. 95.
- [97] S. M. Alessi, "Simulation design for training and assessment," in *Aircrew training and assessment*, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2000, pp. 197–222.
- [98] D. Liu, N. D. Macchiarella, and D. A. Vincenzi, "Simulation fidelity," in *Human factors in simulation and training*, CRC Press, 2009, pp. 61–73.
- [99] M. Mori, "Bukimi no tani [the uncanny valley]," *Energy*, vol. 7, pp. 33–35, 1970.
- [100] H. Brenton, M. Gillies, D. Ballin, and D. Chatting, "The uncanny valley: does it exist," Edinburgh, Napier University, 2005.
- [101] H. C. Ellis, *The transfer of learning*. Oxford, England: Macmillan, 1965.
- [102] J. W. Pellegrino, S. M. Doane, S. C. Fischer, and D. Alderton, "Stimulus Complexity Effects in Visual Comparisons: The Effects of Practice and Learning Context," *Journal Exp. Psychol. Hum. Percept. Perform.*, vol. 17, no. 3, pp. 781–791, 1991.

- [103] M. C. D'Angelo, B. Milliken, L. Jiménez, and J. Lupiáñez, "Implementing flexibility in automaticity: Evidence from context-specific implicit sequence learning," *Conscious. Cogn.*, vol. 22, no. 1, pp. 64–81, Mar. 2013, doi: 10.1016/j.concog.2012.11.002.
- [104] D. R. Thomson and B. Milliken, "Contextual distinctiveness produces long-lasting priming of pop-out," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 39, no. 1, pp. 202–215, 2013, doi: 10.1037/a0028069.
- [105] S. M. Alessi, "Dynamic vs. static fidelity in a procedural simulation," 1995.
- [106] D. P. M. van V. R. Fellow and G. Wulf, "Extrinsic feedback for motor learning after stroke: What is the evidence?," *Disabil. Rehabil.*, vol. 28, no. 13–14, pp. 831–840, Jan. 2006, doi: 10.1080/09638280500534937.
- [107] G. Wulf and C. H. Shea, "Understanding the role of augmented feedback: The good, the bad and the ugly," in *Skill Acquisition in Sport: Research, Theory and Practice*, A. M. Williams, N. J. Hodges (Eds.), London: Routledge, 2004, pp. 145–168.
- [108] C. M. Walsh, S. C. Ling, C. S. Wang, and H. Carnahan, "Concurrent Versus Terminal Feedback: It May Be Better to Wait," *Acad. Med.*, vol. 84, no. 10, p. S54, Oct. 2009, doi: 10.1097/ACM.0b013e3181b38daf.
- [109] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Terminal Feedback Outperforms Concurrent Visual, Auditory, and Haptic Feedback in Learning a Complex Rowing-Type Task," *J. Mot. Behav.*, vol. 45, no. 6, pp. 455–472, Nov. 2013, doi: 10.1080/00222895.2013.826169.
- [110] L. R. Wishart, T. D. Lee, S. J. Cunningham, and J. E. Murdoch, "Age-related differences and the role of augmented visual feedback in learning a bimanual coordination pattern," *Acta Psychol. (Amst.)*, vol. 110, no. 2, pp. 247–263, Jun. 2002, doi: 10.1016/S0001-6918(02)00036-7.
- [111] Y. Blandin, L. Toussaint, and C. H. Shea, "Specificity of practice: Interaction between concurrent sensory information and terminal feedback," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 34, no. 4, pp. 994–1000, 2008, doi: 10.1037/0278-7393.34.4.994.
- [112] M. Huet, D. M. Jacobs, C. Camachon, C. Goulon, and G. Montagne, "Self-Controlled Concurrent Feedback Facilitates the Learning of the Final Approach Phase in a Fixed-Base Flight Simulator," *Hum. Factors*, vol. 51, no. 6, pp. 858–871, Dec. 2009, doi: 10.1177/0018720809357343.
- [113] S. P. Swinnen, T. D. Lee, S. Verschueren, D. J. Serrien, and H. Bogaerds, "Interlimb coordination: Learning and transfer under different feedback conditions," *Hum. Mov. Sci.*, vol. 16, no. 6, pp. 749–785, 1997, doi: 10.1016/S0167-9457(97)00020-1.
- [114] G. Wulf, C. H. Shea, and S. Matschiner, "Frequent feedback enhances complex motor skill learning," *J. Mot. Behav.*, vol. 30, no. 2, pp. 180–192, 1998, doi: 10.1080/00222899809601335.
- [115] R. A. Schmidt and G. Wulf, "Continuous Concurrent Feedback Degrades Skill Learning: Implications for Training and Simulation," *Hum. Factors*, vol. 39, no. 4, pp. 509–525, Dec. 1997, doi: 10.1518/001872097778667979.
- [116] Q. Lai and C. H. Shea, "Generalized Motor Program (GMP) Learning: Effects of Reduced Frequency of Knowledge of Results and Practice Variability," *J. Mot. Behav.*, vol. 30, no. 1, pp. 51–59, Mar. 1998, doi: 10.1080/00222899809601322.
- [117] C. J. Winstein and R. A. Schmidt, "Reduced frequency of knowledge of results enhances motor skill learning," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 16, no. 4, pp. 677–691, 1990, doi: 10.1037/0278-7393.16.4.677.
- [118] D. Krause, M. Agethen, and C. Zobe, "Error Feedback Frequency Affects Automaticity But Not Accuracy and Consistency After Extensive Motor Skill Practice," *J. Mot. Behav.*, vol. 50, no. 2, pp. 144–154, Mar. 2018, doi: 10.1080/00222895.2017.1327406.
- [119] R. Eberts and W. Schneider, "Internalizing the System Dynamics for a Second-Order System," *Hum. Factors*, vol. 27, no. 4, pp. 371–393, Aug. 1985, doi: 10.1177/001872088502700403.
- [120] T. A. Ryan, *Intentional behavior: An approach to human motivation*. Oxford, England: Ronald Press., 1970.
- [121] E. A. Locke and G. P. Latham, *A theory of goal setting and task performance*. 1990. Englewood Cliffs, NJ, US: Prentice-Hall, Inc., 1990.

- [122] E. A. Locke and G. P. Latham, "Building a practically useful theory of goal setting and task motivation: A 35-year odyssey.," *Am. Psychol.*, vol. 57, no. 9, pp. 705–717, 2002, doi: 10.1037//0003-066X.57.9.705.
- [123] E. A. Locke, "Toward a theory of task motivation and incentives," *Organ. Behav. Hum. Perform.*, vol. 3, no. 2, pp. 157–189, May 1968, doi: 10.1016/0030-5073(68)90004-4.
- [124] M. Erez and I. Zidon, "Effect of goal acceptance on the relationship of goal difficulty to performance.," *J. Appl. Psychol.*, vol. 69, no. 1, p. 69, 1984, doi: 10.1037/0021-9010.69.1.69.
- [125] E. A. Locke, D.-O. Chah, S. Harrison, and N. Lustgarten, "Separating the effects of goal specificity from goal level," *Organ. Behav. Hum. Decis. Process.*, vol. 43, no. 2, pp. 270–287, 1989, doi: 10.1016/0749-5978(89)90053-8.
- [126] H. J. Klein, M. J. Wesson, J. R. Hollenbeck, and B. J. Alge, "Goal commitment and the goal-setting process: conceptual clarification and empirical synthesis.," *J. Appl. Psychol.*, vol. 84, no. 6, p. 885, 1999.
- [127] G. P. Latham and E. A. Locke, "Self-Regulation through Goal Setting," *Organ. Behav. Hum. Decis. Process.*, vol. 50, no. 2, pp. 212–247, 1991, doi: [https://doi.org/10.1016/0749-5978\(91\)90021-K](https://doi.org/10.1016/0749-5978(91)90021-K).
- [128] G. P. Latham, M. Erez, and E. A. Locke, "Resolving scientific disputes by the joint design of crucial experiments by the antagonists: Application to the Erez–Latham dispute regarding participation in goal setting.," *J. Appl. Psychol.*, vol. 73, no. 4, pp. 753–772, 1988, doi: 10.1037/0021-9010.73.4.753.
- [129] M. Erez, "Feedback: A necessary condition for the goal setting-performance relationship.," *J. Appl. Psychol.*, vol. 62, no. 5, p. 624, 1977, doi: 10.1037/0021-9010.62.5.624.
- [130] L. J. Becker, "Joint effect of feedback and goal setting on performance: A field study of residential energy conservation.," *J. Appl. Psychol.*, vol. 63, no. 4, p. 428, 1978.
- [131] A. Bandura and D. Cervone, "Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems," *J. Pers. Soc. Psychol.*, vol. 45, no. 5, pp. 1017–1028, 1983, doi: 10.1037/0022-3514.45.5.1017.
- [132] H. R. Strang, E. C. Lawrence, and P. C. Fowler, "Effects of assigned goal level and knowledge of results on arithmetic computation: A laboratory study.," *J. Appl. Psychol.*, vol. 63, no. 4, p. 446, 1978, doi: 10.1037/0021-9010.63.4.446.
- [133] R. E. Wood, A. J. Mento, and E. A. Locke, "Task complexity as a moderator of goal effects: A meta-analysis.," *J. Appl. Psychol.*, vol. 72, no. 3, p. 416, 1987, doi: 10.1037/0021-9010.72.3.416.
- [134] R. Kanfer and P. L. Ackerman, "Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition.," *J. Appl. Psychol.*, vol. 74, no. 4, p. 657, 1989, doi: 10.1037/0021-9010.74.4.657.
- [135] D. Winters and G. P. Latham, "The effect of learning versus outcome goals on a simple versus a complex task," *Group Organ. Manag.*, vol. 21, no. 2, pp. 236–250, 1996, doi: 10.1177/1059601196212007.
- [136] G. H. Seijts and G. P. Latham, "The effect of distal learning, outcome, and proximal goals on a moderately complex task," *J. Organ. Behav. Int. J. Ind. Occup. Organ. Psychol. Behav.*, vol. 22, no. 3, pp. 291–307, 2001, doi: 10.1002/job.70.
- [137] A. Bandura and D. H. Schunk, "Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation.," *J. Pers. Soc. Psychol.*, vol. 41, no. 3, pp. 586–598, 1981, doi: 10.1037/0022-3514.41.3.586.
- [138] G. P. Latham and G. H. Seijts, "The effects of proximal and distal goals on performance on a moderately complex task," *J. Organ. Behav. Int. J. Ind. Occup. Organ. Psychol. Behav.*, vol. 20, no. 4, pp. 421–429, 1999, doi: 10.1002/(SICI)1099-1379(199907)20:4<421::AID-JOB896>3.0.CO;2-%23.
- [139] F. M. Andrews and G. F. Farris, "Time pressure and performance of scientists and engineers: A five-year panel study," *Organ. Behav. Hum. Perform.*, vol. 8, no. 2, pp. 185–200, 1972.
- [140] A. C. Bluedorn and R. B. Denhardt, "Time and organizations," *J. Manag.*, vol. 14, no. 2, pp. 299–320, 1988, doi: 10.1177/014920638801400209.

- [141] Y. Fried and L. H. Slowik, "Enriching goal-setting theory with time: An integrated approach," *Acad. Manage. Rev.*, vol. 29, no. 3, pp. 404–422, 2004, doi: 10.5465/amr.2004.13670973.
- [142] B. F. Skinner, *The behavior of organisms: an experimental analysis*. Oxford, England: Appleton-Century, 1938.
- [143] B. F. Skinner, "The Experimental Analysis of Behavior," *Am. Sci.*, vol. 45, no. 4, pp. 343–371, 1957.
- [144] C. B. Ferster and B. F. Skinner, *Schedules of reinforcement*. East Norwalk, CT, US: Appleton-Century-Crofts, 1957.
- [145] G. Yukl, K. N. Wexley, and J. D. Seymore, "Effectiveness of pay incentives under variable ratio and continuous reinforcement schedules," *J. Appl. Psychol.*, vol. 56, no. 1, pp. 19–23, 1972, doi: 10.1037/h0032137.
- [146] G. E. Zuriff, "A Comparison of Variable-Ratio and Variable-Interval Schedules of Reinforcement1," *J. Exp. Anal. Behav.*, vol. 13, no. 3, pp. 369–374, 1970, doi: 10.1901/jeab.1970.13-369.
- [147] L. M. Saari and G. P. Latham, "Employee reaction to continuous and variable ratio reinforcement schedules involving a monetary incentive," *J. Appl. Psychol.*, vol. 67, no. 4, pp. 506–508, 1982, doi: 10.1037/0021-9010.67.4.506.
- [148] R. V. D. Luca and S. W. Holborn, "Effects of a Variable-Ratio Reinforcement Schedule with Changing Criteria on Exercise in Obese and Nonobese Boys," *J. Appl. Behav. Anal.*, vol. 25, no. 3, pp. 671–679, 1992, doi: 10.1901/jaba.1992.25-671.
- [149] G. Richter, D. R. Raban, and S. Rafaeli, "Studying Gamification: The Effect of Rewards and Incentives on Motivation," in *Gamification in Education and Business*, T. Reiners and L. C. Wood, Eds. Cham: Springer International Publishing, 2015, pp. 21–46.
- [150] A. Francisco-Aparicio, F. L. Gutiérrez-Vela, J. L. Isla-Montes, and J. L. G. Sánchez, "Gamification: Analysis and Application," London, 2013, pp. 113–126, doi: 10.1007/978-1-4471-5445-7\_9.
- [151] E. L. Deci, "Intrinsic motivation, extrinsic reinforcement, and inequity," *J. Pers. Soc. Psychol.*, vol. 22, no. 1, pp. 113–120, 1972, doi: 10.1037/h0032355.
- [152] J. Cameron and W. D. Pierce, "Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis," *Rev. Educ. Res.*, vol. 64, no. 3, pp. 363–423, 1994, doi: 10.2307/1170677.
- [153] R. M. Ryan, V. Mims, and R. Koestner, "Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory," *J. Pers. Soc. Psychol.*, vol. 45, no. 4, pp. 736–750, 1983, doi: 10.1037/0022-3514.45.4.736.
- [154] D. D. Deshler, G. R. Alley, M. M. Warner, and J. B. Schumaker, "Instructional Practices for Promoting Skill Acquisition and Generalization in Severely Learning Disabled Adolescents," *Learn. Disabil. Q.*, vol. 4, no. 4, pp. 415–421, Nov. 1981, doi: 10.2307/1510744.
- [155] M. Csikszentmihalyi, "Play and intrinsic rewards," *J. Humanist. Psychol.*, vol. 15, no. 3, pp. 41–63, 1975, doi: 10.1177/002216787501500306.
- [156] S. T. Klapp, C. A. Boches, M. L. Trabert, and G. D. Logan, "Automatizing alphabet arithmetic: II. Are there practice effects after automaticity is achieved?," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 17, no. 2, pp. 196–209, 1991, doi: 10.1037/0278-7393.17.2.196.
- [157] A. D. Fisk and W. Schneider, "Control and Automatic Processing During Tasks Requiring Sustained Attention: A New Approach to Vigilance," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 23, no. 6, pp. 737–750, Dec. 1981, doi: 10.1177/001872088102300610.
- [158] W. Schneider and J. M. Chein, "Controlled & automatic processing: behavior, theory, and biological mechanisms," *Cogn. Sci.*, vol. 27, no. 3, pp. 525–559, May 2003, doi: 10.1207/s15516709cog2703\_8.
- [159] A. D. Fisk and W. Schneider, "Type of Task Practice and Time-Sharing Activities Predict Performance Deficits Due to Alcohol Ingestion," in *Proceedings of the Human Factors Society Annual Meeting*, Sage CA: Los Angeles, 1982, vol. 26, pp. 926–930.
- [160] T. R. Kirchner and M. A. Sayette, "Effects of alcohol on controlled and automatic memory processes," *Exp. Clin. Psychopharmacol.*, vol. 11, no. 2, pp. 167–175, 2003, doi: 10.1037/1064-1297.11.2.167.

- [161] M. Lorains, C. MacMahon, K. Ball, and J. Mahoney, "Above Real Time Training for Team Invasion Sport Skills," *Int. J. Sports Sci. Coach.*, vol. 6, no. 4, pp. 537–544, Dec. 2011, doi: 10.1260/1747-9541.6.4.537.
- [162] D. C. Donderi, K. K. Niall, K. Fish, and B. Goldstein, "Above-Real-Time Training (ARTT) Improves Transfer to a Simulated Flight Control Task," *Hum. Factors*, vol. 54, no. 3, pp. 469–479, Jun. 2012, doi: 10.1177/0018720812439711.
- [163] D. Guckenberger, K. C. Uliano, and N. E. Lane, "Training high performance skills using above real-time training," NASA-CR-192616, NAS 1.26:192616, Jan. 1993. Accessed: Jul. 03, 2019. [Online]. Available: <https://ntrs.nasa.gov/search.jsp?R=19930015003>.
- [164] M. Lorains, K. Ball, and C. MacMahon, "An above real time training intervention for sport decision making," *Psychol. Sport Exerc.*, vol. 14, no. 5, pp. 670–674, Sep. 2013, doi: 10.1016/j.psychsport.2013.05.005.
- [165] J. A. Bergstrom, "Experiments upon physiological memory by means of interference of associations," *Am. J. Psychol.*, vol. 5, no. 3, pp. 356–369, 1893.
- [166] J. A. Bergstrom, "The relation of the interference to the practice effect of an association," *Am. J. Psychol.*, vol. 6, no. 3, pp. 433–442, 1894, doi: 10.2307/1411651.
- [167] G. Müller and A. Pilzecker, "Experimental contributions to the theory of memory," *Z Psychol Z Angew Psychol*, vol. 1, pp. 1–288, 1900.
- [168] B. J. Underwood, "Interference and forgetting," *Psychol. Rev.*, vol. 64, no. 1, pp. 49–60, 1957, doi: 10.1037/h0044616.
- [169] J. Ceraso, "The Interference Theory of Forgetting," *Sci. Am.*, vol. 217, no. 4, pp. 117–127, 1967.
- [170] I. M. Bilodeau and H. Schlosberg, "Similarity in stimulating conditions as a variable in retroactive inhibition," *J. Exp. Psychol.*, vol. 41, no. 3, pp. 199–204, 1951, doi: 10.1037/h0056809.
- [171] C. E. Osgood, "The similarity paradox in human learning: a resolution," *Psychol. Rev.*, vol. 56, no. 3, pp. 132–143, 1949, doi: 10.1037/h0057488.
- [172] J. T. Wixted, "The role of retroactive interference and consolidation in everyday forgetting," in *Forgetting*, New York, NY, US: Psychology Press, 2010, pp. 285–312.
- [173] J. W. Krakauer, C. Ghez, and M. F. Ghilardi, "Adaptation to Visuomotor Transformations: Consolidation, Interference, and Forgetting," *J. Neurosci.*, vol. 25, no. 2, pp. 473–478, Jan. 2005, doi: 10.1523/JNEUROSCI.4218-04.2005.
- [174] T. Brashers-Krug, R. Shadmehr, and E. Bizzi, "Consolidation in human motor memory," *Nature*, vol. 382, no. 6588, pp. 252–255, Jul. 1996, doi: 10.1038/382252a0.
- [175] R. Shadmehr and T. Brashers-Krug, "Functional stages in the formation of human long-term motor memory," *J. Neurosci.*, vol. 17, no. 1, pp. 409–419, 1997, doi: 10.1523/JNEUROSCI.17-01-00409.1997.
- [176] S. E. Criscimagna-Hemminger and R. Shadmehr, "Consolidation patterns of human motor memory," *J. Neurosci.*, vol. 28, no. 39, pp. 9610–9618, 2008, doi: 10.1523/JNEUROSCI.3071-08.2008.
- [177] G. Cantarero, B. Tang, R. O'Malley, R. Salas, and P. Celnik, "Motor Learning Interference Is Proportional to Occlusion of LTP-Like Plasticity," *J. Neurosci.*, vol. 33, no. 11, pp. 4634–4641, Mar. 2013, doi: 10.1523/JNEUROSCI.4706-12.2013.

## Contents

|  |            |
|--|------------|
| <b>CHAPTER V Validation of the Gamification Models for Motivation and Automaticity Acquisition .....</b> | <b>92</b>  |
| <b>1. Introduction .....</b>   | <b>92</b>  |
| <b>2. State of the Art on the Gamification and Serious Game Models .....</b>                             | <b>93</b>  |
| 2.1. Game Object Model (GOM) .....   | 93         |
| 2.2. Killi's Experiential Gaming Model.....  | 93         |
| 2.3. Rooney's Triadic Model.....   | 93         |
| 2.4. LM-GM Model.....  | 94         |
| 2.5. Activity Theory-based Model for Serious Games (ATMSG).....  | 94         |
| 2.6. The Four-Dimensional Model.....   | 95         |
| 2.7. The RETAIN Model.....   | 95         |
| 2.8. Octalysis .....   | 95         |
| 2.9. Annetta's "I's" Model .....   | 96         |
| 2.10. Cognitive Behavioral Game Design.....  | 96         |
| 2.11. P-III Model .....  | 96         |
| 2.12. General Framework for Digital Game-Based Training Systems .....                                    | 96         |
| 2.13. Van Eck's Problem Type-Gameplay Type Framework.....  | 97         |
| <b>3. Game Design Based on Motivation Theories .....</b>   | <b>97</b>  |
| 3.1. Integration of Self-Determination and Self-Efficacy in Game Design .....                            | 98         |
| 3.2. Game Design Process.....  | 99         |
| 3.3. Experiment Design.....  | 101        |
| 3.4. Data Analysis and Results.....  | 102        |
| <b>4. Game Design for Automaticity Acquisition .....</b>   | <b>105</b> |
| 4.1. Automaticity Gamification Model .....   | 105        |
| 4.2. Game Design Process.....  | 106        |
| 4.3. Experiment Design.....  | 108        |
| 4.4. Data Analysis and Results.....  | 110        |
| <b>5. Conclusion .....</b>   | <b>111</b> |
| <b>6. References.....</b>  | <b>112</b> |



## CHAPTER V

### Validation of the Gamification Models for Motivation and Automaticity Acquisition

#### 1. Introduction

Acquisition of reading, particularly for opaque orthographies, takes years of practice. Hence, any reading program has to consider the motivational factors affecting the engagement and adherence of the learners. Motivation is crucial for in-home technology-based reading programs that work without the presence of teachers. Video games have been known to produce a surge in motivation and engagement levels for a long time [1], [2]. The motivational boost produced by video games motivated scientists and researchers to exploit game elements and mechanics for non-game purposes. This led to the emergence of gamification and serious games [3], [4].

With the artistic aspects of graphic design and storytelling, video game design has always been partly in the realm of art. Researchers of gamification and serious games have used scientific theories to model the arbitrary process of game design. This chapter has highlighted 13 of these proposed models.

There are many disadvantages in the current gamification models. The first and most important drawback is the lack of experimental evaluation of their effectiveness. Furthermore, the role of important motivational theories, such as self-efficacy [5], [6], and self-determination [7], [8], are not fully explored in current models. Moreover, the game design for different learning categories, such as declarative knowledge and automaticity acquisition, is not differentiated. Learning declarative knowledge requires different types of practice than acquiring an automatic skill. Automaticity acquisition, which is among the main focuses of this thesis, evolves through different phases. The effective training and motivational mechanisms differ in each phase. Hence, automaticity training requires a tailored temporally sensitive game design.

For addressing these disadvantages, two gamification models were proposed in this chapter, and their efficacies were evaluated through randomized controlled trials. The first experimental study evaluated the game design based on the integration of Self-Efficacy Theory (SET) and Self-Determination Theory (SDT). In the first step, the game elements were mapped to each component of the two theories. Then, an off-the-shelf game was selected, and three game design variations were developed. Two of the designs were based on either of the theories, while the third design integrated both Self-Efficacy and Self-Determination theories. The effectiveness of these designs was evaluated by involving 46 participants.

In the second experimental study of this chapter, a gamification model was developed for training the automaticity acquisition. This model was designed based on the training model of automaticity acquisition presented in the previous chapter. The effectiveness of this gamification model was evaluated through a randomized controlled trial. An off-the-shelf game with a task that could be automatized was selected. This game was redesigned based on the

automaticity acquisition gamification model. Then, the new design was compared to the original design through an experimental study that involved 49 participants. A multi-tasking challenge was integrated into both designs to evaluate the automaticity level at the end of the experiment.

In the following section, the current state of the art on the gamification and serious game design models are summarized. Section 3 presents the gamification model based on the integration of self-efficacy and self-determination theories. It also presents the experimental study for evaluating the effectiveness of this model. Section 4 presents a stage-based gamification model for automaticity acquisition and the experimental study for evaluating the effectiveness of this model.

## **2. State of the Art on the Gamification and Serious Game Models**

In this section, 13 of the current gamification and serious game models are presented. Most of the models have focused on the gamification of learning.

### **2.1. Game Object Model (GOM)**

Amory et al. [9], [10] proposed a serious game design model called Game Object Model (GOM) for the design and evaluation of the educational games. This model was designed based on the concept of Object-Oriented Programming (OOP). It consists of Objects and interfaces. Objects represent the main components of an educational game, and each of them can contain other interfaces and objects. Interfaces are smaller constructs that describe an object, and they are divided into abstract and concrete categories. Abstract interfaces represent the pedagogical and theoretical elements, while the concrete interfaces refer to the game design elements. In a later study, Amory presented the Game Object Model Version II (GOM II) [11] to include more objects and interfaces. Social aspects of educational game design were added to the previous model. In addition, the interfaces were categorized into six core concepts: game definition, authentic learning, narrative, gender, social collaboration, and challenges-puzzles-quests.

### **2.2. Killi's Experiential Gaming Model**

Kiili [12] proposed a model for designing educational games. It was based on the experiential learning theory [13] and the flow theory [14], [15]. It mixed the stages of Kolb's experiential model with the key requirements of Csikszentmihalyi's flow theory. The experiential theory stages are: concrete experience, reflective observation, abstract conceptualization, and active experimentation. Flow theory requirements are clear goals, immediate feedback, and balance between challenges and skills. However, this game design model was not evaluated.

### **2.3. Rooney's Triadic Model**

Rooney [16] proposed a triadic model for the design of educational serious games. The model consisted of three components of pedagogy, play, and fidelity. The pedagogy component

was based on the constructivist learning model [17]. The constructivist model emphasizes not only the learning experience but also the authenticity and fidelity to real-world situations. However, they argued that too much physical and functional fidelity in game design could be detrimental to the play experience. Therefore, a balance between the play experience and the pedagogical goals should be reached. Nevertheless, this model remained theoretical, and its effectiveness was not evaluated.

#### **2.4. LM-GM Model**

Suttie et al. [18] introduced a gamification framework called LM-GM that maps game mechanics onto learning mechanics. The authors used the learning mechanics and game mechanics existing in the scientific literature. Then, they linked the two sets through a classification based on Bloom's higher-order thinking skills [19], [20]. These thinking skills are creating, evaluating, analyzing, applying, understanding, and retention. Arnab et al. [21] have explained further the details of this model. They evaluated the design of the Re-Mission game based on the LM-GM model [22]. In addition, this model was compared to the GOM model. The preliminary evaluation revealed that LM-GM provided more knowledge for recognizing the pedagogical and game patterns. However, the effectiveness of the model in creating a game was not evaluated.

#### **2.5. Activity Theory-based Model for Serious Games (ATMSG)**

Carvalho et al. [23], [24] presented a model for designing and evaluating educational serious games based on the activity theory [25]. This Activity Theory-based Model for Serious Games (ATMSG) consisted of the three distinct activities of gaming, learning, and instructional. For each activity, exhaustive lists of possible actions, tools, and goals were provided. In the first step, the model asks the designers or evaluators to describe the activities and determine the subjects and their motives. The three types of subjects are learners, instructors, and game designers. The second step is to represent the structure of the game in a sequence visualized in Unified Modeling Language (UML) format. The third step is to identify the actions, tools, and goals for each activity. Finally, the fourth step is to group the actions, tools, and goals of an activity and to provide a more detailed description of its implementation. When these four steps are completed and combined, a document containing a thorough analysis of the game will be acquired. This model was evaluated and compared to the LM-GM model. The results proved that ATMSG model provided a more comprehensive and detailed analysis of the games, however, it was at the expense of usability, as the users with less gaming experience struggled to implement the model.

Callaghan et al. [26] presented an extension to the ATMSG model in which they included the analytics part of the game. The authors proposed using a set of standard traces to capture the relevant data during the gameplay. It included the traces related to the game's start, quit, end, phase change, input, and other variables related to serious games. They also suggested adding another step to map actions to appropriate traces. Furthermore, game traces were added to the current activities of learning, gaming, and instructional. Finally, the authors explored the design of a serious game called *Circuit Warz* by using this extended ATMSG model. However,

they did not conduct any experimental study to evaluate the effectiveness of a game designed based on this model.

## 2.6. The Four-Dimensional Model

De Freitas et al. [27] proposed a four-dimensional framework to evaluate game-based and simulation-based learnings. These four dimensions are context, learner specification, pedagogic consideration, and mode of representation. The authors provided a checklist table with four columns, each corresponding to a dimension of the framework. Inside each checklist, there are questions about the details of that dimension. For example, for the context dimension, there are questions such as “What is the context for learning?”, “Does the context affect learning?”. They ultimately tested this framework by evaluating two games called *MediaStage* and *Savannah*. However, like the other serious game models, the effectiveness of the model in designing a game was not evaluated.

## 2.7. The RETAIN Model

Gunter et al. [28] proposed the RETAIN model to design and evaluate educational games. It is abbreviated from these six words: Relevance, Embedding, Transfer, Adaptation, Immersion, and Naturalization. Relevance is the consideration of the learner’s needs, learning style, and previous knowledge. Embedding determines how well educational content fits into the game’s fantasy and narrative. Transfer is the application of the acquired knowledge to new situations. Adaptation is going beyond transfer and creating new knowledge based on previously acquired knowledge. Immersion refers to “the active creation of belief in the game’s fantasy context”. Finally, Naturalization refers to reaching the mastery level in the acquired knowledge.

After identifying the six criteria, the authors provided a table with four levels for each criterion. Then, by collecting the experts’ opinions, they determined a weight for each criterion. Finally, they evaluated the two games of “*Math Blaster*” and “*Where in the World is Carmen Sandiego?*” using the RETAIN model. However, the model was not adopted to create serious games, and its effectiveness was not evaluated.

## 2.8. Octalysis

Chou, in his book *Actionable Gamification: beyond points, badges, and leaderboards* [29], has presented a model for gamification called Octalysis. It classified the game mechanics into the eight categories of Meaning, Empowerment, Social Influence, Unpredictability, Avoidance, Scarcity, Ownership, and Accomplishment. The author called each of these categories core drives that were presented on an octagon. Each core drive was positioned on a vertex of the octagon.

The right side of the octagon represents the game mechanics that tend to foster intrinsic motivation. The left side represents the game mechanics that result in extrinsic motivation. Furthermore, the top side of the octagon represents the game mechanics that create positive types of motivation such as gaining points and progress. The bottom half represents the game

mechanics that produce negative types of motivation, such as loss avoidance. Despite the practicality of this gamification model, despite it was not scientifically evaluated.

### **2.9. Annetta's "I's" Model**

Annetta [30] proposed a model with six nested elements for educational serious games. The first element is Identity (e.g., avatar, character), which is fundamental to other elements. The second element is Immersion, which is achieved by an enhanced sense of presence and engagement in the content. The third element is Interactivity, which could foster engagement. The fourth element is Increased Complexity, which refers to offering a balance between the abilities of the user and challenges in the game. As the user's abilities improve, the game should become increasingly challenging. The fifth element is Informed Teaching, which refers to the mechanisms for recording the user's behavior inside the game. Collecting performance-related data creates an opportunity for analyzing and improving the educational game. Finally, the last element that encompasses all the aforementioned elements is Instructional, which is the ultimate goal of any educational serious game. However, this model was not evaluated.

### **2.10. Cognitive Behavioral Game Design**

Starks [31] presented a serious game design model called Cognitive Behavioral Game Design (CBGD). This model was the combination of the Social Cognitive Theory (SCT) developed by Bandura [32], the Multiple Intelligence Theory of Gardner [33], and abstract game elements. The social cognitive elements were placed on the top and included knowledge, goals, outcome expectations, encouragement, and barriers. The abstract game elements were placed on the bottom and included engagement, challenge, flow, persistence, and mastery. Finally, the multiple intelligence elements were placed in the middle and included graphics, music/sound, physical movement, humor, space/positioning, narrative, logic/pattern, nature, relationships/role models, math/numbers, word/language, and personal reflection. The authors did not report any validation study of this model.

### **2.11. P-III Model**

Abeele et al. [34] proposed the P-III model for designing and developing serious games. This model defines the pillars and flow of serious game design. The four pillars specified in the model were player-centered design, iterative development, interdisciplinary team, integrated play, and learning. The serious game design flow had three phases of Concept Design, Game Design, and Game Development. Concept Design included user & task analysis, participatory design, game concept definition, and outsider expert discussion. Game Design included storyboards & focus groups, paper prototypes & play tests, game design document, and outsider expert review. Game Development included art & software development, lo-fi to hi-fi prototypes & playtests, final prototype, and outsider expert testing.

### **2.12. General Framework for Digital Game-Based Training Systems**

Brennecke et al. [35] proposed a framework for game design made of three main components. The Authoring component is for the teacher who has to provide the students with

a back story (introductory text or visuals introducing the players to the pre-plot of the game). The Training component is for the students who should play the main story (training session). The Reviewing component refers to the communication and interaction between the student and the teacher, which generate a visual review log that can be used to evaluate and assess the training session. Finally, this proposed framework was used to implement a serious game for crime scene investigation training called *OpenCrimeScene* [36]. However, the authors did not provide any information about the effectiveness of this model.

### **2.13. Van Eck's Problem Type-Gameplay Type Framework**

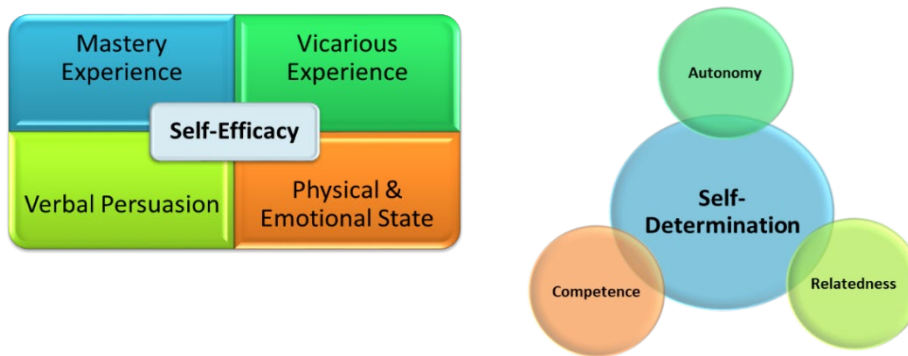
Van Eck et al. [37] proposed a framework for designing educational games that promote problem-solving. Using Jonassen's problem types [38], they mapped the problem types to the gameplay types. Problem types included logical, algorithmic, story, decision-making, troubleshooting, diagnosis-solution, strategic performance, case analysis, design, and dilemma. Gameplay types included adventure, puzzle, action, strategy, role-playing, and simulations. In addition, the possible knowledge and cognitive processes involved in solving each type of problem were provided. This included domain-specific knowledge, higher-order thinking, psychomotor skills, attitude change. The same as the previous serious game design models, the effectiveness of this model was not evaluated.

## **3. Game Design Based on Motivation Theories**

Motivational theories, such as Self-Determination Theory (SDT), tried to explain the main components of motivation and what drives us to action [8]. SDT is a macro-theory framework for studying human motivation; it was first proposed by Deci and Ryan [7], [8], [39]. SDT states that the need for the three factors of relatedness, competence, and autonomy generates intrinsic motivation that can push us to accomplish our goals [40]. Autonomy concerns the sense of free will and being the agent of our own decisions. Competence is the need to be effective and competent in a task. Relatedness is the need to interact with people, feeling attached, or belonging to some groups.

Bandura [41] has defined Self-Efficacy (SET) as “the conviction that one can successfully execute the behavior required to produce the outcomes”. Therefore, in this theory, the self-perceived judgment of one's capabilities is more relevant than the individual's actual capability [42]. Bandura suggested four different self-efficacy sources: Performance Accomplishments, Vicarious Experience, Verbal Persuasion and Emotional Arousal. Performance accomplishment is related to the past successes and failures on a certain task and the perceived capability of accomplishing it. Vicarious experience happens when people observe other people similar to them performing that specific task without too much hardship, this adds to the self-perceived capabilities. In addition, verbal persuasion from others can increase self-efficacy, although the effect might be limited. Finally, emotional arousal relates to the emotional and physiological state of the person in face of a task, and people rely on their emotional arousal state to judge their self-efficacy. The components of these two theories are illustrated in Figure 1. The following sections present the study of the game design based on

the integration of self-determination and self-efficacy theories. The effectiveness of the integrated game-design model was evaluated using a randomized controlled trial.



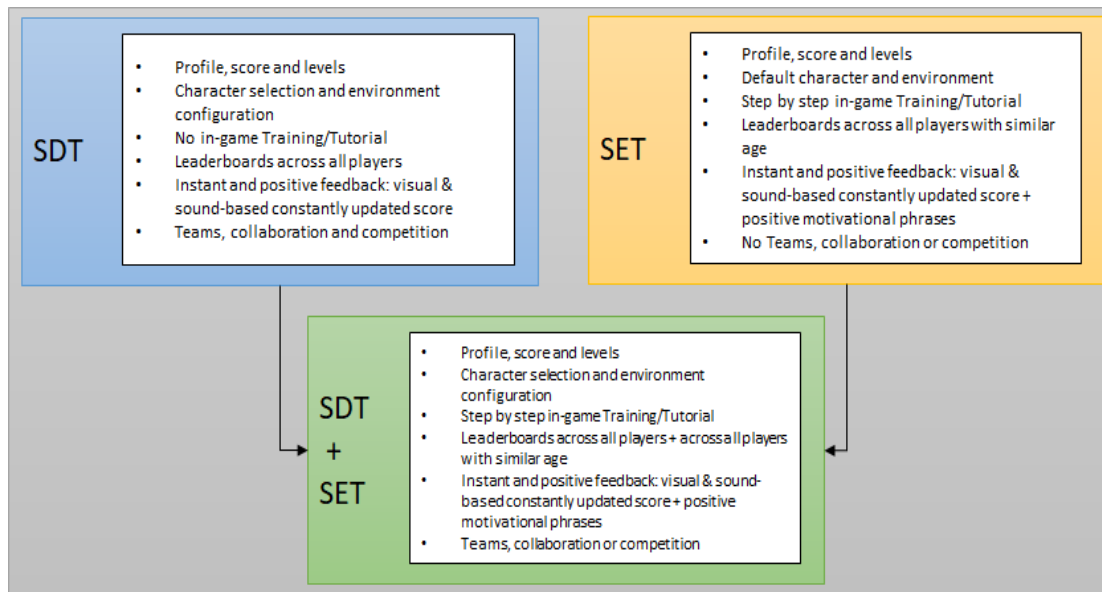
**Figure 1.** The components of self-efficacy and self-determination theories

### 3.1. Integration of Self-Determination and Self-Efficacy in Game Design

Among the main objectives of serious games and gamification is the increase of motivation. Therefore, using motivational theories to design games is rational. Self-determination and self-efficacy are established scientific theories of motivation. However, SDT is a macro theory of human motivation, and its three outlined components are broad [43]. Thus, some researchers have integrated SDT with other psychological theories such as Self-Efficacy Theory (SET) to create a more concrete, practical, and effective theory [41], [44].

Self-efficacy can be compared to the competence part of the self-determination theory. As mentioned previously, competence is the need to be effective and competent in a task. The two concepts seem similar, and both highlight the need for self-belief in one's abilities in a specific task before acting. This has led some researchers to replace SDT's competence component with the four components of SET [44], [45]. The result is the six components of autonomy, relatedness, mastery experience, vicarious experience, verbal persuasion, and emotional arousal.

The gamification model, based on the integration of SDT and SET, classifies each game mechanic into one or multiple of the six components. For example, the personalization option satisfies the need for autonomy, while chat options or team works satisfy relatedness. Figure 2 demonstrates the game mechanics classification into SDT, SET, and the combination of the two theories. In the context of this research and based on the previous psychological findings, it is hypothesized that combining self-determination and self-efficacy could lead to enhanced motivation and better performance. The following research question is stated: Does the integration of self-determination and self-efficacy enhance motivation and performance? For confirming this hypothesis, an experiment, which involved forty-six participants, was designed to evaluate the effectiveness of three game designs. The first two designs are based on either SDT or SET, and the third design is based on the integration of both theories. The following section presents the details of the game design process.

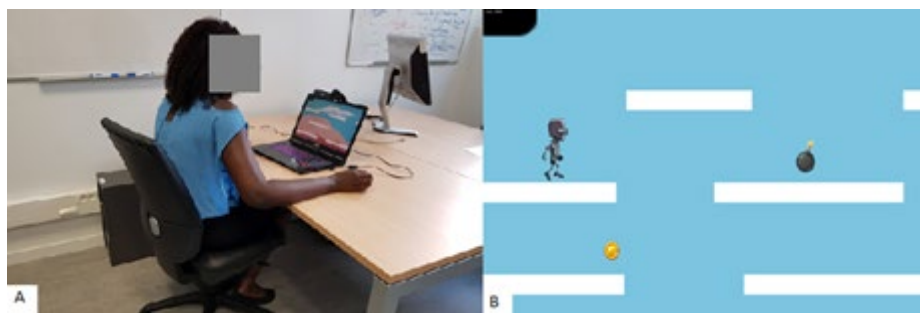


**Figure 2.** Game mechanics implemented for SDT, SET, and the integration of SDT+SET

### 3.2. Game Design Process

A video-game system based on a platformer infinite-running game was developed to study the effects of SDT and SET on player performance and engagement. The player had to collect the maximum points by lasting as much time as possible on the platform. The interaction is limited to pressing one button to jump for avoiding the obstacles and falling from the platform.

The game was adapted from the tutorial “Let’s Make a Game: Infinite Runner” included in Unity Tutorials [23]. As shown in Figure 3, a series of platforms were randomly generated, and the player can jump between them while picking up coins and avoiding bombs. The player’s score is increased based on how much time the user keeps the character alive and the number of coins he/she collects. The game session lasts until a bomb is touched or the character falls to a pit. The player can perform a mid-air jump, but he/she cannot control the force nor modify the character’s starting trajectory. Three different levels of difficulty were presented to the player to select: easy, normal, and hard.



**Figure 3.** A) User playing the game. B) Screenshot of the game

Three different versions of the game were developed. The first design had game mechanics classified for SET. The game mechanics classified for SDT were implemented in the second design. For the third design, the game mechanics relating to both theories were implemented. These three categories represented the three experimental conditions of this study. Figure 2



illustrates the implementation details and the difference between the developed game mechanics, which are explained as follows:

Profiles, score, and levels: These three mechanics were implemented in the same way for all three game versions. To create a profile, the user had to provide a nickname, age, sex, dominant hand, whether he/she wears glasses, indicate his/her prior experience with infinite runner type of videogames. This information was used in the analysis and in the leaderboards. The scores were displayed during gameplay and at the end of each session. The scores increased with the time in which the character was alive and by obtaining coins. Finally, the levels implemented are easy, normal, and hard. The difference between these levels is that the game's speed is increased, and hazards are spawned more often.

Character selection and environment configuration: To satisfy the player's need for autonomy, character selection, and environment configuration mechanics were provided. The player had the possibility to select one of the seven different characters, one of four possible backgrounds, and one of two different songs. However, these personalization options were only aesthetic and did not confer any advantage or disadvantage in the game sessions. These two mechanics were implemented in the SDT and SDT+SET versions of the game but they were not included in the SET version in which the player had to keep the default options.

Training/Tutorials: Tutorials and particularly interactive tutorials are effective in promoting SET's Mastery Experiences. In the SET game version, the player could not proceed to the game before completing the tutorial. The tutorial was divided into five successive stages; each stage taught the player a single gameplay rule. Unlike the game sessions, the tutorial stages had a starting point and a goal. Reaching the goal allowed the access to the next tutorial. After completing the final tutorial stage, the player could play the game session with easy difficulty. Once the user played at least one easy session, he/she could select the normal difficulty. Similarly, when one session with the normal level was played, the user could select the hard difficulty level.

The SDT version of the game did not include the tutorial. In addition, for promoting autonomy, the player could select any preferred difficulty level. The SDT+SET version combines the other two game versions. The tutorial was provided but its completion was not mandatory. Hence, the user can benefit from mastering the game experience without limiting autonomy.

Leaderboards: This mechanic was included to promote SDT's Competence and SET's Vicarious Experiences. In all the experimental conditions, the leaderboards are displayed at the end of each game session. Leaderboards show the name, age, sex, and score of the top players. For SDT game mode, the top ten players for each difficulty level were displayed. For the SET condition, leaderboards were also provided for each difficulty level. However, to promote vicarious experience in the SET group, the leaderboard showed only the high scores of people having the age close to the player's age. The performance in this game depended on reaction times. Older individuals generally have a slower reaction time than younger players. Hence,

comparing older and younger players would not produce the vicarious experience. However, comparing similar-aged people could produce the vicarious experience. A range of three years for each player was considered to identify similar-aged players. The SDT+SET condition showed both leaderboards.

Instant and positive feedback: This mechanic was used to strengthen the SDT's Competence and SET's Social Verbal Persuasion. In all three game modes, instant and positive feedback was provided. The time-based score was displayed in the upper left part of the screen, and when the player picks up a coin, a number (+200) was displayed and added to the score. In order to implement the verbal persuasion, voices of natural spoken speech were recorded, saying positive motivational phrases such as "well done" and "perfect". These voices were randomly played when the player obtained between three and five coins.

Teams, collaboration, and competition: These mechanics were selected to fulfill the player's psychological need for relatedness in the SDT and SDT+SET versions of the game. At the start of the experiment, to foster a sense of belonging, the player had to choose to be part of the yellow or the red team; Belonging to a team allows collaboration with teammates and competing against other teams. Each time a player completed a game session, his/her score was added to his/her team's total score. To create a sense of competition, at the end of each session, the scores of both teams were displayed.

It is important to note, for implementing the leaderboards and team mechanics, previously recorded data of real people were used in the game. Hence, the leaderboards shown to all the players were kept constant. This was to avoid any bias in the experimentation by players trying harder to beat the latest high score.

### **3.3. Experiment Design**

For studying the effect of game design based on motivational theories of self-efficacy and self-determination, an experiment was designed to evaluate three design variations. The first two designs were based on either self-efficacy or self-determination, while the third design was based on the integration of the two theories.

#### **3.3.1. Participants**

Forty-six volunteers participated in this experiment and were assigned randomly, 16 to SDT condition, 15 to SET, and 15 to SDT+SET. The subjects were researchers, university personnel, and students. Fifty-nine percent of the participants were females. The age of the participants ranged between 17 and 70 years (Mean = 30.93, SD = 13.78). Forty-five subjects were right-handed (98 %). Twenty-eight participants wore glasses (61 %). All the participants were asked to rate their prior experience with infinite runner type of video games on a scale ranging from 0 to 10 (where 0 represents no experience and 10 represents a very experienced user). The reported prior experience had a Mean of 3.22, and a standard deviation of 2.83.

### **3.3.2. Procedure**

Three experimental conditions were tested: SDT, SET, SDT+SET. The experiment was conducted individually in equal set-up conditions across the three experimental conditions. All participants were positioned in front of the same laptop in which the game was running. The laptop was a Lenovo B50-80, with a processor of Intel Core i5 2.20GHz, 8GB RAM, Intel HD Graphics 5500 graphics card, 15.6" HD screen.

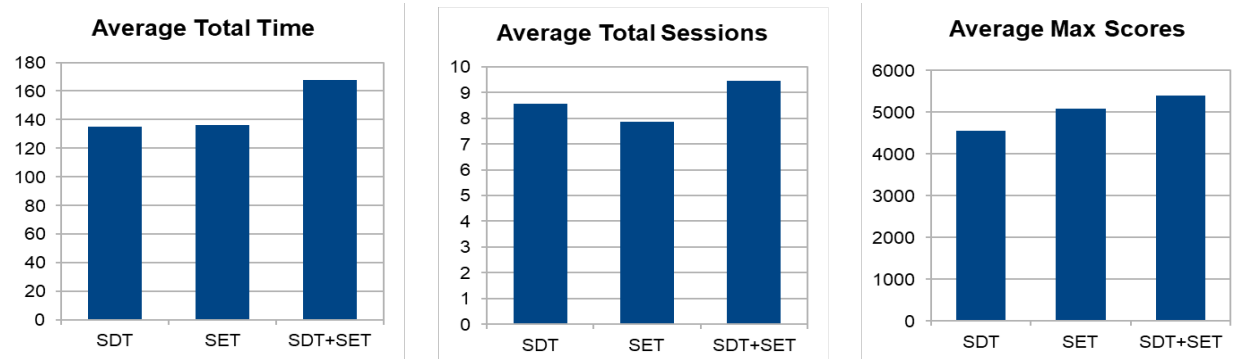
A between-subjects design was adopted for this experimental design. The volunteers arrived according to their availability and they were assigned to the experimental conditions randomly. To avoid the bias caused by completing the experiment in a rush, the subjects were asked to be available for 30 minutes, although the estimated experimentation time for each subject was a maximum of 15 minutes. Each subject was assigned to one of the three experimental conditions randomly. The basic rules of the game were briefly explained to each subject before starting the experiment: "one-click to jump, one-click to perform a mid-air jump, try to collect coins, avoid bombs and pits, and the playing time is not limited". Subsequently, they were assisted in creating their profiles in the game. Then, they played the game until they decided to stop playing.

### **3.3.3. Measures**

In order to evaluate the user performance and the effect of integrating SET and SDT, the following measures were used in this experimental study. The scores of the player was automatically saved and used for analyzing the player performance. The number of played sessions and the total time spent playing in each difficulty level were saved and used to assess the player's engagement. The Situational Motivation Scale (SIMS) was applied to capture the perceived motivation [46]. This scale measured the player's self-perception on Intrinsic Motivation, Identified Regulation, External Regulation, and Amotivation. This test was administered in the form of a pre-test post-test. The Usability of the system was assessed by applying a short version of the System Usability Scale (SUS). This test was applied at post-test [47]. Furthermore, a questionnaire was devised to assess the motivational impact of each game mechanic.

## **3.4. Data Analysis and Results**

The 46 participants formed the three experimental groups of SDT, SET, and SDT+SET with 16, 15, and 15 subjects, respectively. For each subject, a set of subjective and objective data was collected. For each subject, three main outcomes of the objective data were extracted: Total Sessions (N; the sum of the number of sessions played for each difficulty level), Total Time (T; the sum of the duration of times spent in playing each difficulty level), and Max Score (HS; the maximum of recorded high scores throughout all the difficulty levels). The results showed that the SDT+SET condition resulted in higher average of Total Time, Total Session and Max Score. This is illustrated in Figure 4.



**Figure 4.** The difference between the experimental conditions in terms of the average total time, average total sessions and average total max scores

For each of these dependent variables, the normality test of Shapiro-Wilk was conducted. This test showed that the data relating to Total Sessions and Total Time were not normally distributed for all the experimental conditions. However, the data for Max Score was normally distributed. Additionally, Levene's test of homogeneity of variances was conducted for Max Score, and the results confirmed the homogeneity of this data across the different groups ( $p=0.433$ ).

The Shapiro-Wilk test was also applied to the dependent variables based on the independent variables of Age Group, Gender, and Experience Level. The data groups relating to Total Sessions and Total Time were not normally distributed but the data groups relating to Max Score followed the normal distribution. Hence, for the normal data, One-Way ANOVA and the Post-Hoc test of Scheffe were conducted. The non-parametric test of Kruskal-Wallis was applied for the data not normally distributed.

The results of the Kruskal-Wallis test on the variables of Total Sessions and Total Time between the three experimental groups (SDT, SET, and SDT+SET) were not statistically significant ( $H(2) = 0.422$ ,  $p = 0.810$  for Total Sessions,  $H(2) = 0.666$ ,  $p = 0.717$  for Total Time). In addition, the result of One-Way ANOVA on Max Score among the three experimental conditions was not statistically significant ( $F(2,45) = 0.987$  and  $p = 0.381$ ). Moreover, One-Way ANOVA and Kruskal-Wallis tests were performed on each dependent variable categorized based on Age Group, Gender, and Experience Level. The result of the One-Way ANOVA test on Max Score categorized based on the Experience Level was statistically significant ( $F(2, 45) = 8.713$  and  $p = 0.001$ ). This indicated a reasonable expected positive relationship between the prior experience in this type of game and the user performance. However, no significant effect of the Gender and Age group was found on the three dependent variables. These results are summarized in Table 1.

**Table 1.** Results of the statistical analysis

| Variables          | Normality Test | Analysis                        | Results                         |
|--------------------|----------------|---------------------------------|---------------------------------|
| Total Sessions     | Shapiro-Wilk   | Kruskal-Wallis                  | $H(2) = 0.422$ , $p = 0.810$    |
| Total Time         | Shapiro-Wilk   | Kruskal-Wallis                  | $H(2) = 0.666$ , $p = 0.717$    |
| Max Score          | Shapiro-Wilk   | One-Way ANOVA                   | $F(2,45) = 0.987$ , $p = 0.381$ |
| Pass (Significant) |                | Does Not Pass (Not Significant) |                                 |

The results from pre-test post-test questionnaires were extracted for the variables of Intrinsic Motivation, Identified Regulation, External Regulation, and Amotivation. The difference between pre-test and post-test was calculated, and the outcome was tested for normality and homogeneity. The variable of External Regulation did not pass the Shapiro-Wilk test for normality. Hence, a One-Way ANOVA test was conducted for the other variables, and the results did not indicate any significant difference among the three experimental conditions: Intrinsic Motivation  $F(2,45)=0.243$  and  $p=0.785$ , Identified Regulation  $F(2,45)=0.356$  and  $p=0.703$ , and Amotivation  $F(2,45)=0.080$  and  $p=0.923$ . Furthermore, the non-parametric test of Kruskal-Wallis was applied for the External Regulation data. The results did not show any significant difference among the three experimental conditions ( $H(2)=0.716$  and  $p=0.699$ ). Finally, the Paired-Samples T-Test was conducted to analyze the difference in intrinsic motivation before and after the participation to the experiment. This test was carried out separately for each of the experimental condition, and the results showed a significant difference only for the SET group ( $t(14)=-2.559$  and  $p=0.023$ ). Cohen's  $d$  effect size value ( $d=0.66$ ) suggested a moderate to high effect size.

The usability was assessed with the short post-test of SUS. The used scale ranged from one to five. When the participants were asked if they would like to use the system frequently, if it is available, they scored an average of 3.3. When they were asked if they think that the game was easy to use, they produced an average score of 4.39. When the participants were asked if they think that most users will learn quickly to use the game, they answered with an average score of 4.60. Finally, a score of 3.86 was given when they were asked if they felt confident when using the game.

Finally, a questionnaire was used to assess which game mechanics were perceived as more motivating. The scale ranged from -3 (Very demotivating) to 3 (very motivating). The game mechanics that were perceived as more motivating were: watching the high score in the player's range of age (mean=2.36), being able to choose between doing and not doing the tutorial (mean=2.2), and being able to select your character (mean=2.12). Being forced to complete the tutorial before playing the game was the least motivating game mechanic (mean=0.66).

The results did not confirm the hypothesis that integrating self-determination and self-efficacy in game design would enhance the motivation and performance. However, this does not imply that the hypothesis is not plausible. Specific attributes of the experiment design in this research (e.g., type of game, game design, and environment of the experiment) might have contributed to these results. Thus, additional investigation and experimentation are necessary to explore the effect of SET and SDT on motivation and the user performance in the gamified content.

## 4. Game Design for Automaticity Acquisition

Training design for automatic skills has a vast application domain, such as education, physical and cognitive rehabilitation, sports, arts, and professional training. Motivation is a key part of any training program. Particularly, for automaticity acquisition, going through the countless repetitions requires a high degree of motivation. Serious games, and gamification as methods to increase engagement and adherence, can be effective in learning and skill acquisition [48]–[51].

The process of automaticity acquisition goes through phases. Current game design models did not take into account the temporal specificity of the game elements for automaticity acquisition training. For addressing this issue, the following section presents the stage-based game design model for automaticity acquisition. This model is designed based on the principles of automaticity acquisition presented in the previous chapter. This game design model classifies the game mechanics and maps them onto each phase of automaticity acquisition. An experimental study is designed to evaluate this model's effectiveness by involving 49 participants in a randomized controlled trial.

### 4.1. Automaticity Gamification Model

Automaticity acquisition is a temporally sensitive process. It evolves through phases [52]–[54]. There is currently no game design model for the acquisition of automaticity. Hence, there is no game design model that is temporally sensitive to the phases of automaticity acquisition. A *stage-based gamification model for the acquisition of automaticity* is proposed in this section. It classifies the game elements into abstract categories and maps them onto the phases of automaticity acquisition. This model is designed based on the training model for automaticity acquisition presented in the previous chapter.

The purpose of this game design model is to provide a flexible basis for designing serious games or gamified approaches that target automaticity acquisition. This gamification model is presented in Table 2. The main contribution of this model is taking into account the temporal-specificity of the game elements during the three phases of automaticity acquisition. It is built upon the second step of the automaticity training model presented in the previous chapter. As illustrated in the table below, the implementation of some game mechanics can overlap between two or all three stages. In addition, these game mechanics are classified into seven abstract categories of feedback, goals, challenges, achievements and rewards, help, identity, and social interaction. The game mechanics in this table have been selected from various sources in the literature, notably [21], [24], [55]–[60].

**Table 2.** Stage-based gamification model for the acquisition of automaticity

| Abstract Categories      | Fast Learning  | Slow Learning  | Automatization   |
|--------------------------|--|--|--|
| Feedback                 | Process Feedback - Frequent Feedback   | Outcome Feedback - Intermittent Feedback               | Summary Feedback - Event-triggered Feedback - Macro Leaderboards - Assessment    |
|                          | Micro Leaderboards   |  |  |
| Goals                    | Short-term Goals   | Mid-term Goals   | Long-term Goals  |
|                          |  | Checkpoints  |  |
|                          | Progress Bars - Checklist  |  |  |
| Challenge                | -  | Small Challenges                                       | Big Challenges (Boss Fight) – Combo – Disincentives - Intentional Loss (Betting) |
|                          |  | Countdown - Action Points - Self-Competition (Records) |  |
| Achievements and Rewards | Continuous Rewards   | Frequent Variable Rewards                              | Infrequent Variable Rewards - Unlocking  |
|                          |  | Lottery /Chance/Dice                                   |  |
|                          | Experience Points - Level up - Scheduled Rewards - Physical Rewards - Badges/Trophies – Tokens - Redeemable Points |  |  |
| Help                     | Cascading Information Theory (Tutorial/Cinematic/Cutscenes) - Beginner’s Luck - Safe Haven                         | Catch-up/Free Lunch                                    | -  |
|                          | Extra Lives - Help (Hints & Cues/ Warning Messages)  |  |  |
| Identity                 | Meaning (Backstory), Avatar  | Meaning (Narrative/Plot)                               |  |
|                          | Loyalty - Customization/Tailoring - Status   |  |  |
| Social Interaction       | Mentorship   | Teams - Team Rewards - Cheer up - Raids                |  |
|                          | Friends - Sharing (Bragging) - Conversation - Gifts  |  |  |

## 4.2. Game Design Process

This section presents an experimental study for evaluating the effectiveness of the proposed stage-based gamification model. An off-the-shelf video game was selected and redesigned based on the proposed model. An experimental study involving 49 subjects was conducted to determine the effects of the model on automaticity acquisition. In this game, the player has to control a character that is constantly flying forward while being pulled down by a gravitational force (Figure 5). The player has to tap on the screen to make the character flap its wings, lifting it slightly. The objective of the game is to fly as far as possible. However, to make this task

more challenging, some obstacles appear in the form of tubes with openings at varying heights. The game ends when the character touches an obstacle, the stage ceiling, or the ground. When the game is over, the score of the player is displayed. The score represents the total distance traveled by the character in kilometers. The control group used the original design of the game, and the experimental group used the redesigned game based on the stage-based gamification model. In the experimental design, a set of key game mechanics was selected and implemented according to automaticity acquisition phases. Each of these game mechanics is explained below.



**Figure 5.** Screenshot of the game: The character is shown inside the yellow circle, and the obstacles are inside the red rectangles

Experience points and levels: Based on the number of kilometers traveled by the character, the player is rewarded with experience points. The player's level is increased based on the experience points. The level and experience points are displayed on the main menu, and they represent the achievements and the mastery level of the game. In addition, each level unlocked a list of advantages to help through the slow and fast learning phases. These advantages were available from level one to level six. The experience points and levels were used to implement the principles of reinforcement and feedback. During the fast learning phase, fewer experience points were necessary to level up, providing a high reinforcement frequency. During the slow learning phase, leveling up required more experience points. Hence, the reinforcement became less frequent. Finally, in the automatization phase, leveling up required the highest experience points, which result in infrequent reinforcement.

In-game help: In-game parameters relating to game speed and distance between obstacles were modified to make the game easier. The adaptation of the game speed was based on the principle of time pressure. During the fast learning phase, the game was slower, but the game's speed increased as the user progressed. The principle of fidelity was implemented through the modifications of the in-game speed and the distance between obstacles. During the fast and slow learning phases, the system maintained low fidelity by reducing the game speed and increasing the distance between the obstacles. During the automatization phase, these parameters were set to achieve a high level of fidelity. Conversely, the control group design was always under the high-fidelity condition. Finally, a visual cue aid was integrated to implement the principles of explicit instruction and feedback. During the fast and slow learning phases, this visual aid explicitly instructed the player when he/she had to tap the screen to avoid



the obstacles. It acted as concurrent and high-frequency feedback since it constantly informed the player whenever the character's flight trajectory was not correct.

Extra Lives: From level one to five, the player had the possibility to earn extra lives; this allowed the player to touch an obstacle without triggering a game over. If this extra life was used, it could only be recovered by passing successfully through an increasing number of obstacles based on the player's level. Lives were implemented in the game following the principle of fidelity. During the fast learning phase, fidelity should be low; hence, lives were provided from the beginning and then easily recovered. In the slow learning phase, extra lives were still present, but the player had to play more to earn them. When the player reached the automatization phase, the fidelity was high; therefore, extra lives were no longer provided. Since earning the extra lives was rewarding, they also acted as reinforcement. In line with the model, they were earned frequently during the fast learning phase, and as the game progresses, they became infrequent.

Milestones: The game provides the player with goals and challenges. These milestones consist of reaching a fixed number of traveled kilometers depending on the player's level. A progress bar indicated how close the player was to complete a milestone. The milestones mechanic implemented the principle of goals in the proposed model. During the fast learning phase, goals should be proximal, requiring a low number of kilometers to be completed. During the slow learning phase, the number of kilometers increased as the goals became gradually distal. Finally, in the automatization phase, all the goals were distal by default.

Leaderboards: At a game over, the game displays the traveled distance for the session, their longest flight, and leaderboards. Leaderboards showed the high scores of players of the same level. This was implemented as a form of micro-leaderboard. In the automatization phase, the macro leaderboard was added to the micro leader board by presenting the all-time high scores. Leaderboards could also act as both feedback and reinforcement. In terms of feedback, they acted as terminal feedback, and in terms of reinforcement, they reinforced the user's playing behavior by instigating a sense of competitiveness. At the fast and slow learning phase, the all-time high scores were not shown to keep the users from demotivation. However, during the automatization phase, the player had already reached a mastery level, which produces enough self-efficacy to compete with the highest scores.

Personalization: The users had the possibility to customize the character by choosing its in-game appearance. Three options were available: a crow, a helicopter, and a balloon. This mechanic corresponds to the identity section of the game design model.

### **4.3. Experiment Design**

A between-subjects experiment was performed to study the effects of the proposed stage-based gamification model. A control group and an experimental group were involved in the experiment by playing the game. The control group tested the version of the game that included the original design. The experimental group used the redesigned game based on the stage-based

gamification model. Both versions of the game were implemented in a single Android application. It was freely distributed worldwide. Subjects from Belgium, Brazil, Chile, France, India, Iran, Mexico, Netherlands, Russia, and the USA were invited to participate in this experiment as volunteers.

#### 4.3.1. Participants

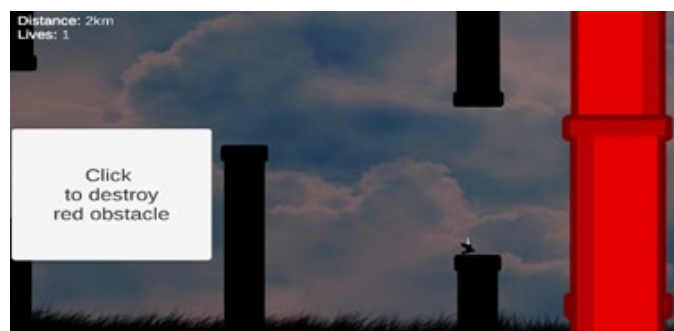
Forty-nine volunteers participated in the study; each participant was assigned randomly by the system to the experimental group or the control group. Hence, 25 subjects were assigned to the experimental group, and 24 subjects were assigned to the control group. They were 71% male and 29% female, and 88% were right-handed and 12% left-handed. The mean age was 28.2 (SD = 7.6) for the experimental group and 29.8 (SD = 5.9) for the control group. Prior experience with video gaming for the experimental group had a mean of 5.08 (SD = 2.6) and a mean of 5.87 (SD = 2.6) for the control group.

#### 4.3.2. Procedure and Measures

To participate in the experiment, the subjects were asked for their agreement about the declaration of consent displayed in the game. After consenting, they had to create an in-game username and provide information about their age, gender, experience playing video games, and the dominant hand. This process created an account linked to the player's device ID. Finally, the players were randomly assigned to either the control or the experimental group.

The participants were requested to play the game at least once per day for seven consecutive days. After playing the game for seven days, another challenge was added to the game, and the subjects were invited to play the game at least one more time. This new game challenge was a multi-tasking challenge for testing automaticity acquisition among the participants (Figure 6). It is an additional obstacle that can only be eliminated by pressing a button appearing on the left side of the screen. This new challenge should be performed simultaneously with the main task of controlling the character, for this reason it is called a multi-tasking challenge.

For simplifying the data collection, the data relating to the user performance and engagement were recorded automatically in the game and transferred to the cloud. Finally, an in-game button was made available for answering to a questionnaire. The questionnaire collected information about the system usability and user's motivation, as measured by the system usability scale (SUS) [47] and the situational motivation scale (SIMS) [46].



**Figure 6.** Screenshot of the game with the introduction of the seventh-day challenge: The new obstacle in red color (right), the button used to destroy the obstacle (left)

#### 4.4. Data Analysis and Results

The objective and subjective data were collected from this experiment. The objective data included information relating to the user performance. The data did not pass the normality test of Shapiro-Wilk; therefore, the non-parametric test of Kruskal-Wallis was performed to test the significance of the difference between the two groups. The average number of sessions played (N) by the experimental group was  $N_E = 71.8$  (SD = 71), as opposed to the control group with the average of  $N_C = 26.2$  (SD = 29.4). The result of the statistical analysis of Kruskal-Wallis showed this significant difference,  $\chi^2(1) = 11.22$ ,  $p = 0.0008$ , between the experimental and control groups. This represents an important finding showing that the experimental group played the game more than the control group. This shows that the game design based on the proposed model allowed better user engagement. This is also showed by the average of total time spent playing the game (T):  $T_E = 1992.9$  seconds (SD = 2658.5) for the experimental group and  $T_C = 458.58$  (SD = 720) for the control group with a significant difference of playing time:  $\chi^2(1) = 15.44$ ,  $p = 0.00008$ .

In general, a higher level of engagement should result in better performance. This is confirmed by the results from the high scores (Maximum of the distance flown in one session) and total distance (sum of the distances of all the sessions). The average of high score (HS) for the experimental group is  $HS_E = 57.2$  (SD = 65.7) in comparison to  $HS_C = 17.8$  (SD = 26.4) for the control group. The statistical analysis showed a significant difference between the two groups:  $\chi^2(1) = 15.80$ ,  $p = 0.00007$ . The same result was found for the total distance (D) with an average of  $D_E = 621.6$  (SD = 844.9) for the experimental group, and  $D_C = 135$  (SD = 221.9) for the control group, with a significant difference given by Kruskal-Wallis test:  $\chi^2(1) = 16.08$ ,  $p = 0.00006$ .

From the 49 participants, 24 played for at least seven days, and therefore, they reached the multi-tasking challenge. Out of these 24 participants, 16 were from the experimental group, and eight participants from the control group. The double dropout in the control group showed that the game design is important for keeping high user engagement. The design based on the proposed model in this study led to higher engagement, attracting participants to make more repetitions. This is expected to generate a higher level of automaticity. To ascertain this assumption, the data collected after introducing the multi-tasking challenge was analyzed. At this design stage, the game's difficulty level is identical for all users, and the multi-tasking challenge could be a good indicator of automaticity. For the analysis of automaticity acquisition, the data of the dropped-out users were excluded, and the data of the 24 remaining participants were analyzed.

The first variable to explore is the number of sessions (N) played after introducing the multi-tasking challenge. This was on average  $N_E = 18.2$  (SD = 14.7) for the experimental group, and  $N_C = 12.6$  (SD = 12) for the control group. The users in the experimental group played more on average. However, the Kruskal-Wallis test showed that this difference was not statistically significant:  $\chi^2(1) = 1.028$ ,  $p = 0.311$ . Regarding the performance after introducing the multi-tasking challenge, the high score and total distance data showed an advantage for the experimental group. For the high score (HS), the experimental group obtained the mean  $HS_E = 33.2$  (SD = 25.4) and  $HS_C = 7.2$  (SD = 8.3) for the control group. This difference is also

statically significant:  $\chi^2(1) = 5.794$ ,  $p = 0.016$ . For the total flew distance (D) after the introduction of the multi-tasking challenge, the average for the experimental group was  $D_E = 161.5$  ( $SD = 141$ ) as opposed to  $D_C = 42.1$  ( $SD = 59$ ). Kruskal-Wallis test showed that this difference was relatively significant:  $\chi^2(1) = 3.607$ ,  $p = 0.058$ .

The experimental group showed a clear advantage in the performance of the multi-tasking challenge. However, this difference might be caused by the fact that they played more times after introducing the challenge. To investigate this issue, the data of the total distance (after the introduction of the multi-tasking challenge) was divided by the number of played sessions. This resulted in the average distance per session (AD) for each user. The mean of this value for the control group is  $AD_C = 2.5$  ( $SD = 1.5$ ) and  $AD_E = 9.0$  ( $SD = 5.9$ ) for the experimental group. Finally, the statistical analysis proved that this difference was significant between the two groups:  $\chi^2(1) = 6.010$ ,  $p = 0.014$ . This result showed that when the challenge was introduced in the game, the experimental group users were more skilled in playing and progressed more towards the acquisition of automaticity.

Furthermore, the subjective data collected from the questionnaires confirmed the same findings. The result of the usability (U) test was  $U_E = 73.3$  for the game design based on the stage-based gamification model, and  $U_C = 66.8$  for the initial game design. The situational motivation scale results also showed that Intrinsic Motivation (IM) was higher for the experimental group with  $IM_E = 79.2\%$  and  $IM_C = 61.6\%$  for the control group. The Identified Regulation (IR) was  $IR_E = 83.8\%$  for the experimental group and  $IR_C = 58.3\%$  for the control group. The External Regulation (ER) was  $ER_E = 76.1\%$  for the experimental group and  $ER_C = 78.3$ , for the control group. Finally, the Amotivation (AM) was  $AM_E = 46.9\%$  for the experimental group and  $AM_C = 48.3\%$  for the control group. However, no effect of gender nor dominant hand was observed but the prior experience in video gaming showed a significant difference in the overall high score:  $\chi^2(2) = 6.543$ ,  $p = 0.038$ .

## 5. Conclusion

The acquisition of automaticity in reading requires both intensive and extensive practice. Hence, adherence and engagement to the reading program are key issues. In the case of technology-based in-home programs, the need for motivational aspects is crucial. Serious game and gamification concepts are emerged to exploit game elements for increasing motivation in non-game contexts. With game design being partly arbitrary and partly artistic, researchers have tried to model serious game design inside the fame of scientific theories.

This chapter presented the current state of the art on the gamification and serious game models for learning. The important weakness of all the reviewed models was the lack of evaluation and validation through experimental studies. Moreover, important motivational theories such as self-efficacy and self-determination were not included in the game design models. Furthermore, the learning models were too broad, and did not differentiate learning categories such as declarative knowledge and automaticity acquisition.

Based on these identified gaps, two gamification models were proposed in this chapter. The first model was based on the integration of self-efficacy and self-determination theories. The second model was based on the training model for automaticity acquisition presented in the

previous chapter. It mapped game mechanics onto the three phases of automaticity acquisition. For evaluating the effectiveness of the two proposed gamification models, two experimental studies were conducted. Off-the-shelf games were selected and designed based on the proposed models for both studies.

In the first experimental study, three versions of a game were designed. The first two designs were based on either self-efficacy theory or self-determination theory, and the third design was based on integrating the two theories. These three game designs were studied with 46 volunteers who participated in the study. The statistical analysis revealed that there was not any significant difference between the groups in terms of the maximum score, number of sessions, and total time spent playing the game. These results did not confirm the hypothesis considering that integrating self-determination and self-efficacy would lead to enhanced performance and engagement.

Furthermore, the analysis of the feedback obtained from the SIMS questionnaire did not reveal any significant difference among the three experimental conditions. However, a significant difference was found between the pre-test and post-test questionnaires of SET condition on Intrinsic Motivation. This effect was not observed in the other experimental conditions. Finally, the participants perceived the developed system as easy to use. The most motivating feature was comparing high scores between players of the same range of age; the least motivating feature was being forced to complete the tutorial before playing the game.

In the second experimental study, the game redesigned based on the stage-based gamification model was compared to its original design. Forty-nine volunteers participated in this experiment. The data analysis showed that the subjects who played the game based on the stage-based gamification model had significantly higher engagement and performance. Similar results were found for the multi-tasking challenge integrated into the game for evaluating the automaticity level. The results from the system usability scale and the situational motivation scale showed a similar trend. The participants attributed a higher score to the game designed based on the gamification model in both usability and intrinsic motivation. These results validated the effectiveness of the proposed stage-based gamification model for the acquisition of automaticity.

## 6. References

- [1] J. P. Gee, "What video games have to teach us about learning and literacy," *Comput. Entertain. CIE*, vol. 1, no. 1, pp. 20–20, 2003.
- [2] R. M. Ryan, C. S. Rigby, and A. Przybylski, "The motivational pull of video games: A self-determination theory approach," *Motiv. Emot.*, vol. 30, no. 4, pp. 344–360, 2006.
- [3] S. Deterding, M. Sicart, L. Nacke, K. O'Hara, and D. Dixon, "Gamification. using game-design elements in non-gaming contexts," in *CHI'11 extended abstracts on human factors in computing systems*, 2011, pp. 2425–2428.
- [4] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: defining 'gamification,'" in *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11*, Tampere, Finland, 2011, p. 9, doi: 10.1145/2181037.2181040.
- [5] A. Bandura, "Self-efficacy," in *Encyclopedia of human behavior*, V. S. Ramachaudran (Ed.), vol. 4, New York: Academic Press, 1994, pp. 71–81.

- [6] A. Bandura, "Self-Efficacy," in *The Corsini Encyclopedia of Psychology*, I. B. Weiner and W. E. Craighead, Eds. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2010, pp. 1–3.
- [7] E. L. Deci and R. M. Ryan, "The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior," *Psychol. Inq.*, vol. 11, no. 4, pp. 227–268, 2000.
- [8] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being," *Am. Psychol.*, vol. 55, no. 1, p. 68, 2000.
- [9] A. Amory, K. Naicker, J. Vincent, and C. Adams, "The use of computer games as an educational tool: identification of appropriate game types and game elements," *Br. J. Educ. Technol.*, vol. 30, no. 4, pp. 311–321, Oct. 1999, doi: 10.1111/1467-8535.00121.
- [10] A. Amory, "Building an educational adventure game: Theory, design, and lessons," *J. Interact. Learn. Res.*, vol. 12, no. 2, pp. 249–263, 2001.
- [11] A. Amory, "Game object model version II: a theoretical framework for educational game development," *Educ. Technol. Res. Dev.*, vol. 55, no. 1, pp. 51–77, Jan. 2007, doi: 10.1007/s11423-006-9001-x.
- [12] K. Kiili, "Digital game-based learning: Towards an experiential gaming model," *Internet High. Educ.*, vol. 8, no. 1, pp. 13–24, Jan. 2005, doi: 10.1016/j.iheduc.2004.12.001.
- [13] D. A. Kolb, *Experiential learning: Experience as the source of learning and development*. .
- [14] M. Csikszentmihalyi, *Finding flow: The psychology of engagement with everyday life*. Basic Books, 1997.
- [15] M. Csikszentmihalyi, *Flow: The psychology of happiness*. Random House, 2013.
- [16] P. Rooney, "A Theoretical Framework for Serious Game Design: Exploring Pedagogy, Play and Fidelity and their Implications for the Design Process," *Int. J. Game-Based Learn.*, vol. 2, no. 4, pp. 41–60, Oct. 2012, doi: 10.4018/ijgbl.2012100103.
- [17] L. P. Steffe and J. E. Gale, *Constructivism in education*. Lawrence Erlbaum Hillsdale, NJ, 1995.
- [18] N. Suttie *et al.*, "In Pursuit of a 'Serious Games Mechanics': A Theoretical Framework to Analyse Relationships Between 'Game' and 'Pedagogical Aspects' of Serious Games," *Procedia Comput. Sci.*, vol. 15, pp. 314–315, 2012, doi: 10.1016/j.procs.2012.10.091.
- [19] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, "Taxonomy of educational objectives: the classification of educational goals. Handbook I: cognitive domain. New York: David McKay Company," *Inc7th Ed*. 1972, 1956.
- [20] L. W. Anderson *et al.*, "A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, abridged edition," *White Plains NY Longman*, 2001.
- [21] S. Arnab *et al.*, "Mapping learning and game mechanics for serious games analysis: Mapping learning and game mechanics," *Br. J. Educ. Technol.*, vol. 46, no. 2, pp. 391–411, Mar. 2015, doi: 10.1111/bjet.12113.
- [22] P. M. Kato, S. W. Cole, A. S. Bradlyn, and B. H. Pollock, "A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial," *Pediatrics*, vol. 122, no. 2, pp. e305–e317, 2008.
- [23] M. B. Carvalho *et al.*, "An activity theory-based model for serious games analysis and conceptual design," *Comput. Educ.*, vol. 87, pp. 166–181, Sep. 2015, doi: 10.1016/j.compedu.2015.03.023.
- [24] M. B. Carvalho, "Serious games for learning: a model and a reference architecture for efficient game development," Eindhoven University of Technology - University of Genoa, 2017.
- [25] Y. Engeström, "Expansive Learning at Work: Toward an activity theoretical reconceptualization," *J. Educ. Work*, vol. 14, no. 1, pp. 133–156, Feb. 2001, doi: 10.1080/13639080020028747.
- [26] M. Callaghan, N. McShane, A. G. Eguíluz, and M. Savin-Baden, "Extending the Activity Theory Based Model for Serious Games Design in Engineering to Integrate Analytics," *Int. J. Eng. Pedagogy IJEP*, vol. 8, no. 1, p. 109, Feb. 2018, doi: 10.3991/ijep.v8i1.8087.
- [27] S. de Freitas and M. Oliver, "How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?," *Comput. Educ.*, vol. 46, no. 3, pp. 249–264, Apr. 2006, doi: 10.1016/j.compedu.2005.11.007.
- [28] G. A. Gunter, R. F. Kenny, and E. H. Vick, "Taking educational games seriously: using the RETAIN model to design endogenous fantasy into standalone educational games," *Educ. Technol. Res. Dev.*, vol. 56, no. 5–6, pp. 511–537, Dec. 2008, doi: 10.1007/s11423-007-9073-2.

- [29] Y. Chou, *Actionable Gamification: Beyond points, badges and leaderboards*. Packt Publishing Ltd, 2019.
- [30] L. A. Annetta, "The 'I's' Have It: A Framework for Serious Educational Game Design," *Rev. Gen. Psychol.*, vol. 14, no. 2, pp. 105–113, Jun. 2010, doi: 10.1037/a0018985.
- [31] K. Starks, "Cognitive behavioral game design: a unified model for designing serious games," *Front. Psychol.*, vol. 5, 2014, doi: 10.3389/fpsyg.2014.00028.
- [32] "Toward a Psychology of Human Agency - Albert Bandura, 2006." <https://journals.sagepub.com/doi/10.1111/j.1745-6916.2006.00011.x> (accessed May 23, 2019).
- [33] H. Gardner, *Frames of mind: The theory of multiple intelligences*. Hachette UK, 2011.
- [34] V. Vanden Abeele *et al.*, "P-III: A Player-Centered, Iterative, Interdisciplinary and Integrated Framework for Serious Game Design and Development," in *Serious Games: The Challenge*, vol. 280, S. De Wannemacker, S. Vandercruysse, and G. Clarebout, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 82–86.
- [35] A. Brennecke and H. Schumann, "A general framework for digital game-based training systems," University of Rostock, 2009.
- [36] A. Brennecke, S. Schlechtweg, and H. Schumann, "Game-Based Training in an Interactive Environment by means of OpenCrimeScene.," in *TPCG*, 2008, pp. 167–170.
- [37] R. Van Eck and W. Hung, "A Taxonomy and framework for designing educational games to promote problem solving," p. 19.
- [38] D. H. Jonassen and W. Hung, "All problems are not equal: Implications for problem-based learning," *Essent. Read. Probl.-Based Learn.*, pp. 7–41, 2015.
- [39] E. L. Deci and R. M. Ryan, *Intrinsic Motivation and Self-Determination in Human Behavior*. Boston, MA: Springer US, 1985.
- [40] E. L. Deci and R. M. Ryan, "Overview of self-determination theory: An organismic dialectical perspective," *Handb. Self-Determ. Res.*, pp. 3–33, 2002.
- [41] A. Bandura, "Self-efficacy: Toward a unifying theory of behavioral change," *Psychol. Rev.*, vol. 84, no. 2, pp. 191–215, 1977, doi: 10.1037/0033-295X.84.2.191.
- [42] M. Bong and R. E. Clark, "Comparison between self-concept and self-efficacy in academic motivation research," *Educ. Psychol.*, vol. 34, no. 3, pp. 139–153, Jun. 1999, doi: 10.1207/s15326985ep3403\_1.
- [43] E. L. Deci and R. M. Ryan, "Self-determination theory: A macrotheory of human motivation, development, and health.," *Can. Psychol. Can.*, vol. 49, no. 3, p. 182, 2008.
- [44] S. N. Sweet, M. S. Fortier, S. M. Strachan, and C. M. Blanchard, "Testing and integrating self-determination theory and self-efficacy theory in a physical activity context.," *Can. Psychol. Can.*, vol. 53, no. 4, p. 319, 2012.
- [45] A. A. Techatassanasoontorn and A. Tanvisuth, "The integrated self-determination and self-efficacy theories of ICT training and use: the case of the socio-economically disadvantaged," *GlobDev 2008*, p. 19, 2008.
- [46] F. Guay, R. J. Vallerand, and C. Blanchard, "On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS)," *Motiv. Emot.*, vol. 24, no. 3, pp. 175–213, 2000, doi: 10.1023/A:1005614228250.
- [47] J. Brooke, "SUS-A quick and dirty usability scale," in *Usability evaluation in industry*, London: Taylor & Francis, 1996, pp. 189–194.
- [48] M. Qian and K. R. Clark, "Game-based Learning and 21st century skills: A review of recent research," *Comput. Hum. Behav.*, vol. 63, pp. 50–58, Oct. 2016, doi: 10.1016/j.chb.2016.05.023.
- [49] A. I. Abdul Jabbar and P. Felicia, "Gameplay Engagement and Learning in Game-Based Learning: A Systematic Review," *Rev. Educ. Res.*, vol. 85, no. 4, pp. 740–779, Dec. 2015, doi: 10.3102/0034654315577210.
- [50] M. Lister, "Gamification: The effect on student motivation and performance at the post-secondary level," *Issues Trends Educ. Technol.*, vol. 3, no. 2, pp. 1–22, Dec. 2015.
- [51] M. Sailer and L. Homner, "The Gamification of Learning: a Meta-analysis," *Educ. Psychol. Rev.*, pp. 1–36, Aug. 2019, doi: 10.1007/s10648-019-09498-w.

- [52] J. Doyon and H. Benali, "Reorganization and plasticity in the adult brain during learning of motor skills," *Curr. Opin. Neurobiol.*, vol. 15, no. 2, pp. 161–167, Apr. 2005, doi: 10.1016/j.conb.2005.03.004.
- [53] P. M. Fitts and M. I. Posner, *Human performance*. Oxford, England: Brooks/Cole, 1967.
- [54] D. M. Neves and J. R. Anderson, "Knowledge compilation: Mechanisms for the automatization of cognitive skills," *Cogn. Sci. Their Acquis.*, pp. 86–102, 1981.
- [55] Y.-K. Chou, *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. Octalysis Media, 2016.
- [56] K. Werbach and D. Hunter, *For the Win: How Game Thinking Can Revolutionize Your Business*. Wharton Digital Press, 2012.
- [57] J. M. Kumar and M. Herger, *Gamification at Work: Designing Engaging Business Software*. Interaction Design Foundation, 2013.
- [58] K. Seaborn and D. I. Fels, "Gamification in theory and action: A survey," *Int. J. Hum.-Comput. Stud.*, vol. 74, pp. 14–31, Feb. 2015, doi: 10.1016/j.ijhcs.2014.09.006.
- [59] J.-N. Proulx, M. Romero, and S. Arnab, "Learning Mechanics and Game Mechanics Under the Perspective of Self-Determination Theory to Foster Motivation in Digital Game Based Learning," *Simul. Gaming*, vol. 48, no. 1, pp. 81–97, Feb. 2017, doi: 10.1177/1046878116674399.
- [60] D. Holmes, D. Charles, P. Morrow, S. McClean, and S. McDonough, "Rehabilitation Game Model for Personalised Exercise," in *2015 International Conference on Interactive Technologies and Games*, Oct. 2015, pp. 41–48, doi: 10.1109/iTAG.2015.11.





## Contents

|   |            |
|---|------------|
| <b>CHAPTER VI Validation of the Individualized Reading System .....</b> | <b>118</b> |
| <b>1. Introduction .....</b>  | <b>118</b> |
| <b>2. The Implementation of the Proposed Approach .....</b>             | <b>119</b> |
| 2.1. The Architecture of the System.....                                | 119        |
| 2.2. Reading Tasks .....  | 120        |
| 2.3. Serious Game Design .....  | 122        |
| 2.4. Individualized Intelligent Module.....                             | 126        |
| 2.5. Threshold Adaptivity.....  | 130        |
| 2.6. Development and Implementation .....                               | 132        |
| 2.7. Cloud Databases .....  | 133        |
| <b>3. Methodology and Data Collection.....</b>                          | <b>135</b> |
| 3.1. French Lexical Mega-Studies.....                                   | 135        |
| 3.2. French Lexical and Sub-Lexical Variables .....                     | 135        |
| 3.3. Forward Stepwise Method .....                                      | 137        |
| 3.4. Calculating the Accuracy of Lexical difficulty Models .....        | 138        |
| 3.5. The Procedure of the Usability Experiments .....                   | 138        |
| <b>4. Experiments and Results.....</b>                                  | <b>138</b> |
| 4.1. Case Study of the Optimization Model .....                         | 139        |
| 4.2. Sensitivity Analysis of the Genetic Algorithm .....                | 139        |
| 4.3. Usability Experiments .....  | 144        |
| 4.4. Regression Models for Auditory Word Recognition .....              | 147        |
| 4.5. Regression Models for Visual Word Recognition .....                | 149        |
| 4.6. Regression Models for Word Spelling .....                          | 150        |
| 4.1. Regression Model for Word Naming .....                             | 152        |
| <b>5. Conclusion .....</b>  | <b>153</b> |
| <b>6. References.....</b>   | <b>154</b> |

## CHAPTER VI

### Validation of the Individualized Reading System

#### 1. Introduction

Designing an intelligent gamified reading system requires a multidisciplinary research work that includes linguistics, educational psychology, game design, and computer science. The multidisciplinary approach developed in this research is described in this chapter. The main components of the implemented system are the reading tasks, serious game design, the individualized intelligent module, threshold adaptive mechanisms, software development, and cloud databases.

Four reading tasks were implemented. Three of them were designed to cover the orthographic, phonological, and semantic aspects of word recognition. The fourth task was designed based on the Reading Acceleration Program (RAP) [1] to focus on the reading rate. These tasks were designed based on the training principles of automaticity acquisition. The words that contain the same graphemes were grouped to increase the consistency of grapheme-phoneme conversion. Time pressure was applied to push the users to decode faster. In addition, blocked and random practice formats were implemented for the initial and advanced phases of training.

An optimization model based on the knapsack problem was used to tailor the content for each learner. It optimized the training sessions by maximizing the educational value of each session without surpassing the allowed difficulty level. A genetic algorithm was implemented to solve this optimization model. It was deployed in a cloud computing service, and its most efficient setup was determined by conducting various sensitivity analyses.

Furthermore, two usability experiments were conducted to validate the functionality of the games. Based on the feedback received in the usability experiments, an iterative design approach was adopted. The usability experiment results validated the functionality of the games and the performance of the optimization model. However, it showed that word length, and word frequency were not adequate in predicting word difficulty, and a combination of lexical variables was necessary for its prediction. This led to the creation of lexical difficulty models based on artificial neural network and linear regression.

Lexical difficulty depends on many lexical and sub-lexical variables. Unraveling the weight of each variable and the intertwined relationships between them is complex. In addition, lexical difficulty greatly varies for different tasks. A word may be easy to read but difficult to write. For addressing this problem, four lexical skills were identified. They covered most of the reading and spelling tasks. These skills were auditory and visual word recognition (represented by auditory and visual lexical decision task), word decoding (represented by word naming), and spelling.

In order to build the lexical difficulty models, two categories of databases were used. The first category was the databases of the studies that tested the lexical tasks with thousands of individuals. These studies are referred to as lexical mega-studies. The second category was the

databases containing lexical and sub-lexical variables for thousands of words. These two categories of databases were combined to acquire the training data required for creating the lexical difficulty models. For auditory and visual lexical decisions, the MEGALEX [2] database was used. The database published in the EOLE book [3] was used for spelling and the Chronolex database [4] was used for word naming. Furthermore, 55 lexical and sub-lexical variables were extracted from Lexique 3.83 [5], [6], and Lexique Infra [7] databases.

The overlapping data created a database with 25776 words for the visual lexical decision, 15842 words for the auditory lexical decision, 11373 words for the spelling, and 1481 words for the word naming. The number of independent variables was high (55). Therefore, the forward stepwise analysis was used to identify the key variables for linear regression and artificial neural network models. For each model, the top 10 lexical and sub-lexical variables were identified. Then the linear regression and ANN models were built with these ten variables. The performance of the models was evaluated through 10-fold cross-validation process using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

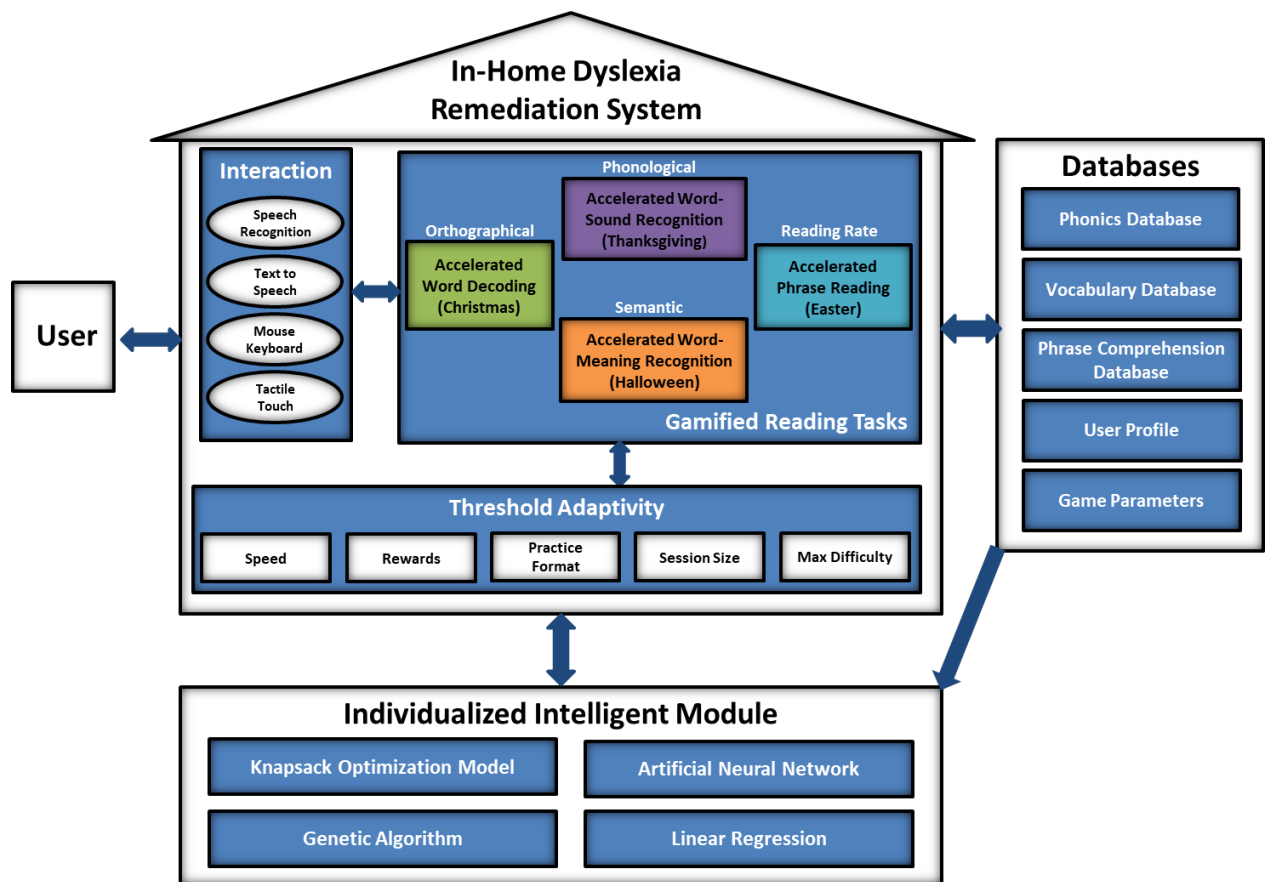
## **2. The Implementation of the Proposed Approach**

In this section, the implementation of the proposed approach is described in full detail. It covers the reading tasks, serious game design, the individualized intelligent module, threshold adaptive mechanisms, software development, and databases. The following section represents the general architecture of the system.

### **2.1. The Architecture of the System**

Based on the scientific approach presented in Chapter 3, an intelligent gamified system was designed and implemented for home-based remediation of dyslexia. It focuses on fostering the automaticity of decoding and providing a basis for extensive vocabulary instruction. The home-based approach was proposed to provide sufficient training time required for the development of automaticity and vocabulary. Gamification was proposed to increase motivation, engagement, and the adherence of the users to the training program. Furthermore, an intelligent individualized module was developed based on the combination of an optimization model with regression models. The optimization model maximized each session's educational value while controlling the difficulty level for each user. The regression models estimated the lexical difficulty levels through artificial neural networks and linear regression.

One of the important challenges in designing educational games is the collection and management of educational content. This content should be organized and indexed carefully to enable it for the use inside the intelligent component. This led to the creation of various cloud databases. Furthermore, natural user interfaces, such as speech recognition and speech synthesis, were implemented to improve interaction quality. Finally, threshold adaptive mechanisms were used to tailor the games' difficulty to the users' performance. Each of these implemented aspects of the training program is explained in more detail below. The general architecture of the system is presented in Figure 1.



## 2.2. Reading Tasks

Based on the identified research gaps and the proposed approach, four reading tasks were designed. Three tasks were focused on word recognition from the orthographic, phonological, and semantic aspects, and another task was focused on the reading rate. Each of these reading tasks is described below. Furthermore, the term *Accelerated* was used to signify the time pressure aspect that pushes the user to decode the words increasingly faster. Time pressure is one of the essential attributes of automaticity training [8], [9].

### 2.2.1. Accelerated Word Decoding (AWD)

In this task, a word was presented and the user had to pronounce it aloud. The speech recognition engine was used to detect the pronunciation and decide whether it was correct or not. Therefore, the user had to convert graphemes to phonemes. The objective was to train the learners to decode grapheme-phonemes automatically. Each grapheme is a written unit that represents a phoneme. For example, 'sh' is a grapheme representing the phoneme /ʃ/ in words such as *shoe* or *ship*. A skilled reader recognizes these graphemes automatically and unconsciously converts them to sounds [10]. To train this skill, for each grapheme-phoneme association, a word list including the association was collected. Presenting a list of words with the same grapheme-phoneme association helped induce the consistency of the stimulus-response required for automaticity acquisition. As discussed in Chapter 4, consistency is

paramount to automaticity acquisition [11], [12]. Furthermore, presenting the words based on grapheme-phonemes associations increased the contiguity of these decoding rules, and contiguity is a key issue in the implicit learning of rules and regularities [13], [14].

For training automaticity, and especially for children, it is better to start with blocked practice and then move on to random practice [15]. Therefore, in the initial phase, the words including each grapheme-phoneme association should be practiced together. Then, in the advanced phase, the words including different associations could be practiced in random order.

The intelligent component in the system assured that the optimal set of word lists were selected. However, the words in each list had to be presented one by one, and the order of the presentation was important. The words were sorted descendingly based on the frequency values, and the system initially presented the words with the highest frequency. Using this approach, the user had a higher chance of recognizing the first presented words. This helped the learner to guess the pronunciation of the next less common words that included the same grapheme-phoneme associations. Furthermore, time pressure was used to enhance automaticity acquisition [9], [16]. The task was designed to offer only a limited time for decoding each word. The time length varied as a function of the ability of the user.

### **2.2.2. Accelerated Word-Sound Recognition (AWSR)**

This task was designed to work as the opposite of the previous task. The learner heard the pronunciation of a word via text-to-speech technology, and he/she had to choose between four choices of words. Therefore, this task focused on converting phonemes to graphemes. For training automaticity in this conversion, a similar approach to the previous task was followed:

- Lists of words based on phoneme-grapheme associations were collected and used.
- The intelligent component selected the optimal training content.
- The training was designed to start by blocked practice and move to random practice.
- Words in the same list were sorted based on the frequency so that the common words were shown first, and they hinted at the pronunciation of the infrequent words
- Time pressure was used for facilitating the automaticity acquisition.

The distractors in this task could be chosen from orthographically similar words that have different pronunciations. However, this made the task difficult for beginner readers. As an alternative solution, the distractors were chosen randomly from the words in the database. This led to the selection of dissimilar words that eased the task for the users.

### **2.2.3. Accelerated Word-Meaning Recognition (AWMR)**

This task was based on cloze question. It focused on the user's vocabulary knowledge. A cloze question is a sentence with a missing word and four possible choices. For each word, three example phrases were collected. The training was designed to go through three phases, and in each phase, a different example sentence was used. It is already an established knowledge that words are better learned in context [54]. Therefore, providing more phrases including the same word helps in acquiring a word.

Time pressure technique was implemented in this task but it was less emphasized than in the design of the other tasks to allow enough time for full comprehension of the questions. The

applied time pressure in this task did not relate to the vocabulary acquisition process, it was implemented to enhance the user's reading rate. Similarly, other automaticity principles such as consistency were not required in this task because vocabulary is a declarative knowledge, and not a procedural skill that should be automatized.

#### **2.2.4. Accelerated Phrase Reading (APR)**

This task was based on a text-fading method called Reading Acceleration Program (RAP) developed by Breznitz et al. [1]. A cross-language study of RAP shown its effectiveness on children with reading disabilities in English and Hebrew [17]. It was also shown to improve reading comprehension [18], executive functions [19], and word decoding rate [20].

In this task, a phrase was presented, and the text disappeared letter by letter from the start. After the text had fully disappeared or the user had signaled that he/she had finished reading, a four-choice WH comprehension question was presented to the user. Depending on the answer, the text-fading speed was increased or decreased. This kept the users on the edge of their reading rate and pushed them to read the phrases faster. The preliminary usability tests with users showed that when the text was presented, the user needed a waiting time to focus his/her gaze on the first word before starting the text-fading process. The waiting time and the text-fading rate were adapted to the reading speed of the users. Each time a comprehension question was answered correctly, the initial waiting time was decreased, and the text-fading rate was increased by small increments. If the first answer was incorrect, but the second answer was correct, the initial waiting time and the text-fading rate remained the same. If the user answered incorrectly on the first and second trials, the waiting time was increased, and the text-fading rate was decreased by double the size of the previous increment.

### **2.3. Serious Game Design**

The four reading tasks designed for this system were transformed into four mini-games. As the system targeted young children, the theme of holidays (Christmas, Halloween, Easter, and Thanksgiving) was chosen for the games. The scenarios of the games are briefly described below.

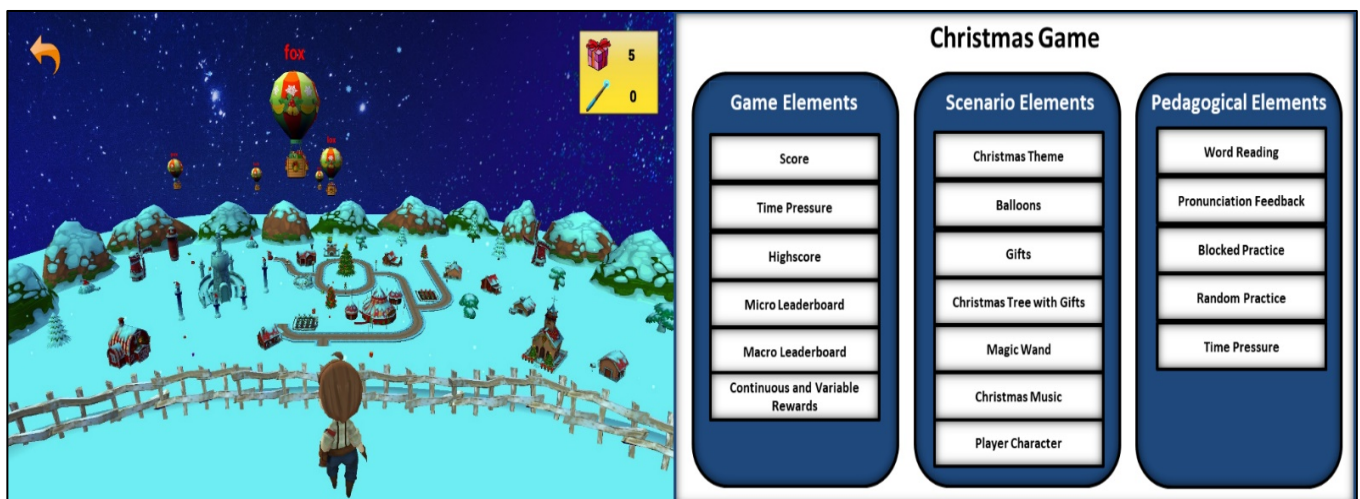
#### **2.3.1. Main Menu**

Once logged in, users can play four games. A progress bar was designed to show the progress in playing each game. In the middle, the image of a holiday superhero was placed in black and white. However, as the user progressed, it became colorful, starting from the bottom and ending at the top. This superhero acted as the overall progress in the games.



### 2.3.2. Christmas Game

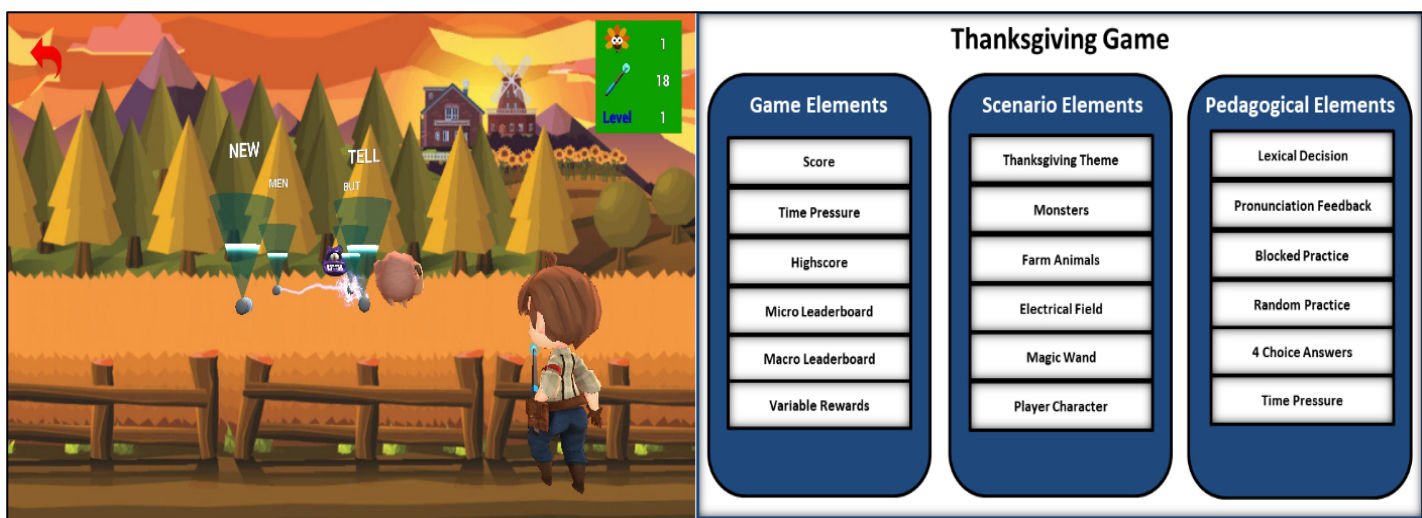
The AWD task was created with the Christmas theme. A set of balloons were travelling over a Christmas town. On top of each balloon, a word was written and each balloon contained gifts. As the balloons came forward, their gifts fell on the Christmas town one by one. By saying the correct pronunciation of the word written on top of the balloon, the user could win the balloon's remaining gifts. Therefore, the sooner the users pronounced the words aloud, the more gifts they received. If a balloon did not receive a correct pronunciation, the game stopped temporarily, and the correct pronunciation of the word was played in a simplified scene. For every ten gifts collected by the users, they won a magic wand. Each magic wand could be used at the end of the game with a tree full of lucky gifts that contained different scores. After spending all the magic wands, a leaderboard containing both the micro and macro leaderboards was displayed.





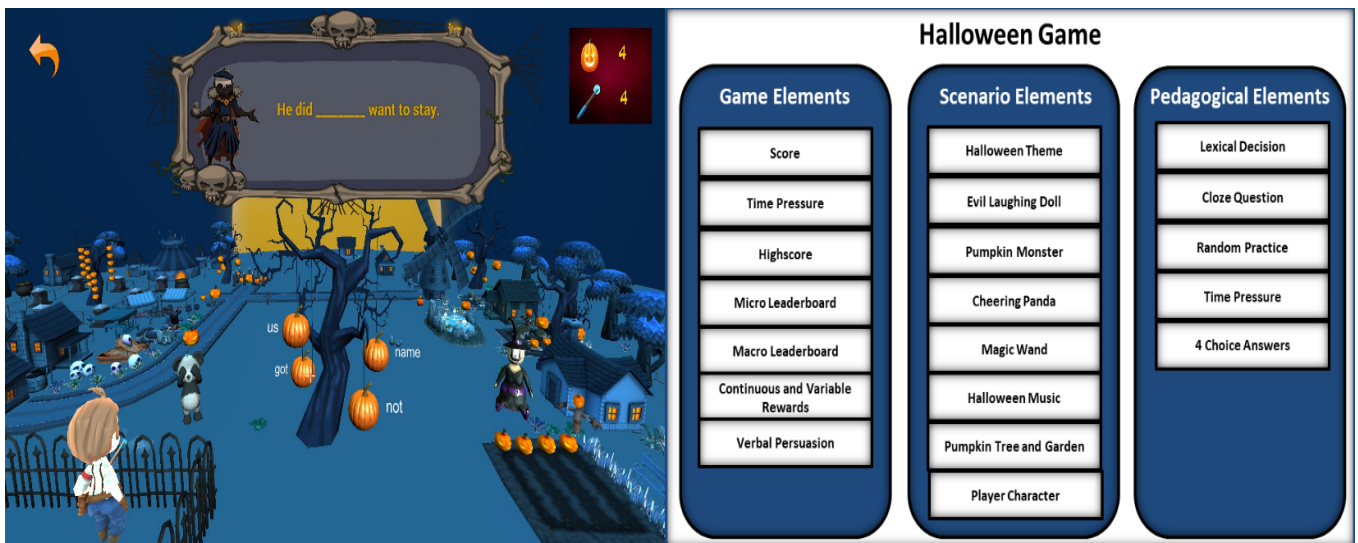
### 2.3.3. Thanksgiving Game

The AWSR task was gamified with a thanksgiving theme. In this game, the player should stop monsters coming towards the character. For stopping them, the player should have launched an electrical field. This electrical field could be activated only if the player chose the correct word (among three distractors) pronounced by the game. Then the activated electrical field could transform the monsters into farm animals. Alternatively, the players could transfer monsters directly out of the game with the three trials of a magic wand. The player had a limited number of magic wands. By leveling up more wands could be earned. The speed and number of monsters was increasing by each level. Each farm animal that the user acquired had a score, and when one of the monsters managed to escape the user, the game was ended, and a leaderboard containing both the micro and macro leaderboards was displayed.



### 2.3.4. Halloween Game

The AWMR task was gamified with a Halloween theme. A cloze question with four-word choice were presented. Each word was written besides a pumpkin hanging from a tree. If they selected the correct answer, a panda bear applauded them, and they could win a magic wand. In addition, with each correct answer, a pumpkin was planted in the pumpkin garden. However, if the answer was wrong, a Halloween doll laughed sinisterly. In addition, the pumpkin containing the right answer was transformed into a monster pumpkin jumping into the scene, and trying to eat the planted pumpkins. To avoid that, the user should have used the magic wand to transfer the monster pumpkin out of the game. When the questions are finished, the surviving planted pumpkins were transformed into magic wands. The player character was teleported to another scene with a tree full of lucky pumpkins, each having different scores. They could choose the lucky pumpkins by using their magic wands. After spending all the magic wands, a leaderboard containing both the micro and macro leaderboards was displayed.

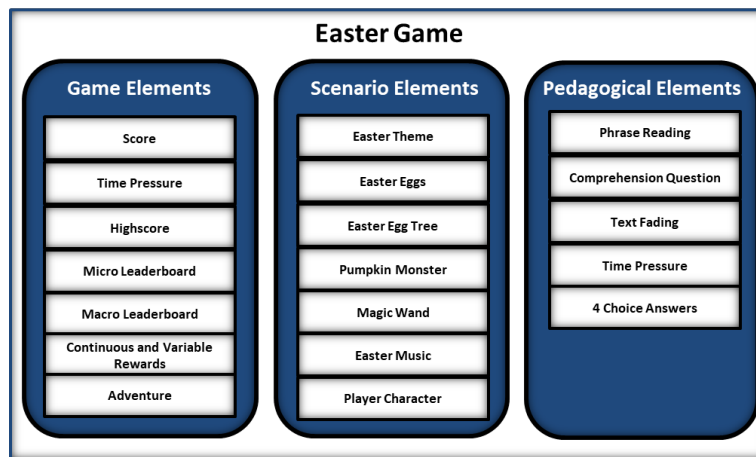


### 2.3.5. Easter Game

The APR task was gamified with an Easter theme. This game was happening in an Easter town that had a maze-like environment. The player had to find the Easter eggs placed in random corners of the town. With each egg found by the player, a text-fading exercise with a comprehension question was presented in a canvas. The initial text-fading speed was determined by a calibration during the first session. However, the users could change the calibration at any time. Each egg contained three or four exercises. Depending on the accuracy of the answers, the user could win a number of magic wands. Once all the Easter eggs were collected, the player was teleported to a scene that had a tree full of lucky eggs having different scores. The lucky eggs could be acquired by the magic wands. After spending all the magic wands, the user's score was recorded, and a leaderboard containing both the micro and macro leaderboards was displayed.

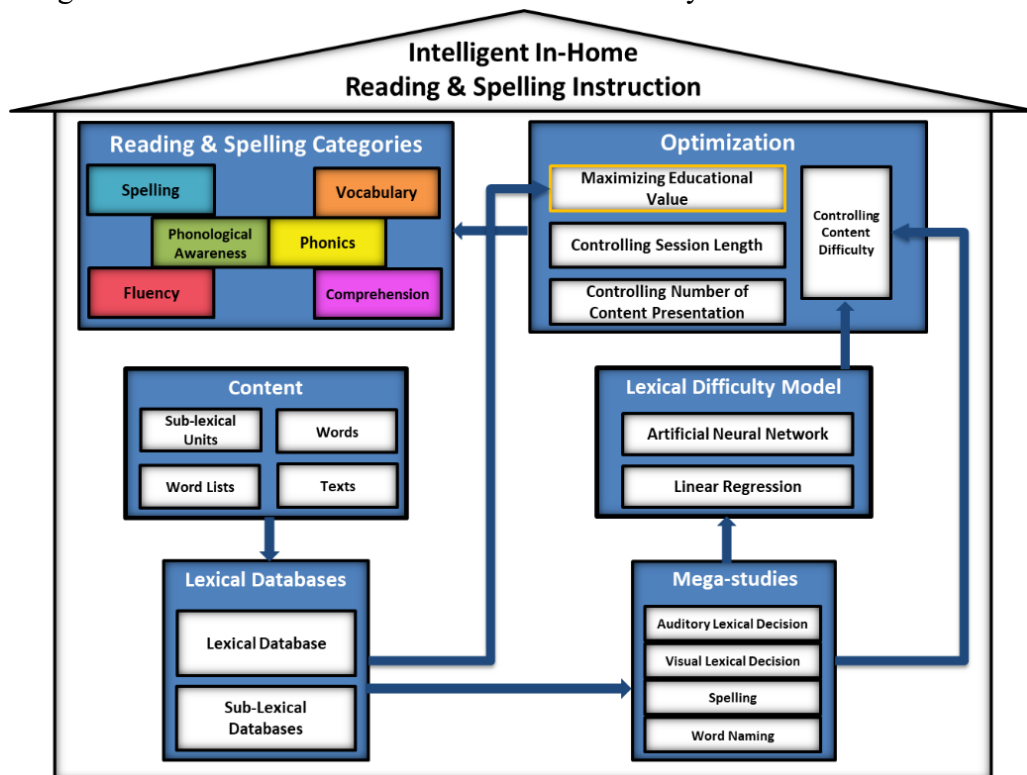


Figure 9. Screenshots from the Easter game



## 2.4. Individualized Intelligent Module

It is highly recommended for reading and writing programs to be systematic [21], [22]. To comply with this recommendation, an intelligent model was developed in this research. The architecture of this intelligent model is illustrated in the diagram below. Each part of this architecture is explained in full detail in the following sections. This intelligent component was designed based on the combination of an optimization model with regression models. The optimization model maximized each session's educational value without surpassing the allowed difficulty level for each user. The regression models estimated the lexical difficulty levels for visual and auditory word recognition, word decoding, and spelling. Artificial neural networks and linear regression were used to build these lexical difficulty models.



**Figure 11.** The architecture of the individualized intelligent module

### 2.4.1. Optimization

According to the Lev Vygotsky's Zone of Proximal Development, most of the development happens when the instructional content is slightly more difficult than what the learner can do on his/her own [23]. In addition, the flow theory proposed by Mihaly Csikszentmihalyi confirms that the peak of motivation is reached when there is a balance between the task's difficulty and the individual's abilities [24]. When the material is too difficult, the user will be frustrated or anxious; however, if the learning material is too easy, the user will be bored. Therefore, the learning program should provide the content with maximum educational value while remaining close to the user's ability level. These intelligent characteristics are essential for any in-home interventional programs.

Maximizing the educational value of a training session while keeping the session's difficulty level within the learner's abilities is an optimization problem. Modeling and solving this optimization problem requires deep linguistic knowledge and high computation power, which cannot be expected from teachers.

The knapsack problem is an optimization method already used in previous research on education [25]. This model has the appropriate characteristics for modeling the reading and spelling tasks proposed in this thesis. In the knapsack problem, for a given set of items, each with a weight and value, we should choose the items that offer the highest total value while not surpassing the weight limit. For modelling the reading tasks into a knapsack problem, every content material was marked by a value (representing the educational value) and a weight (representing the difficulty level).

The objective function of the knapsack model maximized the total value of the selected items. The constraint on maximum difficulty ensured that the selected items remained in the defined weight limit. This constraint included an Upper Weight (UW) for the upper limit and a Lower Weight (LW) for the lower limit. UW limited the content's difficulty to avoid frustration, and LW limited its easiness to avoid boringness. An additional constraint maintains the number of selected items within the defined boundaries. The Upper Number (UN) was the upper limit on the maximum number of items that could be selected, and Lower Number (LN) was the lower limit. This constraint ensured that the length of the sessions were uniform. The third constraint limited the Number of Presentations (P) for each content material to a fixed maximum value so that no material would be presented more than that maximum. The proposed optimization model can be presented by:

The objective of maximizing the value of each training session:

$$\begin{aligned} &\text{Maximize } \sum_{i=1}^n v_i x_i \\ &\text{Subject to} \end{aligned} \tag{1}$$

The constraint for the difficulty level of each training session:

$$LW < \sum_{i=1}^n w_i x_i \leq UW \tag{2}$$

The constraint for the number of items in each training session:

$$LN \leq \sum_{i=1}^n x_i \leq UN \quad (3)$$

The constraint for the number of Presentations for each reading material:

$$\forall_i p_i x_i \leq P \quad (4)$$

Where:

$$x_i \in \{0, 1\} \quad (5)$$

The table below presents the variables and parameters of the optimization model.  $x_i$  is the variable and the other elements are the parameters of the model.

**Table 1:** The notation used for the optimization model

|       |   |
|-------|---|
| $x_i$ | Item $i$ with a value of 0 or 1                         |
| $v_i$ | Value of item $i$                                       |
| $w_i$ | weight of the item $i$                                  |
| $p_i$ | Number of Presentations of the item $i$                 |
| $LW$  | The minimum allowed weight                              |
| $UW$  | The maximum allowed weight                              |
| $LN$  | The lower limit on the number of selected item          |
| $UN$  | The upper limit on the number of selected items         |
| $P$   | The upper limit on the presentation count for each item |

#### 2.4.2. Genetic Algorithm

The proposed optimization model was an NP-Complete problem, and when the size of the problem is big, there is no known method to quickly find the best solution. The meta-heuristic method of Genetic Algorithm (GA) is effective in approximating computationally complex models and it was adopted to solve the proposed optimization model. The first step to implement a genetic algorithm was deciding on the nature of the genes. In this context, the genes were binary values, and each gene represented a single instructional material (it could be a single word, word list, question, sentence, or text). If the gene is equal to zero, this means that the material was not selected, and if the gene is equal to one, this means that the material was selected. The second step was to define the chromosome. Here, the chromosome was an array of binary values (genes). The size of the chromosome was identical to the size of the content. The third step was the random creation of chromosomes for the first population. For achieving this step, the chromosomes with all-zero genes were created. For each chromosome, a number of genes (equal to the length of the session) were selected randomly, and they were set to one.

The fourth step was to define the fitness function. For this problem, a piecewise function was developed as the fitness function. It divided the solutions into the categories of **strictly infeasible**, **mildly infeasible**, **feasible but not desired**, and **feasible and desired**. The first category of strictly infeasible happened when the selected items had a total weight more than the maximum allowed weight. For this category, the big-M (penalty) method was used by assigning a big negative number to the fitness function and reducing the total weight.

Therefore, when two infeasible solutions were compared, the one with lower total weight had a better fitness value. Using this method converged the solutions towards feasible areas.

The second category of mildly infeasible referred to the solutions that had respected the weight constraint but had violated the session length constraint. For this category, a small-M (penalty) was used and multiplied by the deviation degree from the allowed session length. Using this method converged the session lengths towards the allowed boundaries.

For the third category of feasible but not desired, the fitness function's value was still negative, but in small quantities in comparison to the big-M or small-M values. This third state occurred when the main constraints on total weight and session length were respected, but the other constraint on the number of presentations was violated. In this situation, the fitness function's value was set to the extent of the deviation from the constraint. Therefore, in the case of two similar solutions that both had violated the constraint on the number of presentations, the solution that deviated less was regarded as a fitter solution. This converged the solutions towards the feasible areas. For the fourth category, which was feasible and desired, the fitness function was always positive. It was equal to the total value of the selected reading materials. The piecewise function used for calculating the fitness of the solutions is presented below:

$$F = \begin{cases} M - w, & w > W \\ m * (LN - n), & n < LN \text{ and } w < W \\ m * (n - UN), & n > UN \text{ and } w < W \\ p - P, & p > P \text{ and } LN < n < UN \text{ and } w < W \\ v, & p < P \text{ and } LN < n < UN \text{ and } w < W \end{cases} \quad (6)$$

**Table 2:** The notation used for the fitness function

|      |  |
|------|--|
| $F$  | <i>Fitness value of the solution</i>                           |
| $w$  | <i>Total weight of the solution</i>                            |
| $v$  | <i>Total value of the solution</i>                             |
| $n$  | <i>Number of selected items in the solution</i>                |
| $p$  | <i>Total number of prior Presentations</i>                     |
| $M$  | <i>Big negative penalty</i>                                    |
| $m$  | <i>Small negative penalty</i>                                  |
| $LW$ | <i>The minimum allowed weight</i>                              |
| $UW$ | <i>The maximum allowed weight</i>                              |
| $LN$ | <i>The lower limit on the number of selected item</i>          |
| $UN$ | <i>The upper limit on the number of selected items</i>         |
| $P$  | <i>The upper limit on the presentation count for each item</i> |

### 2.4.3. Lexical difficulty Models

The knapsack optimization model suits the problem of this s well only if its parameters are set accurately. Lexical difficulty level was considered as the weight parameter of the model. However, estimating lexical difficulty is complex. Tens of lexical and sub-lexical variables affect lexical difficulty. The order of their importance is not fully known, and the relations between them are complex.

Studies have shown that particularly for beginners in reading, word length is a better predictor of lexical difficulty [26], [27]. In addition, the word length effect is more pronounced

in less skilled readers and people with reading disorders such as dyslexia or pure alexia [28]. Word-length can be presented by the number of letters, graphemes, phonemes, syllables, or morphemes. Another important variable is lexical frequency. The effect of lexical frequency on word naming and lexical decision times has been shown [29], [30]. Another deciding variable for the lexical difficulty is orthographic consistency. It includes grapheme to phoneme (GP) mapping consistency (feed-forward consistency) and phoneme to grapheme (PG) mapping consistency (feedback consistency). Both of them have been shown to have effects on word recognition [31]. In addition, the number of consonant clusters [32], the number of neighbors (words that differ only in one letter or one phoneme) [33], [34], and the number of homophones [35], all can affect lexical difficulty.

In addition, lexical difficulty depends on the task. For example, in a reading task, the word may be considered easy, but it may prove difficult in writing. This is particularly the case for French language, which has an asymmetric orthography, meaning that writing is often more challenging than reading [36]. Moreover, one word may be difficult for one age but easy for another age.

These layers of complexity create a challenge in the accurate estimation of lexical difficulty. In this study, regression models based on linear regression and artificial neural networks were used to estimate lexical difficulty in the French language. Linear regression and artificial neural networks are statistical methods to model the relationship between dependent and independent variables. There are advantages and drawbacks for each method. Linear regression requires a relatively low amount of data, it is computationally fast, and it is easy to understand and interpret. However, its main drawback is its lack of accuracy in the case of non-linearity. Most of the real-life cases include non-linear relationships, and linear regression fails to capture their underlying relationships.

On the other hand, artificial neural networks have the capacity to model the most complex problems. It can handle both linearity and non-linearity. Depending on the number of hidden layers and the number of neurons in each layer, it can offer significant computational power. However, it requires considerably more input-output data to reach acceptable accuracy than the linear regression models. In addition, understanding and interpreting its model is much more complex. Furthermore, its training time can be significantly longer than linear regression. Since linear regression and artificial neural networks offer different possibilities, both models were used in this study for the creation of lexical difficulty models. The comparison of their accuracy could help to decide which one to use. These models and the evaluation of their accuracies are presented in Section 4.

## **2.5. Threshold Adaptivity**

The educational games should be adaptable to the learner's ability level to prevent boredom or frustration. In addition to the adaptive content, game mechanics should also adapt to the ability of the learner. In this study, the four games were analyzed, and the game elements that influenced the difficulty of the games were identified. Then, a threshold mechanism was developed for each of the identified game element. These mechanisms adapted the game elements to the player's performance level. Based on each player's maximum weight parameter,



the difficulty level of game elements was adjusted automatically. These threshold parameters were stored in the Game Parameters database in the cloud. This enabled adjusting the parameters immediately according to user feedback. The threshold adaptivity was applied to several game elements in the games, but a similar formula (presented below) was used in most cases.

The following formula calculated the adaptive value of each game element:

$$X = x_b + x_{inc} \quad (7)$$

Where:

$$\begin{aligned} x_{inc} &= (w_{max} - w_b) * c \\ &\text{if } (x_{inc} < x_l): \\ &\quad x_{inc} = x_l \\ &\text{else if } (x_{inc} > x_u): \\ &\quad x_{inc} = x_u \end{aligned} \quad (8)$$

**Table 3.** The notation used for the adaptive mechanism formula

|           |  |
|-----------|--|
| $X$       | <i>The adaptive value of the game element</i>                                |
| $x_b$     | <i>Baseline value of the game element</i>                                    |
| $x_{inc}$ | <i>Incremental value for adding to the baseline value</i>                    |
| $w_{max}$ | <i>Maximum weight calculated based on the prior performances of the user</i> |
| $w_b$     | <i>Baseline of the weight parameter</i>                                      |
| $C$       | <i>Coefficient</i>   |
| $x_l$     | <i>lower limit on the value of the game element</i>                          |
| $x_u$     | <i>Upper limit on the value of the game element</i>                          |

In the Christmas game, three parameters followed the adaptive formula. First parameter was the vertical traveling speed of the balloons behind the mountains in the game scene. The second parameter was their traveling speed over the Christmas town. The third parameter was the waiting time between the introduction of each balloon. In the Halloween game, three parameters followed the adaptive formula. The first parameter was the walking speed of the pumpkin monster. The second was the waiting time between the pumpkin monster's appearance and the time it attacks the pumpkin garden. The third parameter was the waiting time between each attack of the pumpkin monster on the pumpkin garden. In the Thanksgiving game, two parameters followed the adaptive formula. The first parameter was the moving speed of the monsters, and the second was the waiting time between the introduction of the monsters. In the Easter game, the most important parameters were the text-fading speed and the initial waiting time before starting the text-fading. These two variables were first calculated based on the initial calibration carried out by the users. Then, they were adjusted each time, depending on the user performance.



## 2.6. Development and Implementation

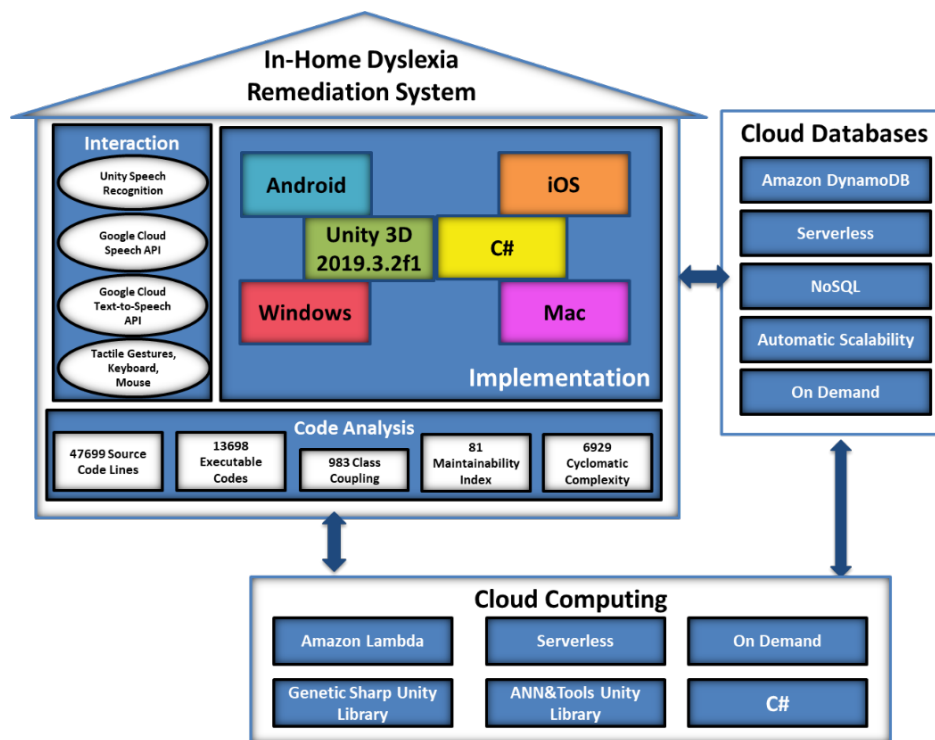
For gamifying the reading tasks, the Unity 3D engine and C# programming language were used. The 3D environment offered a more immersive and engaging experience [37]. Because of the nature of the reading tasks, speech recognition and text to speech (TTS) were used. Despite the higher clarity of human recorded speech, TTS is an agile and practical tool for dealing with thousands of words. For text to speech, the Google Cloud Text to Speech API was used. This Google service provided speech synthesis in many languages, and it worked on cross-platform.

Speech recognition technology was used to evaluate the pronunciation accuracy of the users. Speech recognition engines provide a parameter called “confidence level”. Different levels allow for a looser or stricter recognition. As choosing the right level was a matter of trial and error, this parameter was put on the cloud to enable immediate adjustment following the users' feedback. The Unity Windows speech recognition engine offered instantaneous recognition, which better suited the game's interactivity than the cloud alternatives. The cloud counterparts recorded the voice, sent it to the cloud, and then retrieved the results from the cloud. This generated a delay that affected the user experience. However, ultimately, for making the games cross-platform, Google Cloud Speech Recognition API was used.

The data relating the user profile, the educational contents, and game parameters were stored in the cloud databases. They could be accessed and modified in real-time during the gameplay. This allowed the system to synchronize the users' scores in the leaderboards. However, more importantly, it allowed the use of the system from home and school. The game parameters database in the cloud allowed high control and configurability of the system. It can also be used as a building block for the therapist or instructor's interface, giving them the capability to tweak the games for each student.

AWS DynamoDB service was used to host the cloud databases. This was a NoSQL serverless service that worked on-demand and scaled automatically. For the cloud computing, another AWS service called Lambda was used. This was also a serverless service that worked on-demand. The Genetic Sharp library was used to implement the genetic algorithm in C#, although it was modified to suit the problem in this study. Similarly, the ANN&Tools library was used for building the artificial neural network in C#. However, the regression models were tested with Scikit-learn and Keras libraries of Python to confirm the acquired results.

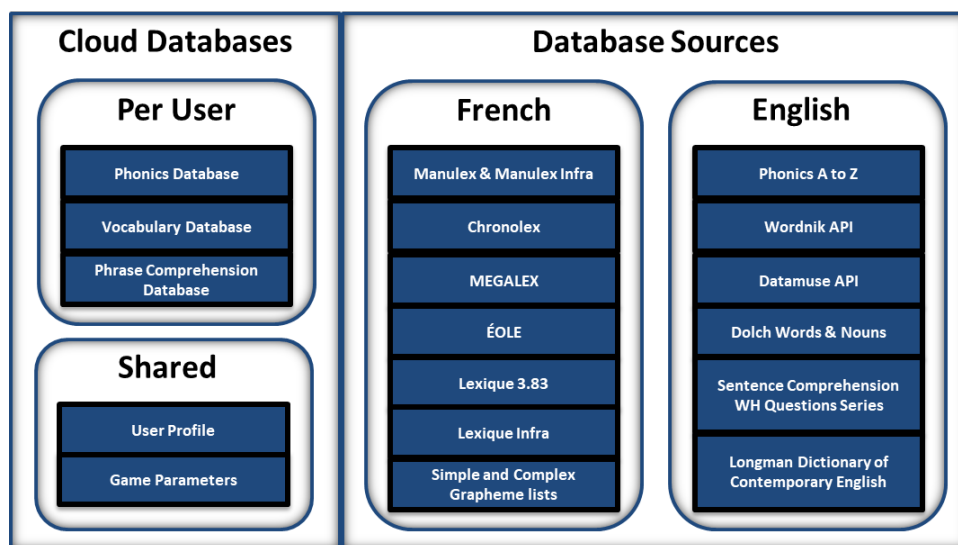
Finally, the games were published for Microsoft Windows, macOS, iOS, and Android. The result of the Microsoft Visual Studio code analysis showed that there were 47699 lines of code in the project, of which 13698 were executable code. Furthermore, 983 class couplings existed in the totality of the project. It scored 81 on the maintainability index, and the Cyclomatic Complexity was 6929. The summary of the technologies used to develop the system is illustrated in the diagram below. The next section details the cloud databases developed in this study.



**Figure 12.** Technologies used in the development of the system

## 2.7. Cloud Databases

One of the main challenges in developing educational serious games is the creation of databases for educational content. Most of the valuable contents are scattered in books, and a small percentage of them are available online. The second problem is that in general, educational contents are produced by teachers who do not have expert knowledge, and the extra information for indexing and quantified analysis is not provided. For this research, various databases were created and deployed in the AWS DynamoDB cloud service. The following sections present five of the most important databases integrated into the system.



**Figure 13.** Databases used in the system and their sources

Phonics Database: This database was constructed for the two games that target systematic phonics instruction (Christmas and Thanksgiving). It included the word lists, listed based on common grapheme-phoneme associations. Each word and word list were marked with parameters of value and difficulties, as well as other parameters for analyzing user performance. This database was used for both random and blocked practices. The content of this database was collected from *Phonics from A to Z* [38] for the English language, and *Simple and Complex Grapheme* lists [39] for the French language. The lexical data for the English words were collected from the Wordnik API [40], and the lexical data for the French words were collected from the Manulex database [41].

Vocabulary Database: The vocabulary database was used in the Halloween game to target vocabulary knowledge. Each word in the database had a cloze question and three distractors. This content was either generated by the authors or collected from the example sentences of various resources, notably the Longman Dictionary of Contemporary English [42]. For each word, three example phrases were collected to be used as cloze questions. Encountering multiple phrases in different contexts enforces the vocabulary acquisition [43].

Phrase Comprehension Database: Easter game required a database of phrases or short passages with comprehension questions and four-choice answers. This data was extracted from a series of books on sentence comprehension with WH questions [44]. This resulted in 118 phrases with comprehension questions in the database.

User Profile Database: The user profile database contained the data for accessing, playing, and analyzing the game. The data for accessing the game was a username and password for each user. The sign-in process with username and password was implemented in the game to avoid the collection of personally identifiable information such as Device ID, and also, for accessing the games from several platforms. Several anonymous user information was stored in this database. Notably, it included the Max Weight parameter, high score values, time spent playing in each game, number of playing sessions, and age.

Game Parameters Database: The developed system and its games have many parameters. Hard coding these parameters had few drawbacks. First, for changing any of them, a new version of the games should be released. This slowed down the feedback-driven development. Releasing multiple versions of the game at the same time with different settings became error-prone and time-consuming. Therefore, 51 parameters were stored in the Game Parameters database on the cloud. In addition, instead of a global set of parameters for all the users, the parameters could be easily used separately for each individual. Furthermore, the language parameter was stored in this database. For each language, the relevant educational content and the user interface translation could be loaded automatically. The 51 parameters contained in this database were mainly related to the adaptive game mechanics, and the intelligent model.

### 3. Methodology and Data Collection

This section presents the data collection procedure, the methods used for obtaining the results, and the data analysis.

#### 3.1. French Lexical Mega-Studies

For building the lexical difficulty models, the French lexical mega-studies were used to collect the data related to auditory and visual lexical decisions, word naming, and spelling. These databases were the results of behavioral mega-studies that tested the lexical tasks of numerous words with many human subjects. In word naming, the individual has to read a word aloud, and the time it took for reading is recorded. In auditory lexical decision, the pronunciation of a word is played, and the user has to decide whether it is a real word or a pseudoword, and the reaction time is recorded. In the visual lexical decision task, a written word is presented, and the user has to decide whether the word is a real word or a pseudoword. In the spelling task a word is pronounced, and the individual has to write its correct spelling, and the error or success is recorded.

For the spelling task, the database published in the EOLE book [3] was used. This database contained the results of spelling task for around 12000 words tested with numerous students in over 2000 French classrooms. In this work, Béatrice and Philippe Pothier collected the average spelling success-rate of students for all the words across the five grades of the primary school in France. Regarding the auditory and visual lexical decision tasks, the MEGALEX [2] database was selected, it includes the decision times and accuracies of 28,466 words for visual lexical decision and 17,876 words for auditory lexical decision. The Chronolex [10] database was selected for the word-naming task, it includes 1482 monosyllabic and monomorphemic words.

#### 3.2. French Lexical and Sub-Lexical Variables

The French language was used for creating lexical difficulty models. The Lexique 3 database [5], [6], was used for the general lexical variables, and the Lexique-Infra [7] was used for sub-lexical variables. Both of these databases contained the data of around 140000 French words. These two databases were combined, and only quantified variables were selected. The result was 55 lexical and sub-lexical variables. The table below presents the list of these 55 lexical and sub-lexical variables collected from Lexique 3.83 and Lexique-Infra databases. All of the quantified data was normalized to a value between zero and one. The frequency for words was calculated by averaging the frequency values in books and films available in the Lexique 3 database. Since these average frequencies were dispersed, the Standard Frequency Index (SFI) formula was used to homogenize them before the normalization process. The formula below was adopted from the Manulex database [41], originally proposed by John B. Carroll [45]. SFI is the Standard Frequency Index, and  $c$  is the word count.

$$SFI = 10 * (\log_{10} c + 4) \quad (9)$$

**Table 3.** The lexical and sub-lexical variables

| Variable                  | Definition   | Source               |
|---------------------------|--|----------------------|
| <b>Frequency</b>          | Standard Frequency Index (SFI) calculated based on the frequency counts                                | <b>Lexique 3.83</b>  |
| <b>islem</b>              | If the word is a lemma (canonical form), the value is one, and if it is not a lemma, its value is zero |                      |
| <b>nblettr</b>            | Number of letters  |                      |
| <b>nbphons</b>            | Number of phonemes   |                      |
| <b>nbsyll</b>             | Number of syllables  |                      |
| <b>nbmorph</b>            | Number of morphemes  |                      |
| <b>nbhomoph</b>           | Number of homophones   |                      |
| <b>nbhomogr</b>           | Number of homographs   |                      |
| <b>voisorth</b>           | Number of orthographic neighbors   |                      |
| <b>voisphon</b>           | Number of phonological neighbors   |                      |
| <b>puorth</b>             | Orthographic unicity point   |                      |
| <b>puphon</b>             | Phonological unicity point   |                      |
| <b>Freq_GP</b>            | Mean grapheme-phoneme Consistency (Type Count)   | <b>Lexique-Infra</b> |
| <b>FreqTok_GP</b>         | Mean grapheme-phoneme Consistency (Token Count)  |                      |
| <b>Freq_I_GP</b>          | Grapheme-phoneme consistency for the initial grapheme (Type Count)                                     |                      |
| <b>Freq_I Tok_GP</b>      | Grapheme-phoneme consistency for the initial grapheme (Token Count)                                    |                      |
| <b>Freq_M_GP</b>          | Grapheme-phoneme consistency for the middle graphemes (Type Count)                                     |                      |
| <b>Freq_M Tok_GP</b>      | Grapheme-phoneme consistency for the middle graphemes (Token Count)                                    |                      |
| <b>Freq_F_GP</b>          | Grapheme-phoneme consistency for the final grapheme (Type Count)                                       |                      |
| <b>Freq_F Tok_GP</b>      | Grapheme-phoneme consistency for the final grapheme (Token Count)                                      |                      |
| <b>minfreqgraph_GP</b>    | Minimum grapheme-phoneme consistency of the word (Type Count)  |                      |
| <b>minfreqgraphTok_GP</b> | Minimum grapheme-phoneme consistency of the word (Token Count)   |                      |
| <b>countregTy_GP</b>      | Number of grapheme-phoneme irregularities in the word (Type Count)                                     |                      |
| <b>countregTo_GP</b>      | Number of grapheme-phoneme irregularities in the word (Token Count)                                    |                      |
| <b>posregTy_GP</b>        | Position of the first grapheme-phoneme irregularity in the word (Type Count)                           |                      |
| <b>posregTo_GP</b>        | Position of the first grapheme-phoneme irregularity in the word (Token Count)                          |                      |
| <b>Freq_PG</b>            | Mean phoneme-grapheme Consistency (Type Count)   |                      |
| <b>FreqTok_PG</b>         | Mean phoneme-grapheme Consistency (Token Count)  |                      |
| <b>Freq_I_PG</b>          | Phoneme-grapheme consistency for the initial phoneme (Type Count)                                      |                      |
| <b>Freq_I Tok_PG</b>      | Phoneme-grapheme consistency for the initial phoneme (Token Count)                                     |                      |
| <b>Freq_M_PG</b>          | Phoneme-grapheme consistency for the middle phoneme (Type Count)                                       |                      |
| <b>Freq_M Tok_PG</b>      | Phoneme-grapheme consistency for the middle phoneme (Token Count)                                      |                      |
| <b>Freq_F_PG</b>          | Phoneme-grapheme consistency for the final phoneme (Type Count)  |                      |
| <b>Freq_F Tok_PG</b>      | Phoneme-grapheme consistency for the final phoneme (Token Count)                                       |                      |
| <b>minfreqgraph_PG</b>    | Minimum phoneme-grapheme consistency of the word (Type Count)  |                      |
| <b>minfreqgraphTok_PG</b> | Minimum phoneme-grapheme consistency of the word (Token Count)   |                      |
| <b>countregTy_PG</b>      | Number of phoneme-grapheme irregularities in the word (Type Count)                                     |                      |
| <b>countregTo_PG</b>      | Number of phoneme-grapheme irregularities in the word (Token Count)                                    |                      |
| <b>posregTy_PG</b>        | Position of the first phoneme-grapheme irregularity in the word (Type Count)                           |                      |
| <b>posregTo_PG</b>        | Position of the first phoneme-grapheme irregularity in the word (Token Count)                          |                      |
| <b>complexmoygraph</b>    | Average complexity of graphemes (average number of letters for each grapheme)                          |                      |
| <b>LetFreqTy</b>          | Average letter frequency in the word (Type Count)  |                      |
| <b>LetFreqTo</b>          | Average letter frequency in the word (Token Count)   |                      |
| <b>BigrFreqTy</b>         | Average bigram frequency in the word (Type Count)  |                      |
| <b>BigrFreqTo</b>         | Average bigram frequency in the word (Token Count)   |                      |
| <b>TrigrFreqTy</b>        | Average trigram frequency in the word (Type Count)   |                      |
| <b>TrigrFreqTo</b>        | Average trigram frequency in the word (Token Count)  |                      |
| <b>PhonFreqTy</b>         | Average phoneme frequency in the word (Type Count)   |                      |
| <b>PhonFreqTo</b>         | Average phoneme frequency in the word (Token Count)  |                      |
| <b>BiphonFreqTy</b>       | Average biphone frequency in the word (Type Count)   |                      |
| <b>BiphonFreqTo</b>       | Average biphone frequency in the word (Token Count)  |                      |
| <b>GraphFreqTy</b>        | Average grapheme frequency in the word (Type Count)  |                      |
| <b>GraphFreqTo</b>        | Average grapheme frequency in the word (Token Count)   |                      |
| <b>SylphonFreqTy</b>      | Average phonological syllable frequency in the word (Type Count)                                       |                      |
| <b>SylphonFreqTo</b>      | Average phonological syllable frequency in the word (Token Count)                                      |                      |

The lexical and sub-lexical database, with its 55 variables, was combined with the databases of each of the four lexical tasks. The overlapping data generated four databases with 11373 words for the spelling task, 25776 words for the visual lexical decision, 15842 words for the auditory lexical decision, and 1481 words for the word-naming task. Furthermore, the spelling task had only the success rates as the output data, and the word-naming database had only the reaction times. The lexical decision tasks had both reaction times and error rates. However, for the lexical decision tasks, the error rate was not included in building the regression models. This decision was made because the subjects involved in these tasks were adults, and the number of errors for most of the words was significantly low. Hence, it was concluded that reaction time was a better indicator of word recognition difficulty. In addition, for the lexical decision tasks, the standardized reaction times were used instead of the raw reaction times. Finally, all these data were normalized to values between zero and one.

### 3.3. Forward Stepwise Method

The forward stepwise method was used to identify the top 10 lexical and sub-lexical variables for building each lexical difficulty model. If none of the variables were dropped for building the lexical difficulty model, the model could become unnecessarily complex. Some of the independent variables did not significantly affect the dependent variable. In addition, some variables, despite their significant effects, had strong collinearity with other variables. To overcome these problems, the forward stepwise analysis was used to identify the most important variables for each lexical difficulty model. The forward stepwise method started with a model that had no variable (called NULL model). Then, the variable with the most predictive value was selected and added to the model. Then in the next step, the new model was tested with all the remaining variables, and again the variable that added the most predictive value was kept into the model. This process could end when the change in the predictive value became insignificant. The process of forward stepwise analysis was performed for both linear regression and Artificial Neural Network. For the forward stepwise linear regression, the selection criterion was the probability of F (p-value). For the forward stepwise artificial neural network, the selection criterion was the Mean Absolute Error (MAE), which was calculated from the formula below:

$$MAE = \frac{1}{N} \sum_{n=1}^N (|a_1 - b_1|) \quad (10)$$

Where  $a$  is the real value and  $b$  is the predicted value.

However, for evaluating the accuracy of the models, another error rate was used alongside MAE to get a complete picture. This second error rate is the Root Mean Square Error (RMSE), which was calculated from the formula below:

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (a_1 - b_1)^2}{N}} \quad (11)$$

Where  $a$  is the real value and  $b$  is the predicted value.

In this study, extensive training time was allowed to make the model choose the variables accurately. In the technical terms of the artificial neural network, each time the model is trained with the whole dataset, an Epoch has been completed. In this study, the ANN models were allowed to have 1000 epochs for recording their highest accuracy. In each step, all the models were compared, and the model having the better accuracy was selected. The lengthy training time was allowed because, with each step, the differences between the variables narrowed, and the selection of better variable required high precision. For identifying the top 10 variables among 55 lexical variables, the forward stepwise method required training 505 models for each lexical tasks. Building 505 models with 1000 training epochs and shuffling all the training sets for each epoch required extensive computation time.

### **3.4. Calculating the Accuracy of Lexical difficulty Models**

For evaluating the final accuracy of the lexical difficulty models, 10-fold cross-validation method was used. The dataset was divided into 10 equal sized (randomly selected) subsets. Then, one subset (test set) was kept for the validation of the model and the remaining nine folds were used for training the model. When the model was fully trained, the test set was used to evaluate the accuracy of the model and the accuracy metrics of RMSE and MAE was recorded. Then another subset was used as the new test set and the model was trained with the remaining subsets. This process was repeated ten times and the average RMSE and MAE was calculated. This cross validation process was followed in the same way for both linear regression and artificial neural network models.

### **3.5. The Procedure of the Usability Experiments**

Two usability studies were carried out to test the functionality and usability of the games. The two studies were separated by a period of four months to allow conducting an iterative design of the system. For the first study, 15 participants were selected from English as second language learners who volunteered to participate in the functionality testing study. Sixteen other participants who were native English speakers were involved in the second study. The two samples included 18 males (58%), with ages ranging between 7 and 50 (Mean=19.4, SD=8.9). Four of them were left-handed (13%), and eight worn glasses (26%). Each of the volunteers played the games individually, and they were asked to play each game at least once. A combination of subjective data (questionnaires and written feedback) and objective data (player performance and engagement) was used to evaluate the games. System Usability Scale (SUS) [46] was used to evaluate the usability of the gamified reading system, and a short version of the Situational Motivation Scale (SIMS) [47] was administered to evaluate the perceived motivation of the users.

## **4. Experiments and Results**

This section starts by presenting the optimization model's performance evaluations and the sensitivity analysis of the genetic algorithm. Then, the results of the usability experiments are

presented. Finally, the linear regression and ANN forward stepwise analyses of the lexical tasks are presented.

#### **4.1. Case Study of the Optimization Model**

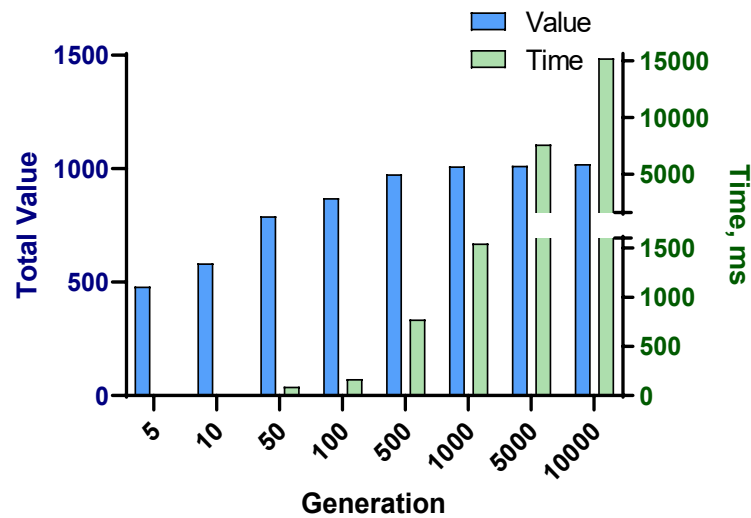
As a case study, a database of 315 Dolch sight words was prepared. The weights were difficulty levels distinguished by the word length, and the values were based on the frequency level of the words. For the Genetic Algorithm, an Adam chromosome was constructed with 315 binary genes, with each gene representing one word. The population size was considered 50 for each generation. Twenty-five percent of the new generation were selected as the fittest solutions of the past generation, which were passed directly to the new generation (Elite Selection), and 75 percent were the result of crossover. Among this 75 percent, there was a 20 percent chance of mutation. A uniform crossover with a mixed probability of 0.5 was used for creating new offsprings. A partial shuffle mutation with the mutation probability rate of 20 percent was used. The iteration number (generation number) was chosen as the termination method of the algorithm. However, it was also possible to terminate the algorithm based on time. A time-based termination strategy guaranteed that the algorithm took the same time to initiate across all the devices. However, for slower devices, the solution could be less optimal. On the other hand, the iteration-based termination offered the same level of performance by the algorithm.

The algorithm was set to terminate after the 100th generation. It returned the fittest chromosome, which was converted to the words. This problem was solved using a laptop (MSI Stealth Pro) with a processor of Intel Core i7 2.80GHz, 16GB RAM, 8 GB GeForce GTX 1070 graphics card. It took, on average, 1070 milliseconds to finish the algorithm. However, for AWS cloud computing service with 3 GB of RAM, it took on average 168 milliseconds, which was more than six times faster. It is important to note that although 1 second is acceptable for the user, the local PC that tested the algorithm had relatively powerful specifications. When the algorithm is performed on mobile devices or any device with low computation power, cloud computing provides a faster and more accurate solution.

#### **4.2. Sensitivity Analysis of the Genetic Algorithm**

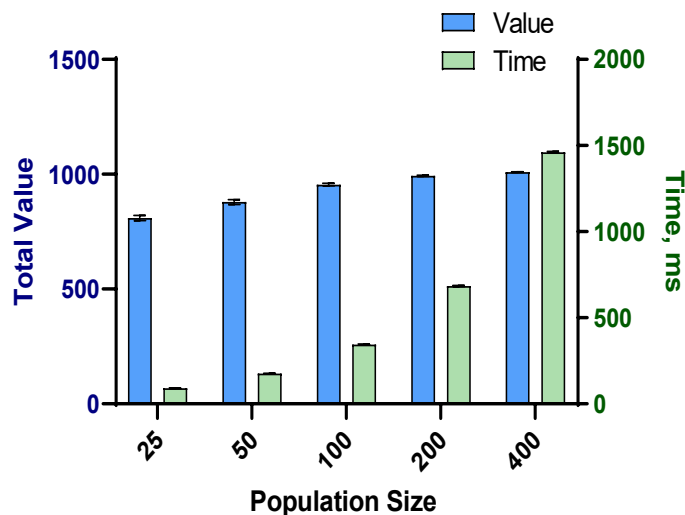
In this section, the efficiency of the structural choices in the genetic algorithm is analyzed. A series of tests were conducted in the cloud-computing version of the optimization model. The first analysis investigated the performance of the GA for different numbers of generations concerning the objective function (total value) and the computation time (milliseconds). For each iteration, the algorithm was performed ten times, and the averages of total value and computation time were recorded. Figure 14 depicts this and shows the linear relationship between the time and the number of generations. However, the optimal value reached an asymptote.



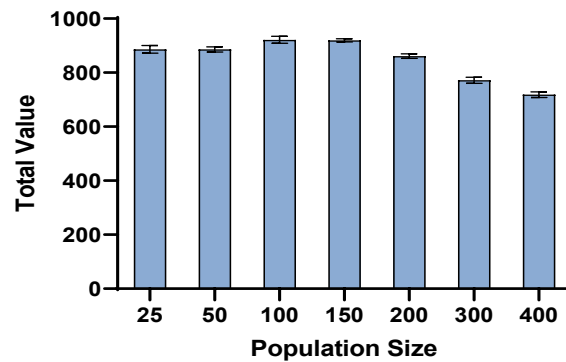


**Figure 14.** The average performance of the model for different generation numbers of the GA

Figure 15 illustrates a similar analysis of the size of the population. The model was executed 20 times for each population, and the average was recorded. As the size of the population grew, the computation time increased. However, the total value reached an asymptote. The important question was which population size was more effective for a fixed amount of computation time. To answer this question, the GA's termination was set to 200 milliseconds, the algorithm with each population size was executed 20 times, and the average total value and computation time was recorded and shown in Figure 16. It shows that the population sizes of 100 and 150 were the most efficient, and when the population size was increased further, the total value started to drop.

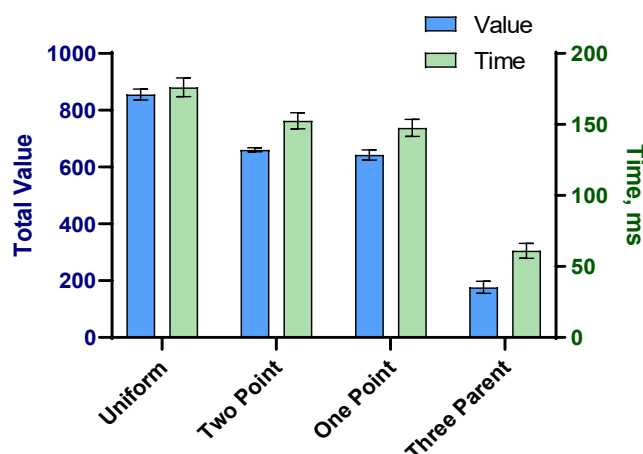


**Figure 15.** The average performance of the model for different population sizes of the GA



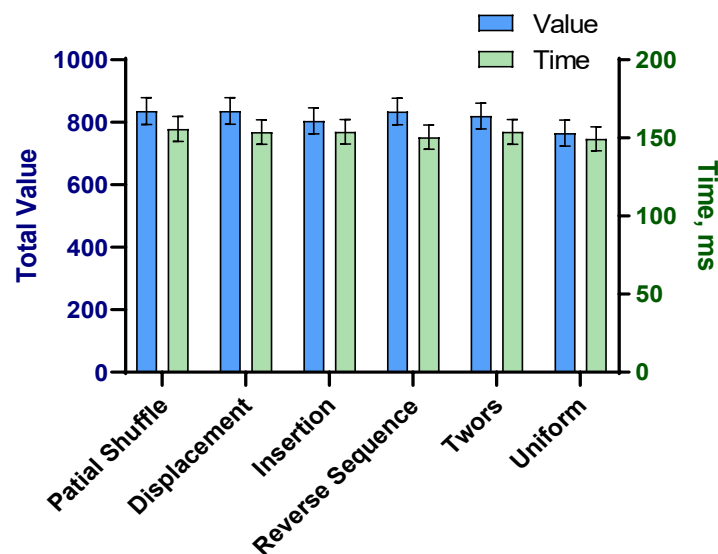
**Figure 16.** Average total value for different population sizes in a fixed time (200ms)

Another analysis was carried out on the performance of the model with different crossover functions. Four crossover functions were studied. In Uniform crossover (with the mixed probability of 0.5), each gene of the offspring was taken from one parent based on a 50-50 percent chance. Therefore, each parent's contribution to the offspring was around 50 percent. In One-Point crossover, one point in both parents was selected randomly, and all the genes beyond that point were swapped between the two parents to produce the two offsprings. In Two-Point crossover, two points were selected in parent chromosomes, and the genes between these two points were swapped between the two parents. In the Three-Parent crossover, the genes of two of the parent chromosomes were compared, and each gene that was the same in the two parents was transferred to the offspring. However, for the different genes, the gene from the third parent was chosen for the offspring. The algorithm with each crossover was performed ten times, and the average numbers were calculated. The results are presented in Figure 17. It shows the advantage of Uniform crossover compared to Two-Point, One-Point, and Three-Parent crossovers. Despite the longer time required to perform Uniform crossover, the increase in performance was considerable.



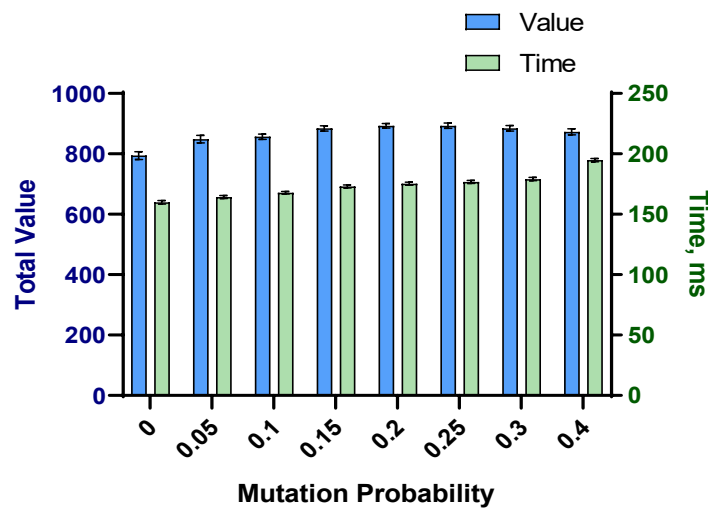
**Figure 17.** Comparison of four crossover functions

The same analysis was performed for the mutation types. Five mutation functions were examined by executing the model 30 times for each mutation type and recording the averages. In Partial Shuffle mutation, two points were randomly chosen, and the sequence between these two points was shuffled. In Displacement mutation, one part of the chromosome (a sequence of genes) was chosen randomly, and then this part was removed and placed at a random position in the chromosome. In Insertion mutation, a single gene was randomly selected and removed from the chromosome and then placed in another random position. In Reverse Sequence mutation, a random part of the chromosome was chosen (a sequence of genes), and their sequence was reversed. Finally, In the Twors mutation, two randomly chosen genes swapped positions. The results of the tests on mutation functions are presented in Figure 18. The results were close, and Partial Shuffle mutation performed slightly better than other functions in terms of the total value. The computation times were close as well, with a minimum of 159 milliseconds and a maximum of 164 milliseconds.



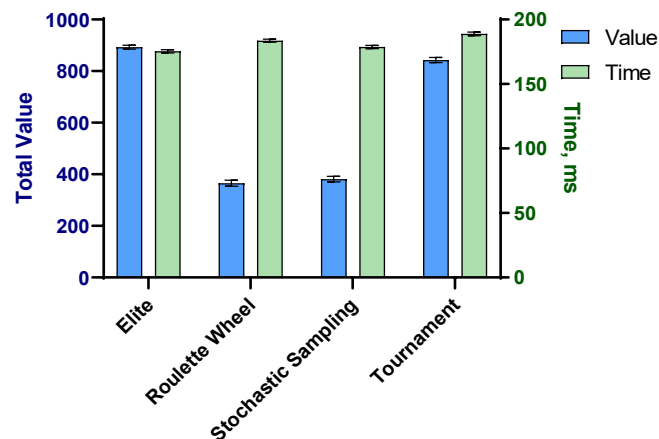
**Figure 18.** Comparison of five mutation functions

Another parameter related to mutation was the mutation probability. For finding the optimal mutation probability, eight probabilities were tested by executing the algorithm 30 times for each probability and calculating the average total values and computation times. The results are presented in Figure 19. It shows that 20 and 25 percent mutation rates produced higher total value. However, the 20 percent mutation rate was slightly faster.



**Figure 19.** Comparison of the performance of the model with different mutation probabilities

The selection type of chromosomes for passing to the next generation was also studied. Four selection types were compared. The first one was Elite selection, in which the fittest chromosomes were selected for passing to the next generation. Roulette Wheel was a fitness-proportionate selection method, in which the probability of selection for each chromosome was based on its fitness. Therefore, the fitter individuals had a higher probability of being selected. Stochastic Universal Sampling was also a fitness-proportionate selection method, which ensured that this proportionality was kept in the selection because the classic Roulette Wheel selection did not guarantee this. In Tournament selection, several competitions were held with a number of chromosomes, and the winning chromosomes was selected as the fittest chromosomes. The key in Tournament selection was the size of the competitions. If the size was small, weak chromosomes had more chance to be selected, and if the size was big, weaker chromosomes had less chance. Each selection type was tested by executing the algorithm 30 times to get the average performance. The results are presented in Figure 20. It shows that the Elite selection method resulted in better total values and faster computation time.



**Figure 20.** Comparison of four selection functions

### 4.3. Usability Experiments

The first usability study was carried out mainly to test the functionality and interactivity of the games. The obtained feedback was valuable in building the second design iteration of the system. The table below contains the most important feedbacks received from the first usability study.

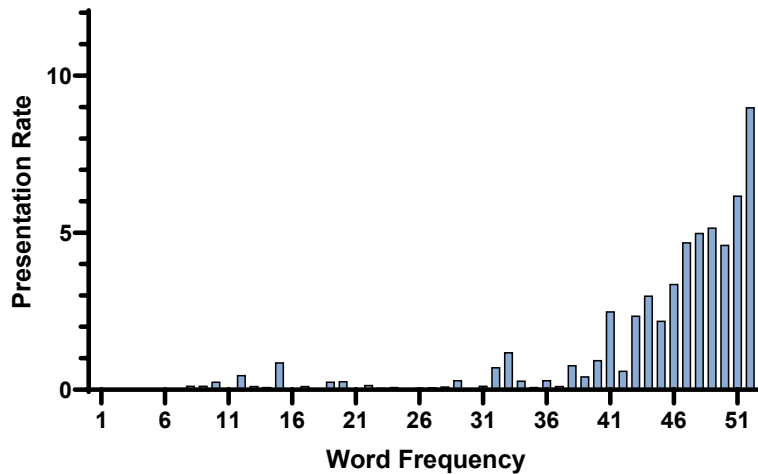
**Table 5.** The collected feedback from the first study

| Feedback  | Action  |
|---|---|
| Playing time of the games were not uniform                          | SD of time decreased from 331s (M=388) in the first experiment to 62 (M=477) in the second experiment   |
| Scores of the games were not uniform                                | SD of scores decreased from 393 (M=525) in the first experiment to 347 (M=626) in the second experiment   |
| Thanksgiving game (AWSR task) was difficult to play                 | <ul style="list-style-type: none"> <li>- Multiple immediate visual and auditory feedbacks were added to the interactions.</li> <li>- The speed of the game was decreased.</li> <li>- A replay function was added for listening to the pronunciation of the words again</li> </ul> |
| The original task in the Easter game was not effective              | This task was completely removed and replaced by a task based on the Reading Accelerated Program (RAP) [25]   |
| Some of the reading materials were similar to the prior sessions    | It was intentionally designed in this way as blocked practice at the first stage of automaticity acquisition, particularly for young children this leads to better results [45]   |
| Halloween game (AWMR task) was difficult for ESL children           | This problem was not reported when testing with native English speakers in the second experiment  |
| Speech Recognition in the Christmas game (AWD task) was not precise | The speech Recognition system was improved by limiting the recognition possibilities to only active words in the game scene   |

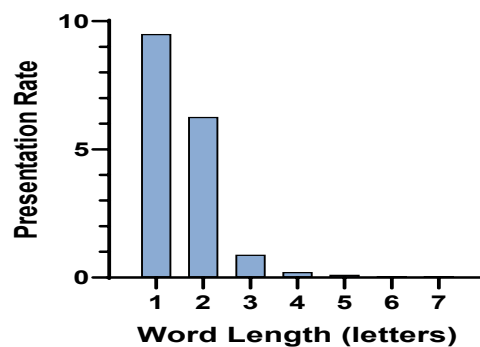
The second usability study was carried out with the improved version of the games. The first analysis was made on the behavior of the optimization model to determine if it produced the expected content. The number of the presentation for each reading material (here called “Presentation”) and each wrong answer of the users (here called “Error”) was recorded for each reading material.

The number of Presentations was analyzed relative to the frequency of the words in the language. The frequency was chosen as the Value parameter of the words in the optimization model. These frequencies were classified into 53 levels. Since the games were using the same optimization model with minor tweaks, the data of the three games were combined to obtain a more statistically significant data. It was expected from the system to start by selecting the words or word lists having the highest frequency level. However, because the number of words in each frequency level was not equal, the number of “Presentations” had to be moderated. The number of “Presentations” for each frequency level was divided by the number of words (or word lists) existing at that level. The result is shown in Figure 21. Since users often played the game only once or twice, it was expected that the model selects more often the words having the highest frequency levels and less often the words with the next neighboring frequency levels. Similarly, the model prioritized the content with lower difficulty levels, which was marked by word length. Figure 22 illustrates this tendency. However, the number of

Presentations for each word length level had to be moderated by the number of words in the level. This was important because, for example, the only single-letter words in the English Language are “a” and “I”. Therefore, the number of Presentations in each word length level was divided by the number of words (or word lists) at that level.

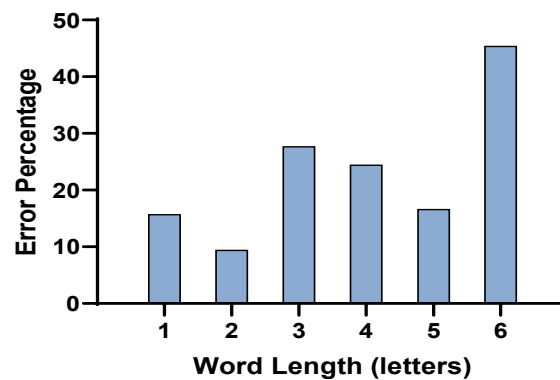


**Figure 21.** Presentation rate of word frequency levels



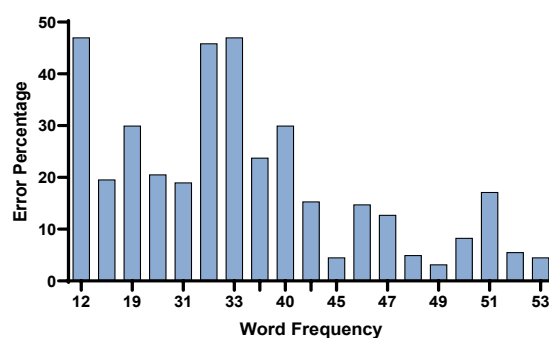
**Figure 22.** Presentation rate of word length levels

Word length is not the only attribute that can be used as an indicator of word difficulty. For the investigating this choice, the data relating to the wrong answers (error rate) was analyzed. To reach statistical significance, these automatically recorded data were combined across the three games that dealt with single words. Letter-number was selected for the weight (difficulty) parameter in the optimization model. The error rates (percentage of user errors in the number of presentations) for each letter-number are presented in Figure 23. As it appears in the diagram, word-length was not a perfect predictor of word difficulty. The linear regression analysis also confirmed that letter-number was not a statistically significant predictor of word difficulty ( $F(1, 4) = 3.90$ ,  $p = .119$ ), and it accounted for only 36.7% of the variability in error percentages. The lack of statistical significance in the linear regression model may be due to the low number of samples (in this case, the letter numbers).



**Figure 23.** The percentage of errors for different word lengths.

Moreover, the same analysis was carried out for the frequency of words. The result showed that the least frequent words produced more errors. Figure 24 shows this inverse relationship between word frequency and the percentage of errors. The linear regression analysis showed that word frequency was a statistically significant predictor of error percentage ( $F(1, 17) = 16.64, p = .001$ ). However, it only accounted for 46.5% of the variability in error percentages. Therefore, neither word length (R-Squared = 36.7%) nor word frequency (R-Squared = 46.5%) was sufficiently accurate predictors of word difficulty. To build a more reliable word-difficulty prediction model, a combination of lexical information was required. This included lexical variables such as word length, word frequency, consistency of mapping between graphemes and phonemes, and the player's grade level. It is also worth mentioning that the frequency levels with less than 20 Presentations were discarded for the diagram below because they were too sparse to be reliable as an average of the error rates.



**Figure 24.** The percentage of errors for each frequency level

Furthermore, the additional data were collected from the questionnaires presented to the users. The SUS test assessed the usability. Each question was represented by a score ranging from zero to four. Zero indicated the lowest usability, while four indicated the highest usability. The overall usability score was calculated for both experiments separately. The obtained scores were 75 percent for the first study and 80 percent for the second study. This increase was expected after improving the games according to the feedback received from the first study. The second questionnaire in this study was the Situational Motivation Scale (SIMS) to evaluate the variables of Intrinsic Motivation, Identified Regulation, External Regulation, and

Amotivation. The scores for each of these four variables were calculated and presented in Table 6. In general, participants were more intrinsically motivated in the first study. The biggest gap between the first and the second study came from the external regulation, which was significantly higher for the second usability study. This was expected, as participants' average age in the second study (23.9) was significantly higher than the first study (14.4).

**Table 6.** Situational Motivation Scale for the two usability studies

| Motivation Type              | First Study | Second Study |
|------------------------------|-------------|--------------|
| <i>Intrinsic Motivation</i>  | 77%         | 75%          |
| <i>Identified Regulation</i> | 75%         | 70%          |
| <i>External Regulation</i>   | 37%         | 56%          |
| <i>Extrinsic Regulation</i>  | 40%         | 38%          |

Finally, in the last questionnaire, the participants were asked to rate the games based on their level of difficulty and amusingness. The most difficult game was AWSR (Thanksgiving) with a rating of 3.75 out of 5, and the easiest game was the APR (Easter) with a rating of 4.5. The least amusing game was AWD (Christmas) with a rating of 3.81, and the most amusing was the AWMR (Halloween) game with a rating of 4.25.

#### 4.4. Regression Models for Auditory Word Recognition

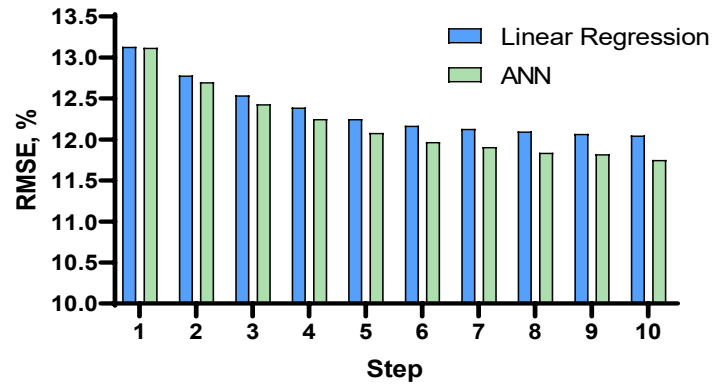
This section presents the forward stepwise analysis of the auditory lexical decision using linear regression and artificial neural networks. The accuracy of the two models was compared with Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Table 7 shows that the number of phonemes in the word was the most important variable in predicting the auditory lexical decision's reaction time. Frequency, number of homophones, number of syllables, and is-lemma (is the word a lemma) were the next important variables. The two forward stepwise models resulted in eight similar variables and two different variables.

**Table 7.** Forward stepwise analysis of the auditory lexical decision task

| Step | <i>Linear Regression</i> |         |         |        | <i>Artificial Neural Network</i> |         |        |
|------|--------------------------|---------|---------|--------|----------------------------------|---------|--------|
|      | Added Variable           | $\beta$ | RMSE, % | MAE, % | Added Variable                   | RMSE, % | MAE, % |
| 1    | nbphons                  | 0.722   | 13.13   | 10.56  | nbphons                          | 13.12   | 10.52  |
| 2    | Frequency                | -0.238  | 12.78   | 10.24  | Frequency                        | 12.70   | 10.15  |
| 3    | nbhomoph                 | -0.157  | 12.54   | 10.00  | nbhomoph                         | 12.43   | 9.88   |
| 4    | nbsyll                   | -0.304  | 12.39   | 9.88   | nbsyll                           | 12.25   | 9.75   |
| 5    | islem                    | -0.126  | 12.25   | 9.75   | islem                            | 12.08   | 9.58   |
| 6    | Freq_PG                  | -0.084  | 12.17   | 9.70   | Freq_PG                          | 11.97   | 9.51   |
| 7    | Freq_F_Tok_GP            | -0.078  | 12.13   | 9.66   | SylphonFreqTy                    | 11.91   | 9.46   |
| 8    | voisphon                 | 0.101   | 12.10   | 9.63   | Freq_F_PG                        | 11.84   | 9.40   |
| 9    | PhonFreqTo               | -0.066  | 12.07   | 9.61   | SylphonFreqTo                    | 11.82   | 9.35   |
| 10   | SylphonFreqTy            | 0.077   | 12.05   | 9.59   | Freq_F_Tok_GP                    | 11.75   | 9.30   |

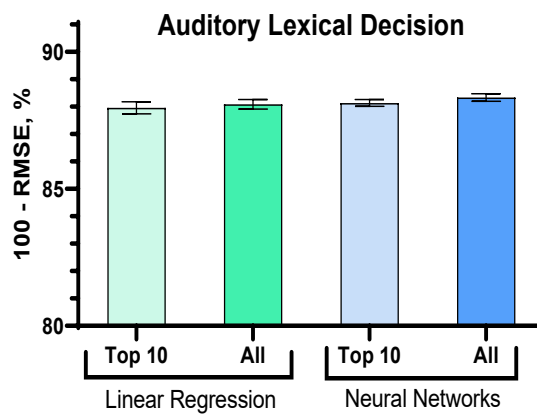


Figure 25 illustrates the progress of the forward stepwise analysis on the measure of Root Mean Square Error (RMSE). The diagram shows that as the steps progressed, the difference between the two models became bigger, and ANN produced a lower error rate. Nevertheless, the difference between the two approaches remained marginal.



**Figure 25.** The Root Mean Square Error percentage for the ten steps of forward stepwise analysis of the auditory lexical decision task

The average error rate of linear regression from the 10-fold cross-validation was  $RMSE_{LR} = 12.05\%$  ( $SD = 0.22$ ) and  $MAE_{LR} = 9.60\%$  ( $SD = 0.16$ ), while the average error rate of ANN (with two hidden layers of 20 neurons) was  $RMSE_{ANN} = 11.87\%$  ( $SD = 0.12$ ) and  $MAE_{ANN} = 9.42\%$  ( $SD = 0.17$ ). The difference of the two models was 0.18 percent in both of the accuracy metrics. This similarity in the linear regression and ANN performance means that either the task has low non-linearity, or the data size is not sufficient for detecting any complex relationship. Furthermore, when the models were trained with all the 55 lexical variables, 10-fold cross-validation error rates of the linear regression model were  $MAE_{LR-All} = 9.49$  ( $SD = 0.17$ ),  $RMSE_{LR-All} = 11.92$  ( $SD = 0.17$ ). These values for the ANN model were  $MAE_{ANN-All} = 9.25$  ( $SD = 0.14$ ),  $RMSE_{ANN-All} = 11.67$  ( $SD = 0.14$ ). Figure 26 illustrates these results.



**Figure 26.** Comparison of the accuracy of the linear regression and ANN models

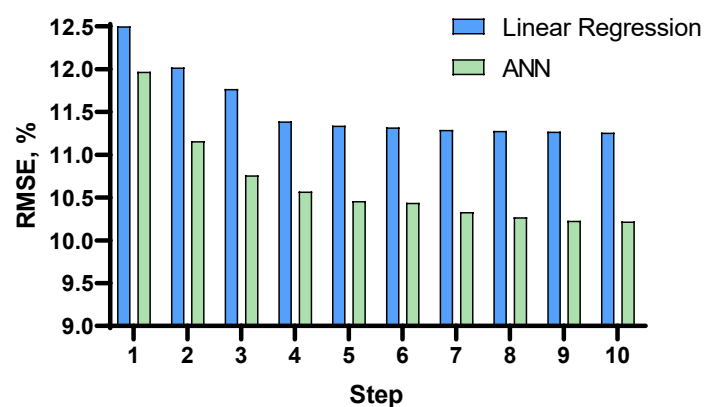
#### 4.5. Regression Models for Visual Word Recognition

Table 8 presents the forward stepwise analysis results on visual lexical decision using linear regression and ANN. The word frequency was the most important variable in predicting the reaction time of the visual lexical decision task. The two forward stepwise models resulted in five similar variables and five different variables. The five similar variables were Frequency, number of homographs, number of syllables, phonological unicuity point, and number of morphemes.

**Table 8.** Forward stepwise analysis of the visual lexical decision task

| Step | Linear Regression |         |         |        | Artificial Neural Network |         |        |
|------|-------------------|---------|---------|--------|---------------------------|---------|--------|
|      | Added Variable    | $\beta$ | RMSE, % | MAE, % | Added Variable            | RMSE, % | MAE, % |
| 1    | Frequency         | -0.561  | 12.50   | 9.95   | Frequency                 | 11.97   | 9.40   |
| 2    | nbhomogr          | -0.173  | 12.02   | 9.55   | puphon                    | 11.16   | 8.70   |
| 3    | nbsyll            | 0.201   | 11.77   | 9.35   | nbhomogr                  | 10.76   | 8.39   |
| 4    | puphon            | -0.231  | 11.39   | 9.00   | nbsyll                    | 10.57   | 8.25   |
| 5    | LetFreqTy         | 0.044   | 11.34   | 8.94   | BigrFreqTy                | 10.46   | 8.12   |
| 6    | Freq_F_Tok_PG     | -0.073  | 11.32   | 8.92   | islem                     | 10.44   | 8.05   |
| 7    | PhonFreqTo        | 0.050   | 11.29   | 8.90   | Freq_F_PG                 | 10.33   | 7.97   |
| 8    | FreqTok_PG        | 0.041   | 11.28   | 8.88   | SylphonFreqTo             | 10.27   | 7.88   |
| 9    | posregTo_GP       | 0.034   | 11.27   | 8.88   | BiphonFreqTy              | 10.23   | 7.86   |
| 10   | nbmorph           | -0.031  | 11.26   | 8.87   | nbmorph                   | 10.22   | 7.82   |

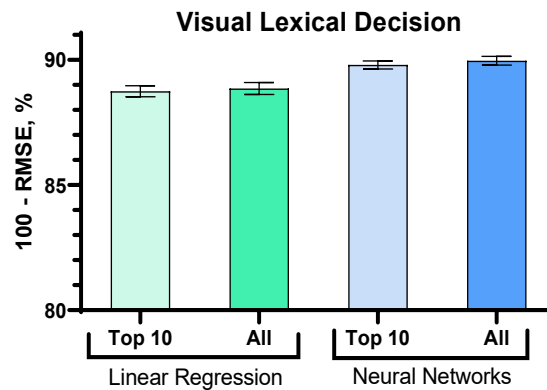
Figure 27 illustrates the progress of the forward stepwise analysis on the measure of RMSE. Similar to the auditory lexical decision, the diagram shows that as the steps progressed, the difference between the two models became bigger, and ANN produced a lower error rate than linear regression. However, the difference between the two models remained marginal.



**Figure 27.** The Root Mean Square Error percentage for the ten steps of forward stepwise analysis of the visual lexical decision task

The average error rates of the 10-fold cross-validation process was  $RMSE_{LR} = 11.26\%$  ( $SD = 0.22$ ) and  $MAE_{LR} = 8.87\%$  ( $SD = 0.17$ ), while the average error rate of the ANN model (with two hidden layers of 20 neurons) was  $RMSE_{ANN} = 10.21\%$  ( $SD = 0.16$ ) and  $MAE_{ANN} = 7.92\%$

(SD = 0.14). The difference was  $RMSE_{DIFF} = 1.05\%$  and  $MAE_{DIFF} = 0.95\%$ . Furthermore, when the models were trained with all the 55 lexical variables, 10-fold cross-validation error rates of the linear regression model were  $MAE_{LR-All} = 8.87$  (SD = 0.18),  $RMSE_{LR-All} = 11.15$  (SD = 0.24). These values for the ANN model were  $MAE_{ANN-All} = 7.82$  (SD = 0.16),  $RMSE_{ANN-All} = 10.04$  (SD = 0.17). Figure 28 illustrates these results.



**Figure 28.** Comparison of the accuracy of the linear regression and ANN models

#### 4.6. Regression Models for Word Spelling

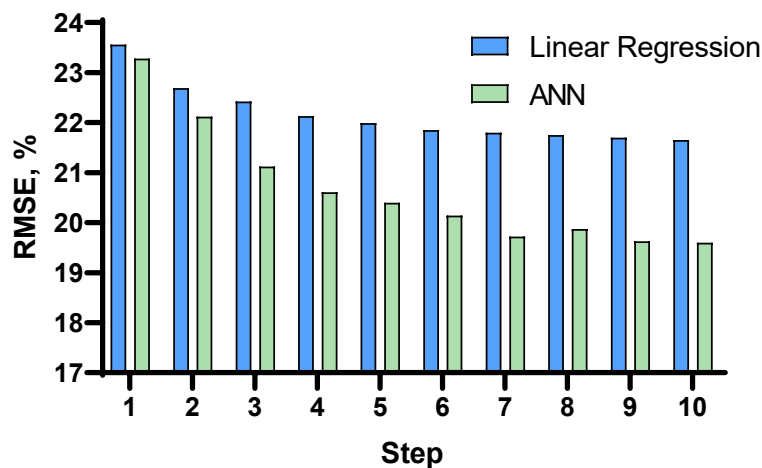
Table 9 presents the forward stepwise analysis results on the spelling task using linear regression and ANN. It should be noted that each word had five spelling success rates, representing the five primary school grades. Therefore, the number of the dataset was 56865. The grade variable was included in the models by default. The two forward stepwise models resulted in six similar variables and four different variables. The six similar variables were the number of phoneme-grapheme irregularity, frequency, minimum consistency of phoneme-grapheme associations, consistency of the last phoneme-grapheme association, number of grapheme-phoneme irregularities, and the average frequency of phonological syllables.

**Table 9.** Forward stepwise analysis of the spelling task

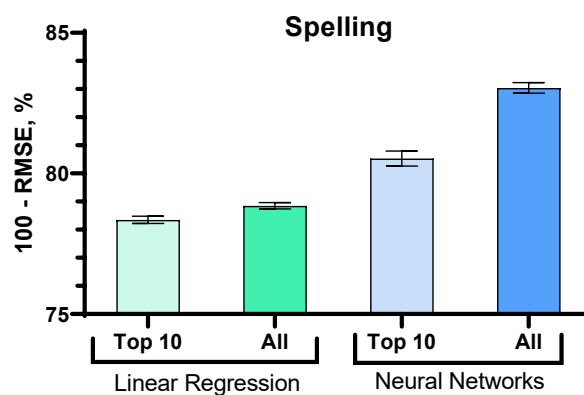
| Step | Linear Regression |         |         |        | Artificial Neural Network |         |        |
|------|-------------------|---------|---------|--------|---------------------------|---------|--------|
|      | Added Variable    | $\beta$ | RMSE, % | MAE, % | Added Variable            | RMSE, % | MAE, % |
| 1    | countregTo_PG     | -0.264  | 23.56   | 19.26  | countregTo_PG             | 23.28   | 18.49  |
| 2    | Frequency         | 0.167   | 22.69   | 18.52  | Frequency                 | 22.12   | 17.48  |
| 3    | minfreqgraph_PG   | 0.206   | 22.42   | 18.23  | minfreqgraph_PG           | 21.12   | 16.71  |
| 4    | Freq_F_PG         | -0.135  | 22.13   | 17.93  | countregTy_GP             | 20.61   | 16.07  |
| 5    | voisorth          | 0.078   | 21.99   | 17.80  | nblettes                  | 20.40   | 15.67  |
| 6    | TrigrFreqTo       | 0.108   | 21.85   | 17.67  | SylphonFreqTy             | 20.14   | 15.51  |
| 7    | nbmorph           | 0.056   | 21.80   | 17.63  | minfreqgraphTok_PG        | 19.72   | 15.34  |
| 8    | SylphonFreqTy     | -0.055  | 21.75   | 17.58  | Freq_PG                   | 19.87   | 15.19  |
| 9    | countregTy_GP     | -0.065  | 21.70   | 17.54  | Freq_F_PG                 | 19.63   | 15.05  |
| 10   | complexmoygraph   | -0.054  | 21.65   | 17.5   | Freq_F_Tok_PG             | 19.57   | 15.01  |

Figure 29 illustrates the progress of the forward stepwise analysis on the measure of RMSE. Similar to the two other tasks, the diagram shows that as the steps progressed, the difference between the two models became bigger, and ANN produced a lower error rate than linear regression. The difference between the two models remained marginal. However, in comparison to the two lexical decision tasks, the difference was bigger. The average error rates of the 10-fold cross validation of the linear regression model was  $RMSE_{LR} = 21.65\%$  (SD = 0.13) and  $MAE_{LR} = 17.50\%$  (SD=0.12), while the average error rate of the ANN model (with two hidden layers of 20 neurons) was  $RMSE_{ANN} = 19.47\%$  (SD = 0.27) and  $MAE_{ANN} = 15.17\%$  (SD = 0.25). The difference was  $RMSE_{DIFF} = 2.18\%$  and  $MAE_{DIFF} = 2.33\%$ .

Furthermore, when the models were trained with all the 55 lexical variables, 10-fold cross-validation error rates of the linear regression model were  $MAE_{LR-All} = 17$  (SD = 0.1),  $RMSE_{LR-All} = 21.15$  (SD = 0.11). These values for the ANN model were  $MAE_{ANN-All} = 13.26$  (SD = 0.14),  $RMSE_{ANN-All} = 16.96$  (SD = 0.19). Figure 30 illustrates these results.



**Figure 29.** The Root Mean Square Error percentage for the ten steps of forward stepwise analysis of the spelling task



**Figure 30.** Comparison of the accuracy of the linear regression and ANN models

The EOLE database [3] had the advantage of being tested with children of the five primary grades in France. Each word in the database included the average spelling success-rate for each of the five grades. The result of the forward stepwise analysis for each grade is presented in Table 10. The results showed that the number of phoneme-grapheme irregularities (countregTo\_PG) was the most important variable for all five grades. The next important variable was word frequency. However, in the first grade, word frequency had less importance, and by looking at the standardized coefficients, it can be noticed that grade by grade, it became more and more important. On the other hand, the voisorth variable, which is the number of orthographic neighbors, was important initially, but it became less important with each grade. The minimum consistency of phoneme-grapheme associations was a stable variable across the five grades. The average trigram frequency in the word (TrigrFreqTo) seemed to be increasing in importance by each grade. Lastly, final phoneme-grapheme consistency was a stable predictor of spelling success in the French language. However, its importance marginally decreases by each grade.

**Table 10.** Grade-based forward stepwise linear regression for the spelling task

| <i>STEP</i> | <i>GRADE 1</i>  |         | <i>GRADE 2</i>  |         | <i>GRADE 3</i>  |         | <i>GRADE 4</i>  |         | <i>GRADE 5</i>  |         |
|-------------|-----------------|---------|-----------------|---------|-----------------|---------|-----------------|---------|-----------------|---------|
|             | Variable        | $\beta$ | Variable        | $\beta$ | Variable        | $\beta$ | Variable        | $\beta$ | Variable        | $\beta$ |
| 1           | countregTo_PG   | -0.14   | countregTo_PG   | -0.29   | countregTo_PG   | -0.33   | countregTo_PG   | -0.34   | countregTo_PG   | -0.31   |
| 2           | voisorth        | 0.13    | Frequency       | 0.16    | Frequency       | 0.22    | Frequency       | 0.25    | Frequency       | 0.27    |
| 3           | minfreqgraph_PG | 0.24    | voisorth        | 0.13    | minfreqgraph_PG | 0.28    | TrigrFreqTo     | 0.12    | TrigrFreqTo     | 0.14    |
| 4           | Freq_F_PG       | -0.17   | minfreqgraph_PG | 0.27    | Freq_F_PG       | -0.15   | minfreqgraph_PG | 0.25    | minfreqgraph_PG | 0.24    |
| 5           | posregTo_PG     | -0.13   | Freq_F_PG       | -0.16   | TrigrFreqTo     | 0.12    | Freq_F_PG       | -0.14   | Freq_F_PG       | -0.12   |
| 6           | Freq_GP         | 0.10    | TrigrFreqTo     | 0.10    | voisorth        | 0.09    | voisorth        | 0.07    | nbmorph         | 0.09    |
| 7           | complexmoygraph | -0.09   | posregTo_PG     | -0.14   | nbmorph         | 0.09    | nbmorph         | 0.08    | nbhomogr        | 0.07    |
| 8           | Frequency       | 0.08    | complexmoygraph | -0.09   | SylphonFreqTy   | -0.09   | SylphonFreqTy   | -0.08   | posregTy_PG     | 0.10    |
| 9           | FreqTok_PG      | 0.11    | Freq_GP         | 0.09    | countregTy_GP   | -0.08   | posregTy_PG     | 0.09    | SylphonFreqTy   | -0.08   |
| 10          | nblettes        | -0.08   | posregTy_PG     | 0.09    | posregTy_PG     | 0.08    | nbhomogr        | 0.06    | countregTy_GP   | -0.05   |

#### 4.1.Regression Model for Word Naming

For the word-naming task, only 1481 words were available in the database. Hence, the ANN model could not produce accurate results, and only the forward stepwise linear regression was carried out. The results are presented in Table 11. The number of letters was the most important variable. The consistency of the initial grapheme-phoneme association was the next important variable. The third variable was the frequency. The three next variables were again related to the initial grapheme-phoneme and phoneme-grapheme associations. It should be noted that this task was performed with only monosyllabic, monomorphemic words, and the results may not generalize to all the words. The results from the 10-fold cross validation produced the average error rates of 16.14% for RMSE (SD = 1.07) and 13.02% MAE (SD = 0.78). Furthermore, 10-fold cross-validation of the linear regression including all the 55 lexical variables produced 16.09% for RMSE (SD = 0.94) and 12.83% MAE (SD = 0.81).

**Table 11.** Forward stepwise analysis of the word-naming task

| <i>Step</i> | <i>Linear Regression</i> |         |          |                 |         |        |
|-------------|--------------------------|---------|----------|-----------------|---------|--------|
|             | Added Variable           | $\beta$ | R Square | R Square Change | RMSE, % | MAE, % |
| 1           | nblettres                | 0.722   | 0.055    | 0.055           | 17.315  | 14.30  |
| 2           | Freq_I_Tok_GP            | -0.238  | 0.094    | 0.039           | 16.955  | 13.80  |
| 3           | Frequency                | -0.157  | 0.112    | 0.018           | 16.795  | 13.68  |
| 4           | Freq_I_GP                | -0.304  | 0.124    | 0.012           | 16.684  | 13.47  |
| 5           | Freq_I_PG                | -0.126  | 0.148    | 0.024           | 16.458  | 13.22  |
| 6           | Freq_I_Tok_PG            | -0.084  | 0.167    | 0.019           | 16.282  | 13.08  |
| 7           | nbhomogr                 | -0.078  | 0.174    | 0.007           | 16.224  | 13.08  |
| 8           | Freq_PG                  | 0.101   | 0.180    | 0.006           | 16.170  | 13.06  |
| 9           | PhonFreqTo               | -0.066  | 0.187    | 0.007           | 16.103  | 12.97  |
| 10          | posregTy_PG              | 0.077   | 0.194    | 0.007           | 16.041  | 12.94  |

## 5. Conclusion

In this research, an intelligent gamified training program for young dyslexics were developed. It focused on the automaticity aspect of reading in languages with opaque orthographies. The training design was based on the principles of automaticity acquisition. Four games were designed to enhance the automaticity of reading and offer a basis for expanding vocabulary knowledge. An optimization model was developed and integrated into the games to offer maximized educational value, taking into account the content difficulty level. The optimization model was solved with a genetic algorithm. This algorithm was analyzed in detail to ensure its efficiency.

Two usability studies were conducted to validate the approach. The data collected from playing the games showed that the optimization model produced the expected content for the games. The contents with the highest value and lowest difficulty were preferred more frequently. Word frequency was a better predictor of word difficulty than letter-number. However, neither letter-number nor word frequency was a perfect predictor of lexical difficulty. Therefore, it was concluded that a combination of lexical information was required for estimating lexical difficulty.

Building lexical difficulty models are complex because many variables are involved. The importance of each variable and their intertwined relationships are not fully known. In addition, the lexical difficulty varies from task to task. A word that is easy to read is not necessarily easy to write and vice versa. In this thesis, four lexical tasks of auditory and visual word recognition (lexical decision), word naming, and spelling were targeted. The databases from lexical mega-studies were combined with lexical and sub-lexical databases. This resulted in 55 lexical and sub-lexical variables, and 15842 words for auditory word recognition, 25776 words for visual word recognition, 1481 words for word naming, and 11373 words for spelling task. Linear regression and artificial neural networks were used for building the lexical difficulty models of each task.

For identifying the most important set of variables, the forward stepwise analysis was carried out for both neural networks and linear regressions. The top 10 variables for each of the lexical tasks were identified, and the accuracies of the models were compared. The neural

network models performed marginally better than the linear regression models. The auditory word recognition task reached 88.13% accuracy calculated by the root mean square error percentage. The visual word recognition reached 89.79%, and the spelling task reached 80.53%. The word-naming task that did not have sufficient data was analyzed only by linear regression, reaching 83.86% accuracy.

The spelling models showed that variables such as the number of irregular phoneme-grapheme associations, word frequency, and minimum phoneme-grapheme associations were important in predicting spelling success. The auditory lexical decision task depended highly on the number of phonemes, homophones, and word frequency. For the visual lexical decision task, word frequency, number of homographs, and syllables were among the most important variables. Finally, for the word naming of monosyllabic words, the number of letters, word frequency, and initial grapheme-phoneme consistency were important.

Finally, the results of the system usability and situational motivation scale were promising. The overall score of the usability test reached 75% in the first usability study and 80% for the second study.

## 6. References

- [1] Z. Breznitz and M. Nevat, *The reading acceleration program (RAP)*. Israel: University of Haifa (The Edmond J. Safra Brain Research Center for the Study of Learning Disabilities), 2004.
- [2] L. Ferrand *et al.*, “MEGALEX: A megastudy of visual and auditory word recognition,” *Behav Res*, vol. 50, no. 3, pp. 1285–1307, Jun. 2018, doi: 10.3758/s13428-017-0943-1.
- [3] B. Pothier and P. Pothier, *Echelle d'acquisition en orthographe lexicale: pour l'école élémentaire: du CP au CM2*. Retz, 2004.
- [4] L. Ferrand *et al.*, “Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from Chronolex,” *Frontiers in psychology*, vol. 2, p. 306, 2011.
- [5] B. New, C. Pallier, L. Ferrand, and R. Matos, “Une base de données lexicales du français contemporain sur internet: LEXIQUE™//A lexical database for contemporary french: LEXIQUE™,” *L'année psychologique*, vol. 101, no. 3, pp. 447–462, 2001.
- [6] B. New, “Lexique 3: Une nouvelle base de données lexicales,” 2006.
- [7] M. Gimenes, C. Perret, and B. New, “Lexique-Infra: grapheme-phoneme, phoneme-grapheme regularity, consistency, and other sublexical statistics for 137,717 polysyllabic French words,” *Behav Res*, May 2020, doi: 10.3758/s13428-020-01396-2.
- [8] D. Guckenberger, K. C. Uliano, and N. E. Lane, “Training high performance skills using above real-time training,” NASA-CR-192616, NAS 1.26:192616, Jan. 1993. Accessed: Jul. 03, 2019. [Online]. Available: <https://ntrs.nasa.gov/search.jsp?R=19930015003>.
- [9] D. C. Donderi, K. K. Niall, K. Fish, and B. Goldstein, “Above-Real-Time Training (ARTT) Improves Transfer to a Simulated Flight Control Task,” *Hum Factors*, vol. 54, no. 3, pp. 469–479, Jun. 2012, doi: 10.1177/0018720812439711.
- [10] S. Dehaene, *Neurones de la lecture (Les): La nouvelle science de la lecture et de son apprentissage*. Odile jacob, 2007.
- [11] W. Schneider and A. D. Fisk, “Degree of consistent training: Improvements in search performance and automatic process development,” *Perception & Psychophysics*, vol. 31, no. 2, pp. 160–168, 1982.
- [12] A. D. Fisk, N. A. Oransky, and P. R. Skedsvold, “Examination of the role of ‘higher-order’ consistency in skill development,” *Human Factors*, vol. 30, no. 5, pp. 567–581, 1988.
- [13] P. Ginns, “Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects,” *Learning and Instruction*, vol. 16, no. 6, pp. 511–525, Dec. 2006, doi: 10.1016/j.learninstruc.2006.10.001.

- [14] M. F. St. John and D. R. Shanks, "Implicit learning from an information processing standpoint," in *How implicit is implicit learning?*, New York, NY, US: Oxford University Press, 1997, pp. 162–194.
- [15] F. Brady, "Contextual Interference: A Meta-Analytic Study," *Perceptual and Motor Skills*, vol. 99, no. 1, pp. 116–126, 2004.
- [16] J. J. Farahani, A. H. Javadi, B. V. O'Neill, and V. Walsh, "Effectiveness of above real-time training on decision-making in elite football: A dose–response investigation," in *Progress in Brain Research*, vol. 234, Elsevier, 2017, pp. 101–116.
- [17] T. Horowitz-Kraus, N. Cicchino, M. Amiel, S. K. Holland, and Z. Breznitz, "Reading improvement in English-and Hebrew-speaking children with reading difficulties after reading acceleration training," *Annals of dyslexia*, vol. 64, no. 3, pp. 183–201, 2014.
- [18] E. Nevo, S. Brande, and S. Shaul, "The Effects of Two Different Reading Acceleration Training Programs on Improving Reading Skills of Second Graders," *Reading Psychology*, vol. 37, no. 4, pp. 533–546, May 2016, doi: 10.1080/02702711.2015.1066911.
- [19] T. Horowitz-Kraus, "Improvement in non-linguistic executive functions following reading acceleration training in children with reading difficulties: An ERP study," *Trends in Neuroscience and Education*, vol. 4, no. 3, pp. 77–86, Sep. 2015, doi: 10.1016/j.tine.2015.06.002.
- [20] Z. Breznitz, S. Shaul, T. Horowitz-Kraus, I. Sela, M. Nevat, and A. Karni, "Enhanced reading by training with imposed time constraint in typical and dyslexic adults," *Nat Commun*, vol. 4, no. 1, p. 1486, Jun. 2013, doi: 10.1038/ncomms2488.
- [21] L. C. Ehri, S. R. Nunes, S. A. Stahl, and D. M. Willows, "Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis," *Review of educational research*, vol. 71, no. 3, pp. 393–447, 2001.
- [22] S. Graham and K. R. Harris, "Strategy instruction and the teaching of writing," *Handbook of writing research*, vol. 5, pp. 187–207, 2006.
- [23] K. Shabani, M. Khatib, and S. Ebadi, "Vygotsky's Zone of Proximal Development: Instructional Implications and Teachers' Professional Development.," *English language teaching*, vol. 3, no. 4, pp. 237–248, 2010.
- [24] M. Csikszentmihalyi, *Finding flow: The psychology of engagement with everyday life*. Basic Books, 1997.
- [25] N. Zandi Atashbar and F. Rahimi, "Optimization of Educational Systems Using Knapsack Problem," *International Journal of Machine Learning and Computing*, vol. 2, no. 5, p. 552, 2012.
- [26] P. Zoccolotti, M. De Luca, E. Di Pace, F. Gasperini, A. Judica, and D. Spinelli, "Word length effect in early reading and in developmental dyslexia," *Brain and Language*, vol. 93, no. 3, pp. 369–373, Jun. 2005, doi: 10.1016/j.bandl.2004.10.010.
- [27] H. S. S. L. Joseph, S. P. Liversedge, H. I. Blythe, S. J. White, and K. Rayner, "Word length and landing position effects during reading in children and adults," *Vision Research*, vol. 49, no. 16, pp. 2078–2086, Aug. 2009, doi: 10.1016/j.visres.2009.05.015.
- [28] J. J. S. Barton, H. M. Hanif, L. Eklinder Björnström, and C. Hills, "The word-length effect in reading: A review," *Cognitive Neuropsychology*, vol. 31, no. 5–6, pp. 378–412, Aug. 2014, doi: 10.1080/02643294.2014.895314.
- [29] D. A. Balota, M. J. Cortese, S. D. Sergent-Marshall, D. H. Spieler, and M. J. Yap, "Visual word recognition of single-syllable words.," *Journal of Experimental Psychology: General*, vol. 133, no. 2, p. 283, 2004.
- [30] M. J. Yap and D. A. Balota, "Visual word recognition of multisyllabic words," *Journal of Memory and Language*, vol. 60, no. 4, pp. 502–529, May 2009, doi: 10.1016/j.jml.2009.02.001.
- [31] I. Lacruz and J. R. Folk, "Feedforward and Feedback Consistency Effects for High- and Low-Frequency Words in Lexical Decision and Naming," *The Quarterly Journal of Experimental Psychology Section A*, vol. 57, no. 7, pp. 1261–1284, Oct. 2004, doi: 10.1080/02724980343000756.
- [32] M. A. D. K. Wijayathilake and R. Parrila, "Predictors of word reading skills in good and struggling readers in Sinhala," *Writing Systems Research*, vol. 6, no. 1, pp. 120–131, Jan. 2014, doi: 10.1080/17586801.2013.846844.



- [33] A. Pollatsek, M. Perea, and K. S. Binder, "The effects of 'neighborhood size' in reading and lexical decision," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 25, no. 4, pp. 1142–1158, 1999, doi: 10.1037/0096-1523.25.4.1142.
- [34] C. Mulatti, M. G. Reynolds, and D. Besner, "Neighborhood effects in reading aloud: New findings and new challenges for computational models.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 32, no. 4, pp. 799–810, 2006, doi: 10.1037/0096-1523.32.4.799.
- [35] L. Ferrand and J. Grainger, "Homophone interference effects in visual word recognition," *The Quarterly Journal of Experimental Psychology Section A*, vol. 56, no. 3, pp. 403–419, 2003.
- [36] J. Ecalte and A. Magnan, *L'apprentissage de la lecture et ses difficultés-2e éd.* Dunod, 2015.
- [37] C. P. Lim, D. Nonis, and J. Hedberg, "Gaming in a 3D multiuser virtual environment: engaging students in Science lessons," *Br J Educ Technol*, vol. 37, no. 2, pp. 211–231, Mar. 2006, doi: 10.1111/j.1467-8535.2006.00531.x.
- [38] W. Blevins, *Phonics from A to Z: A practical guide*. Scholastic Inc., 1998.
- [39] "Listes de mots des graphèmes simples et complexes." Accessed: Jun. 21, 2020. [Online]. Available: <http://monbazarcolore.com/2019/02/langage-listes-de-mots-des-graphemes-simples-et-complexes-documents-a-telecharger-gratuitement/>.
- [40] S. Davidson, "Wordnik," *charleston adv*, vol. 15, no. 2, pp. 54–58, Oct. 2013, doi: 10.5260/chara.15.2.54.
- [41] B. Lété, L. Sprenger-Charolles, and P. Colé, "MANULEX: A grade-level lexical database from French elementary school readers," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 1, pp. 156–166, Feb. 2004, doi: 10.3758/BF03195560.
- [42] B. Duckett, "The Longman Dictionary of Contemporary English (5th ed.)," *Reference Reviews*, vol. 23, no. 6, pp. 32–34, Aug. 2009, doi: 10.1108/09504120910978915.
- [43] S. Webb, "The Effects of Context on Incidental Vocabulary Learning," *Reading in a Foreign Language*, vol. 20, no. 2, pp. 232–245, Oct. 2008.
- [44] M. A. Reeve, "Sentence Comprehension-WH Questions Series," *Teachers Pay Teachers*. <https://www.teacherspayteachers.com/Product/WH-QUESTIONS-Comprehension-product-BUNDLE-2536701> (accessed Jun. 24, 2020).
- [45] J. B. Carroll, "An alternative to Juilland's usage coefficient for lexical frequencies, and a proposal for a standard frequency index (SFI)," *Computer studies in the humanities and verbal behavior*, vol. 3, no. 2, pp. 61–65, 1970.
- [46] J. Brooke, "SUS-A quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.
- [47] F. Guay, R. J. Vallerand, and C. Blanchard, "On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS)," *Motivation and emotion*, vol. 24, no. 3, pp. 175–213, 2000.

## Conclusion and Perspectives of Research

After outlining the research objectives in the first chapter of this thesis, an analytical review of the reading programs was presented Chapter 2. This review identified the gaps in knowledge and the disadvantages of the current practices of dyslexia remediation. In order to address the identified gaps, and based on the automaticity theories of dyslexia, a scientific approach was proposed in Chapter 3. The training principles of automaticity acquisition were thoroughly studied in Chapter 4. Based on this literature analysis, a training model for automaticity acquisition was developed. Chapter 5 was dedicated to the development of the stage-based gamification model for automaticity acquisition. Accomplishing these steps created a solid scientific platform for implementing the dyslexia remediation system described in Chapter 6. In this conclusion, we present the summary of the main findings, as well as the limitations and perspectives of our research.

The main objective of this research was to enhance and facilitate the reading acquisition of opaque orthographies in young dyslexic children. Technology-based or technology-assisted reading intervention systems were classified into the six instructional categories of Phonological Awareness, Phonics, Vocabulary, Comprehension, Fluency, and Multi-Component. The investigations carried out in the chapter 2 led to the main findings:

1. Automaticity was neglected in elementary grades,
2. Technology-based vocabulary programs were overlooked,
3. Fluency interventions were too short compared to the other instructional categories,
4. Fluency approaches were more neglected in languages with opaque orthographies,
5. Gamification was applied less often to fluency and comprehension,
6. Non-computer technologies were overlooked,
7. A holistic approach to fluency was lacking.

Based on the automaticity theories of dyslexia, we focused our research on enhancing the automaticity of grapheme-phoneme and phoneme-grapheme associations. Home-based training was targeted to facilitate access to the system. The Gamification concept was proposed to increase engagement and adherence to the designed training program. The intelligent adaptive models were proposed to fill the missing role of the instructors in home-based training conditions. The proposed intelligent component included an optimization model to offer the optimal content for each learner and the regression models for estimating the difficulty level of the content. The estimated difficulty levels were used as parameters in the optimization model.

Chapter 4 was dedicated to the identification and modeling of effective training principles of automaticity acquisition. Eight key training attributes in automaticity acquisition were identified, and for each of them, a literature overview was performed. Based on this review, a two-step training model for automaticity acquisition was developed. The first step of the model, called Task Analytics was a task-specific procedure that determined the appropriate practice format based on the task's characteristics. The second step of the model consisted of a descriptive and a prescriptive model for the three phases of automaticity acquisition (fast learning, slow learning, and automatization). The descriptive model characterized each phase

of the automaticity acquisition based on psychological and performance-related characteristics. The prescriptive model recommended the appropriate training elements for each phase of automaticity acquisition.

After modeling the training principles of automaticity acquisition, the next step was to model the gamification process. Game design has always been a highly arbitrary process. Framing the process of game design into scientific models was the focus of Chapter 5. Two different gamification models were proposed, and their effectiveness were evaluated through randomized controlled trials. The first model was based on the motivation theories of self-determination and self-efficacy. Three versions of a game were designed; two based on either of the theories, and the third based on the integration of the theories. Forty-six volunteers participated in the study. The group that played the integrated design of the game played for a longer duration on average. However, the results of the data analysis did not show any significant difference between the three designs.

In another experimental study, based on the automaticity acquisition training model, a stage-based gamification model for automaticity acquisition was developed. This model classified the game mechanics into seven abstract categories and placed each mechanic at appropriate phases of automaticity acquisition. This gamification model was implemented in an off-the-shelf game, and its effectiveness was compared to the original design of the game by involving 49 participants. The results validated the effectiveness of the proposed gamification model for the acquisition of automaticity.

The research reported in the first five chapters built a solid platform for implementing the proposed dyslexia remediation approach. The multidisciplinary work required to develop this system is described in Chapter 6 and it included:

1. The detailed pedagogical design of the four reading tasks,
2. The game design highlighting the game scenarios and game elements,
3. The intelligent individualized module consisted of optimization model, genetic algorithm, linear regression, and artificial neural network models,
4. Cloud databases of user profiles, game parameters, and educational contents,
5. Threshold adaptive mechanisms that evolved the gameplay based on user performance,
6. Software development process for building the cross-platform system.

After describing this multifaceted system, multiple usability studies were carried out to evaluate the developed system's functionality and usability. These studies helped in the iterative user-centered design of the system. The analysis of the data validated the behavior of the intelligent model in selecting the appropriate content. Word frequency was a better predictor of word difficulty than letter-number. However, neither of them was sufficient for the accurate estimation of word difficulty. For finding accurate models of word difficulty, linear regression and artificial neural network models were used. Since word difficulty varies from task to task, four main lexical skills of auditory and visual word recognition, word decoding, and spelling were selected to represent most literacy tasks. Therefore, four different models of word difficulty was created. Regression models require sufficient input-output data to perform accurately. The input was acquired from lexical and sub-lexical databases with 55 quantified and normalized variables. The output was acquired from lexical mega-studies of auditory

lexical decisions (15842 words), visual lexical decision (25776 words), word naming (1482 words), and spelling (11373 words for five primary grades). The forward stepwise analysis was carried out in both linear regression and artificial neural networks to identify the top 10 variables in estimating word difficulty. The results 10-fold cross-validation showed that the artificial neural network models performed marginally better than the multiple linear regression models. The accuracy percentage (100 - mean absolute error) of the models were 90.58% for the auditory lexical decision, 92.08% for the visual lexical decision, 84.83% for the spelling task, and 86.98% for the word naming task (due to the limited size of the data, only linear regression was performed for word naming task).

The reported research was a multi-disciplinary work, and in each of the disciplines, there are possibilities for improvement. Having a therapist interface powered by visual analytics is one of the important research directions that could be followed. Another important future research direction is conducting an interventional study that can assess the effectiveness of the created system on the acquisition of reading skills for dyslexic children. Furthermore, in this thesis, a training model and a gamification model were developed for automaticity acquisition. The next step is to create the training and gamification models for learning declarative knowledge such as vocabulary.

For future research, it is also important to acquire the content for other languages. The implemented speech recognition and text-to-speech technologies offer support for tens of languages. Hence, the foundation of a multi-language platform is already built into the system but it requires including the educational content for each language. Additionally, cross-language studies on word difficulty may shed new light on the similarities and differences between orthographies. The implementation of the current games can be further improved by building more options to increase the user's autonomy. In addition, more reading tasks can be developed. The intelligent adaptive model, as well as the training and game design models, could be used in the future creation of reading tasks. Moreover, emotion recognition and eye-gaze technologies could be used to evaluate the user's level of engagement and adjust the system accordingly.

Currently, only a few of the lexical studies have targeted children. Future studies on children's word naming and lexical decision abilities could improve the intelligent model of this system. Furthermore, following the creation of models for the estimation of lexical difficulty, a natural future step is to extend the current study and create models to estimate sentence difficulty. Finally, the literature study of this field has shown that most of the established technology-based reading systems were the result of extensive and iterative research and development. The reading system developed in this research needs further improvement to reach the level of maturity and rigor required to solve the pervasive problem of dyslexia.



## Résumé

La dyslexie est le trouble de l'apprentissage le plus répandu. En l'absence d'interventions thérapeutiques, les enfants dyslexiques peuvent rencontrer des difficultés pour la réussite scolaire. Le manque d'automatisme dans la reconnaissance des mots est l'une des principales caractéristiques de la dyslexie. Cependant, malgré les recommandations fréquentes dans la littérature scientifique, l'automatisme est négligé dans les interventions de lecture. Cette thèse porte sur la remédiation à domicile de la dyslexie en se basant sur l'entraînement à l'automatisme.

Pour acquérir l'automatisme, il faut qu'il y ait une consistance entre le stimulus et la réponse. Cependant, les associations lettres-sons dans les orthographe opaques comme l'anglais et le français ne sont pas consistantes. Par conséquent, l'automatisme ne peut être acquis en pratiquant simplement les associations lettres-sons. Il faut donc cibler des unités plus grandes et plus consistantes, comme les associations graphème-phonème. La connaissance des associations graphème-phonème n'est que la première étape. L'objectif est de pratiquer ces associations jusqu'à l'acquisition de l'automatisme. Cependant, il existe des centaines d'associations graphème-phonème dans des orthographe opaques, et le temps scolaire limité ne permet pas la pratique intensive requise. C'est la raison pour laquelle nous proposons une approche basée sur une technologie utilisable au domicile de l'apprenant.

La pratique de l'automatisme est différente de l'apprentissage des connaissances déclaratives et elle nécessite un régime d'entraînement spécifique. Cependant, les modèles d'entraînement pour l'acquisition de l'automatisme sont rares et la plupart des modèles ont ciblé les aspects plus larges de l'acquisition des compétences. Pour combler cette lacune, une synthèse des principaux attributs de l'entraînement à l'automatisme, tels que la consistance, la contiguïté, la pression temporelle, le format de la pratique et l'interférence, a été réalisée. Sur la base de cette synthèse complète, un modèle d'entraînement à l'automatisme en deux étapes a été développé. La première étape, appelée Task-Analytics, aide les concepteurs de formation à prendre des décisions spécifiques à la tâche. La deuxième étape du modèle est divisée en deux parties : descriptive et prescriptive. La partie descriptive caractérise chaque phase d'acquisition de l'automatisme (apprentissage rapide, apprentissage lent et automatisation) sur la base d'attributs cognitifs liés à la performance. La partie prescriptive, qui constitue la principale contribution de ce modèle, recommande des éléments de formation efficaces pour chaque phase d'acquisition de l'automatisme en fonction de ses caractéristiques.

L'entraînement à l'automatisme nécessite de nombreuses répétitions qui peuvent devenir ennuyeuses pour l'apprenant. L'adhésion est donc la clé du succès de tout programme de formation de lecture. Les concepts de gamification et des jeux sérieux ont le potentiel d'augmenter la motivation, l'engagement et l'adhésion de l'apprenant. Cependant, les modèles actuels de gamification manquent de structure scientifique et de validation expérimentale. Pour résoudre ce problème, deux modèles de gamification ont été proposés et dont l'efficacité a été évaluée par des essais contrôlés randomisés. Le premier modèle proposé est basé sur l'intégration des théories de l'auto-efficacité et de l'auto-détermination. Trois versions d'un jeu ont été conçues : la première version était basée sur l'auto-efficacité, la deuxième sur l'auto-détermination et la troisième était basée sur l'intégration des deux théories. Quarante-six

volontaires ont participé à une étude expérimentale. Les résultats de l'étude expérimentale ont démontré que le groupe qui a été impliqué dans le jeu conçu sur la base de l'intégration des deux théories a fait preuve d'un engagement et d'une performance plus élevés. Cependant, ces différences n'étaient pas statistiquement significatives entre les groupes.

Le deuxième modèle de gamification était basé sur la partie prescriptive du modèle d'entraînement à l'automatisme. Les éléments du jeu ont été classés en sept catégories abstraites, et en outre, ils ont été placés dans des phases appropriées d'acquisition de l'automatisme en fonction de leurs caractéristiques. Sur la base de ce modèle de gamification, un jeu vidéo bien connu a été restructuré et son efficacité a été comparée à celle de la version originale du jeu à travers une étude expérimentale qui a impliqué 49 participants volontaires. Un défi multi-tâches a été intégré aux jeux afin d'évaluer le niveau d'automatisme. L'analyse des résultats a montré un avantage significatif en faveur de la conception basée sur le modèle de gamification de l'automatisme. L'avantage statistiquement significatif a été maintenu après l'introduction du défi multitâche, indiquant un niveau d'automatisme plus élevé.

Dans l'éducation traditionnelle, c'est le rôle de l'enseignant de proposer un contenu d'apprentissage qui ne doit être ni trop difficile, ni trop facile pour les enfants. Les technologies éducatives à domicile doivent être intelligentes pour compenser l'absence de l'enseignant. Dans cette thèse, un module intelligent composé d'un modèle d'optimisation et de modèles prédictifs basés sur les réseaux de neurones artificiels et la régression linéaire a été proposé. Le modèle d'optimisation a pour rôle de maximiser la valeur éducative de chaque session, tout en maintenant le niveau de difficulté du contenu dans les limites du niveau d'aptitude de chaque apprenant.

Pour mesurer le niveau de difficulté du contenu, les réseaux de neurones artificiels et les modèles de régression linéaire ont été utilisés pour estimer la difficulté lexicale. Cependant, comme la difficulté lexicale varie considérablement d'une tâche à l'autre, les modèles ont été créés pour quatre compétences lexicales : la reconnaissance auditive et visuelle, le décodage, et l'orthographe des mots. Les données pour ces tâches ont été collectées à partir de méga-études lexicales françaises testant des milliers de mots avec des sujets humains. Ces données ont été combinées avec 55 variables lexicales et sous-lexicales pour chaque mot. Cependant, puisque le nombre de variables était élevé, des analyses par étapes ont été effectuées à la fois pour la régression linéaire et le réseau neurones artificiels. Les 10 principales variables importantes pour chaque modèle ont été identifiées. La précision des modèles basée sur l'erreur absolue moyenne a atteint 90,58 % pour la reconnaissance auditive des mots, 92,08 % pour la reconnaissance visuelle des mots, 84,83 % pour l'orthographe et 86,98 % pour le décodage des mots.

L'analyse des modèles de difficulté lexicale pour l'orthographe a montré que des variables telles que le nombre d'associations irrégulières phonème-graphème, la fréquence des mots et consistance phonèmes-graphèmes la plus faible dans le mot étaient parmi les variables les plus importantes. La reconnaissance auditive des mots dépendait fortement du nombre de phonèmes, de la fréquence des mots et du nombre d'homophones. Pour la reconnaissance visuelle des mots, la fréquence des mots, le nombre d'homographes et le nombre de syllabes étaient parmi les variables les plus importantes. Enfin, pour le décodage des mots, le nombre de lettres, la fréquence des mots et la consistance de graphème-phonème initiale étaient parmi des critères importants.

---

Sur la base de cette recherche, quatre jeux ont été développés dans le moteur de jeu Unity 3D. Le processus de développement a suivi la conception multiplateforme, et les jeux ont été publiés sur Android, iOS, Microsoft Windows et macOS. Les bases de données éducatives ont été mises en œuvre dans le service de base de données en nuage AWS DynamoDB, et le modèle d'optimisation a été mis en œuvre dans le service informatique en nuage AWS Lambda. Les interactions principales étaient constituées des API de reconnaissance vocale et de synthèse vocale de Google Cloud, avec la prise en charge de dizaines de langues. Le développement a été effectué selon une approche itérative. Les jeux ont été évalués en réalisant plusieurs études d'utilisabilité qui ont confirmé la validité du système intelligent ainsi que l'utilisabilité des jeux dont le score a atteint 80 % sur l'échelle d'utilisabilité du système.



**Titre :** Gamification intelligente pour une remédiation à domicile et individualisée de la dyslexie

**Mots clés :** Dyslexie, Technologie de l'éducation, Système d'apprentissage intelligent, Jeu sérieux, Gamification, Thérapie à domicile

**Résumé :** La dyslexie est le trouble spécifique de langage le plus répandu, le manque d'automatisme dans la reconnaissance des mots est l'une de ses principales caractéristiques. Cette thèse est dédiée à la gamification de la remédiation de la dyslexie à domicile, elle focalise sur le développement de l'automatisme chez les enfants âgés de cinq à neuf ans. Une recherche approfondie sur les méthodes efficaces d'acquisition d'automatisme nous a permis de créer un nouveau modèle de formation pour l'apprentissage de l'automatisme. Cet apprentissage nécessite plusieurs répétitions mais l'adhésion au programme de formation est un critère très important pour l'obtention de bons résultats. En se basant sur le modèle d'apprentissage de l'automatisme, nous avons créé un modèle de gamification qui inclut chaque élément du jeu à l'étape appropriée d'acquisition d'automatisme. L'efficacité de ce modèle de gamification a été validée par un essai randomisé contrôlé.

Afin de fournir un système adaptatif, nous avons développé un modèle d'optimisation qui propose des sessions de formation individualisées en fonction du niveau de l'apprenant. Pour estimer le niveau de difficulté du contenu pédagogique, quatre compétences lexicales ont été modélisées à l'aide de réseaux de neurones artificiels et de régression linéaire. Pour chaque compétence lexicale, les 10 principales variables lexicales ont été identifiées par une analyse de type « Forward Stepwise ». La précision des modèles basés sur l'erreur absolue moyenne a atteint 90,58% pour la reconnaissance auditive des mots, 92,08% pour la reconnaissance visuelle des mots, 84,83% pour l'orthographe et 86,98% pour le décodage des mots. Enfin, sur la base de ces développements, quatre jeux ont été créés et fournis sur différentes plateformes informatiques. L'évaluation de l'utilisabilité a confirmé la viabilité du système intelligent et les jeux ont obtenu un score de 80% sur l'échelle d'utilisabilité du système.

**Title :** Intelligent gamification for individualized dyslexia remediation and in-home therapy

**Keywords :** Dyslexia, Educational Technology, Intelligent Learning System, Serious Games, Gamification, In-Home Therapy

**Abstract:** Dyslexia is the most prevalent learning disability and the lack of automaticity in word recognition is one of its main characteristics. This thesis is dedicated to the gamification of in-home dyslexia remediation by focusing on the development of automaticity for children aged from five to nine years old. An extensive research on the effective principles of automaticity training, led to the creation of a novel training model for automaticity acquisition. Training for automaticity requires countless repetitions, and adherence is crucial. Gamification concept was proposed to increase motivation, engagement, and adherence. Based on the automaticity-training model, a gamification model was developed for placing each game element at its appropriate phase of automaticity acquisition. The effectiveness of this gamification model was validated by a randomized controlled trial.

Furthermore, an optimization model was developed to provide individualized training sessions based on the level of the learner. For estimating the difficulty level of the content, four lexical skills were modelled using artificial neural networks and linear regression. For each lexical skill, the top 10 lexical variables were identified through forward stepwise analysis. The accuracy of the models based on mean absolute error reached 90.58% for auditory word recognition, 92.08% for visual word recognition, 84.83% for spelling, and 86.98% for word decoding. Finally, based on these developments, four games were created and provided on multiple platforms. The evaluation of the usability study confirmed the viability of the intelligent system and the games scored 80% on the system usability scale.