Design, Development, and Evaluation of Research Tools for Evidence-Based Learning: A Digital Game-Based Spelling Training for German Primary School Children

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Heiko Holz
aus Guayaquil, Ecuador

Tübingen
2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	12.11.2020
Stellvertretender Dekan:	Prof. Dr. József Fortágh
1. Berichterstatter:	Prof. Dr. Detmar Meurers
2. Berichterstatter:	Dr. Alexandra Kirsch
3. Berichterstatter:	Prof. Dr. Ulrike Pado

Acknowledgments

I would like to express my sincere gratitude to my supervisors Dr. Alexandra Kirsch and Prof. Detmar Meurers for their continuous support during my Ph.D. and for the opportunity of pursuing my research interests very independently. Your interest and supervision have led me to this point of my academic career, and your guidance kept me focused when I seemed to lose myself in ceaseless ideas and perfectionism.

I am also very grateful to the LEAD Graduate School & Research Network which funded my Ph.D. and enabled me to conduct my own research. Specifically, I would like to thank Mareike Bierlich, Sophie Freitag, and Ulrike Michael for providing such a supportive environment. I also would like to thank all current and former Ph.D. colleagues with whom I could share my struggles and successes and free my mind in visible fun activities. Particularly, I would like to thank Jakob Schwerter for his statistical advise, Tobias Appel and Luzia Leifheit for their adamat positivity, Franz Wortha and Silke Bieck for their like-mindedness, Molly Hammer for her writing support and joyful coffee breaks, Maria Chinkina for her dedication, and Manuel Ninaus for his academic mentorship. Let there be excellence!

A special thanks to the members of the Prosodiya project, particularly Katharina and Jochen Brandelik and Benedikt Beuttler. We have crafted something lovingly and invaluable of which we can be very proud. Thank you, Benedikt, especially for your coding support and for accompanying me during my thesis.

I also would like to thank the ICALL-Research.com Group of showing me the wonders of CALL. Particularly, I would like to thank my friend and colleague Zarah Weiss for daily lunch breaks and for always having an open door for me.

A big thanks also to team TuCAN, particularly Christian Mychajliw, who helped me to jointly give our "satellite" project wings to become real.

My gratitude also goes to Tess Renahan for carefully proofreading my thesis. It is not a matter of course to read and understand a thesis outside your own research.

Above all, I would like to thank my partner and tree star Stefan for his loving and unfailing support. It is impossible to express in writing how much you have supported me over the past years. You encouraged me to go on when I doubted myself and tamed the wild Heikofant when I was about to go mad.
A heartfelt thanks also to Sonja and Gerd for welcoming me into your family so warmheartedly and supporting me these past years.

Finally, I would like to thank my family, especially my sister and paragon Anja for her support, encouragement, patience, and down-to-earthness. Our path has not always been easy, but together we have – and will – overcome each tragedy that life has in store for us. Thank you for being such an outstanding big sister!
I wholeheartedly thank my parents for their love, support, and bravery. I am infinitely thankful for providing me with everything I could ever ask for, and for raising me an inquisitive, open-minded, and devoted person. Without this foundation, I would not be a scientist today, trying to make the world a better place.

Summary

Reading and writing are two of the most important skills acquired by young learners. However, approximately 4–10% of the German population suffers from dyslexia. If not treated adequately, reading and spelling disorders negatively affect children's academic, personal, and social development. Thus, the interest in digital tools to efficiently support literacy acquisition outside of class or therapy has grown.

This thesis addresses three main aspects to expand the current state of research on digital tools to support reading and spelling development. First, I present our novel digital game-based spelling training called "Prosodiya" to improve literacy skills in German primary school children. Prosodiya differs from similar approaches in that it systematically teaches orthographic knowledge in combination with the awareness of syllable stress. I investigate the feasibility, effectiveness, and validity of Prosodiya in a randomized controlled field trial with 116 German second to fourth graders with mainly poor spelling skills. The training was carried out at home over a short period of 9–10 weeks. Results showed significant improvements in syllable stress awareness and spelling abilities for trained and untrained word material. Prosodiya was also reportedly easy to use, motivating, and provided good game experience, proving its feasibility for use at home. Further, the validity of our novel pedagogical approach was confirmed in correlation analyses investigating the relationships among syllable stress awareness, reading, spelling, and training performances. Prosodiya may therefore expand the traditional pool of training methods.

Second, I address one challenge of interaction design in educational applications, namely choosing the appropriate interaction style for use with children – which is not trivial. For this, I compare drag-and-drop, point-and-touch, and simple touch for writing words in a mobile spelling game. I evaluate the perceived workload, user experience, and writing time of 25 German third and fourth graders (8–11 years). We found that touch was the fastest, best rated, and most preferred interaction style. We also found small advantages for drag-and-drop over point-and-touch, which runs counter to some research and recent design guidelines.

Third, I explore the potential and limitations of automatic generation of language learning material in two domains. I investigate the utility of text-to-speech tools to automatically generate minimal pairs (e.g., *beaver* vs. *peaver*) for use in language learning systems to foster learners' phonological awareness. I present our novel approach to improve the pronunciation of artificially generated German pseudowords. The results of an online study showed that distinguishing the lexical word from its pseudoword counterpart was equally successful when the minimal pair was generated by our method or produced by a human. In addition, I further present COAST, a web-based tool for easy and automatic visual enhancement of syllable structure, word stress, and spacing in reading material using resources of natural language processing. COAST's feasibility and usability were validated in user tests.

All aspects contribute to the current state of research on digital learning tools and pose promising results to support children's reading and spelling development.

Zusammenfassung

Lesen und Schreiben gehören zu den wichtigsten von Kindern erworbenen Fähigkeiten. Indes leiden 4–10% der Deutschen Kinder an einer Lese-Rechtschreib-schwäche (LRS). Wenn die LRS nicht angemessen behandelt wird, beeinträchtigt sie die akademische, persönliche und soziale Entwicklung betroffener Kinder erheblich. Daher steigt das Interesse an digitalen Lernwerkzeugen, um die Kinder auch außerhalb oder ergänzend zu Therapie und Schule wirksam zu unterstützen.

Die vorliegende Dissertation konzentriert sich auf drei Hauptaspekte im Bereich digitaler Lernwerkzeuge zur Förderung der Lese- und Rechtschreibentwicklung. Zuerst stelle ich "Prosodiya" vor – unser neuartiges, digitales und spielerisches Rechtschreibtraining für deutsche Grundschulkinder. Prosodiya unterscheidet sich von ähnlichen Ansätzen, indem es systematisch orthographisches Wissen in Kombination mit der Bewusstheit für Silbenbetonung vermittelt. Ich untersuche die Anwendbarkeit, Wirksamkeit und Validität von Prosodiya in einer randomisierten kontrollierten Feldstudie mit 116 deutschen Grundschulkindern mit vorwiegend schwachen Rechtschreibleistungen. Die Kinder übten zu Hause über einen Zeitraum von neun bis zehn Wochen. Die Ergebnisse zeigen signifikante Trainingseffekte im Bereich der Rechtschreibung und der Bewusstheit für Silbenbetonung. Kinder berichteten ebenfalls, dass Prosodiya einfach zu bedienen und motivierend sei, was die Verwendbarkeit des Spiels für zu Hause bestätigt. Weiter konnten wir in Korrelationsanalysen, die den Zusammenhang zwischen der Bewusstheit für Silbenbetonung, Lesen, Rechtschreiben, und Trainingsleistungen betrachteten, unseren neuartigen pädagogischen Ansatz bestätigen. Prosodiya könnte daher den traditionellen Pool an evidenzbasierten Trainingsmaßnahmen erweitern.

Zweitens untersuche ich eine der großen Herausforderung des Interaktionsdesigns von Lernapps: die Wahl des für Kinder am besten geeigneten Interaktionsstils. Ich vergleiche systematisch die Interaktionsstile Drag-and-Drop, Point-and-Touch und Touch in einem mobilen Rechtschreibspiel mit 25 deutschen Grundschülern. Die Ergebnisse zeigen, dass Touch gegenüber Drag-and-Drop und Point-and-Touch am besten abschnitt. Entgegen einiger Forschungen und aktuellen Designrichtlinien fanden wir kleine Vorteile von Drag-and-Drop gegenüber Point-and-Touch.

Drittens exploriere ich Möglichkeiten der automatischen Erzeugung von Sprachlern-Material in zwei Bereichen. Ich untersuche den Nutzen von Text-to-Speech Tools, um automatisch Minimalpaare (z.B. *Biber* vs. *Piber*) für den Gebrauch in Sprachlernsystemen zu erzeugen. Ich stelle unsere neue Methode zur Verbesserung der Aussprache künstlich erzeugter Pseudowörter vor. Ergebnisse einer Online-Studie zeigen, dass die Unterscheidung des lexikalischen Wortes von seinem Pseudowort-Gegenstück gleich gut erfolgt, wenn die Minimalpaare mit unserer Methode erzeugt oder von Menschen gesprochen wurden. Zudem stelle ich COAST vor, eine Webapp zur automatischen und anpassbaren Hervorhebung von Silbenstruktur und Wortbetonung in Lesematerialien. In Nutzertests wurde COAST erfolgreich validiert.

Die Beiträge erweitern den aktuellen Forschungsstand digitaler Lernwerkzeuge und bieten aussichtsvolle Implikationen für die Lese-Rechtschreibförderung von Kindern.

List of Publications

Parts of this thesis have been published elsewhere. A listing of explicit contributorship for each manuscript is given in Chapter 10.

1. **Holz, H.**, Beuttler, B., & Kirsch, A. (2017). Bewegungserkennung mit Wearables für Embodied Trainings in Serious Games [Movement Recognition with Wearables for Embodied Trainings in Serious Games]. In M. Burghardt, R. Wimmer, C. Wolff, & C. Womser-Hacker (Eds.), *Mensch und Computer 2017 – Tagungsband* (pp. 259–262). doi: 10.18420/muc2017-mci-0368
2. **Holz, H.**, Brandelik, K., Brandelik, J., Beuttler, B., Kirsch, A., Heller, J., & Meurers, D. (2017). Prosodiya – A Mobile Game for German Dyslexic Children. In J. Dias, P. A. Santos, & R. C. Veltkamp (Eds.), *Games and Learning Alliance. GALA 2017* (pp. 73–82). doi: 10.1007/978-3-319-71940-5_7
3. **Holz, H.**, Brandelik, K., Beuttler, B., Brandelik, J., & Ninaus, M. (2018). How to train your syllable stress awareness – A digital game-based approach for German dyslexic children. *International Journal of Serious Games*, 5(3), 37–59. doi: 10.17083/ijsg.v5i3.242
4. **Holz, H.**, Beuttler, B., & Ninaus, M. (2018). Design Rationales of a Mobile Game-Based Intervention for German Dyslexic Children. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 205–219). doi: 10.1145/3270316.3272053
5. **Holz, H.**, Ninaus, M., Meurers, D., & Kirsch, A. (2018). Validity and Player Experience of a Mobile Game for German Dyslexic Children. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 469–478). doi: 10.1145/3270316.3271523
6. **Holz, H.**, Chinkina, M., & Vetter, L. (2018). Optimizing the Quality of Synthetically Generated Pseudowords for the Task of Minimal-Pair-Distinction. In *2018 IEE Spoken Language Technology Workshop (SLT)* (pp. 470–476). doi: 10.1109/SLT.2018.8639037
7. **Holz, H.**, Weiss, Z., Brehm, O., & Meurers, D. (2018). COAST – Customizeable Online Syllable Enhancement in Texts. A flexible framework for automatically enhancing reading materials. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 89–100). doi: 10.18653/v1/W18-0509
8. **Holz, H.**, & Meurers, D. (submitted). Interaction Styles in Context: Comparing Drag-and-Drop, Point-and-Touch, and Touch in a Mobile Spelling Game.
9. **Holz, H.**, Ninaus, M., Beuttler, B., Brandelik, K., & Meurers, D. (unpublished). A Digital Game-Based Training Improves Spelling in German Primary School Children – A Randomized Controlled Field Trial.

Contents

Part I

Introduction

Chapter 1

Introduction

"The illiterate of the 21st century will not be those who cannot read and write, but those who cannot learn, unlearn, and relearn."
— Alvin Toffler, *Future Shock* (1970)

"Das stumme h, das ist nicht schwer, steht meist vor l, m, n und r." —
gelbes Kugellicht, *Prosodiya*

Reading and writing belong to the most important skills acquired by young learners. Importantly, this still applies even in the modern era of digital devices that offer spelling correction, read-aloud functionality, or voice user interfaces using speech recognition. Unfortunately, approximately 4–10% of the German population do not master the challenges of learning to read and write appropriately and are diagnosed with developmental dyslexia (Katusic, Colligan, Barbaresi, Schaid, & Jacobsen, 2001; Moll, Kunze, Neuhoff, Bruder, & Schulte-Körne, 2014; Moll & Landerl, 2009). Dyslexia is thus one of the most frequent learning disorders that affects more than 50.000 German children of each birth cohort. The learning disorder is characterized by a specific, isolated impairment of reading and/or spelling which cannot be explained by inadequate schooling, delayed development of cognitive abilities, or low intelligence of a child (Schulte-Körne, 2010). If reading and spelling disorders are not diagnosed and treated adequately, they negatively affect children's academic (Daniel et al., 2006), personal (Schulte-Körne, 2010), and social development (Beddington et al., 2008) in the short and long run.

Compared to their classmates, dyslexic children acquire reading and writing skills at a much slower pace and not as proficiently (Schulte-Körne & Remschmidt, 2003) and suffer massively from their impaired literacy acquisition. If dyslexic children do not receive appropriate treatments, they lose their motivation for the learning process and the faith of being able to develop a comprehension of literacy language (Bender et al., 2017), and are more likely to experience negative thoughts, depression, and school-related anxiety (Schulte-Körne, 2010). Further, not being able to properly read and write heavily impairs children's academic careers, future employment prospects, and personal well-being in the long run (Daniel et al., 2006). The learning disorder negatively impacts not only mental health but also social and cultural participation (Beddington et al., 2008; Daniel et al., 2006). Thus, appropriate treatments and interventions are indispensable to support reading and spelling development of affected children in order to counteract negative consequences in time, and to improve their future prospects.

The effectiveness of traditional teaching methods to improve literacy skills applied in standard classroom or individual learning therapy is widely proven and much is known on effective treatment components of spelling disorders. In addition to traditional learning therapy, digital learning tools are invaluable assets that offer a range of possibilities to support children with reading and spelling disorders. For example, digital trainings and learning environments can be used to support children in or outside of classroom and learning therapy and have shown great promise to support children's literacy acquisition (cf. Holz, Brandelik, Beuttler, Brandelik, & Ninaus, 2018). Moreover, digital game-based trainings exploit the use of various game elements, such as narratives, feedback, or rewards, to boost motivation and to address negative feelings in the learning process (Deterding, Dixon, Khaled, & Nacke, 2011) and support successful learning outcomes (cf. Boyle et al., 2016; Hainey, Connolly, Boyle, Wilson, & Razak, 2016). However, more research is needed in the empirical evaluation of digital game-based spelling trainings to support German (dyslexic) primary school children in the home environment (cf. Holz, Brandelik, et al., 2018; Holz, Ninaus, Beuttler, Brandelik, & Meurers, unpublished).

1.1 Aims and Contributions of this Thesis

The aim of this thesis is to extend the current state of research on the challenges and benefits of digital learning tools to support reading and spelling development of primary school children. That is, this thesis presents the design, development, and evaluation of digital educational tools to facilitate positive learning experiences in reading and spelling. Particularly, I address the following three main aspects:
(1) the development and evaluation of a digital game-based spelling training to improve children's literacy skills in the home environment, (2) the systematic comparison of touch interaction styles for use with children, and (3) the exploration of automatic generation and input enhancement of learning materials.

The main contribution of this thesis is the development and evaluation of the digital game-based spelling training called "Prosodiya" to fill the gap of empirically evaluated and evidence-based spelling trainings for the home environment. The spelling training is designed for German primary school children to improve reading and spelling skills in the home environment. Prosodiya is the first digital approach that systematically teaches spelling skills by focusing on syllable stress awareness and on linking the linguistic features related to syllable stress to orthographic regularities of German orthography. The development and evaluation of Prosodiya follows three major principles: *feasibility, validity,* and *educational effectiveness*. In this context, *feasibility* refers to the game's everyday applicability in the home environment, i.e., that it can be used by children unassisted and that it engages and motivates over a longer time. The feasibility of the training is determined by the game design (i.e., does Prosodiya offer a good game experience?) and implementation of the educational content (i.e., do children perceive a positive influence on their literacy skills?). The *validity* is determined by the integrity of the design and implementation of the education content, which is based on scientific evidence. Lastly, *educational effectiveness* means that the spelling training evidentially supports improving children's literacy skills. The effectiveness is determined by the trainings' validity and feasibility. To evaluate the feasibility, validity, and effectiveness of Prosodiya, we carried out a randomized controlled field trial with 116 German primary school children.

For the second contribution of this thesis, I zoom in on one of the major challenges of interaction design in educational applications. That is, I address the question whether to implement a drag-and-drop, point-and-click, or click interaction style to move and interact with objects. In particular, I address a central game mechanic of Prosodiya, namely the implementation how to interact and move letters in our touch-based spelling game. As I will argue in this thesis, the design decision is non-trivial and the state of research seems puzzling and contradictory. To determine the most appropriate interaction style for the touch-based spelling game, we systematically compare the drag-and-drop, point-and-touch, and touch interaction style in a within-subject experiment with 25 German primary school children.

The third aspect of this thesis addresses the automatic generation and input enhancement of learning material in two specific domains of language learning. The provision of age- and skill-appropriate content is costly and thus limits the available materials. In particular, I explore the utility of state-of-the-art technology to automatically generate speech material and for the automatic visual input enhancement of reading material. On one hand, I explore the use of text-to-speech tools to automatically generate speech material for Prosodiya. In particular, I present and evaluate our novel approach to improve the pronunciation of synthetically generated pseudowords for the use in minimal pair distinction tasks (e.g., distinguishing *beaver* and *peaver*). Further, I address the issue of the scarcity of age- and skill-appropriate visually enhanced reading material. For this, I present and evaluate COAST – a web-based system to automatically enhance syllable structure, word stress, and spacing in texts.

To summarize, this thesis is highly interdisciplinary and covers aspects of different disciplines and topics ranging from computer science, psychology, and educational science to computational linguistics. Thus, the results of this thesis may have practical implications for all these disciplines.

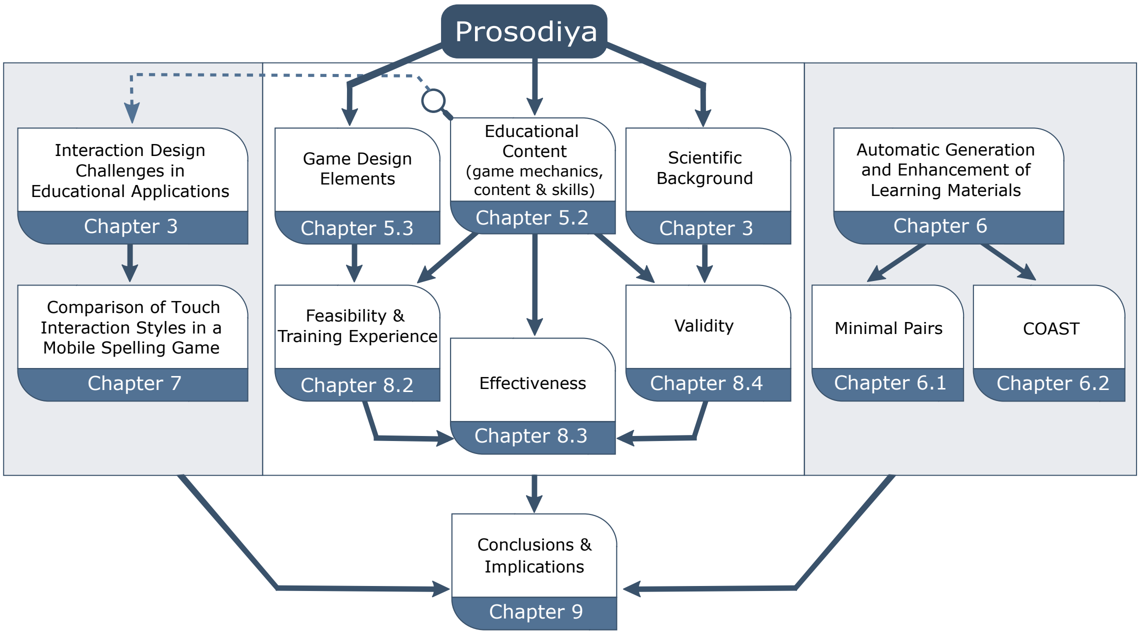


Figure 1.1: Overview of this thesis with regard to the three main contributions: (1) design, development, and evaluation of the digital game-based spelling training "Prosodiya", (2) interaction design challenges in education application with focus on interaction styles for use with children, and (3) automatic generation and enhancement of learning material in two domains of language learning.

1.2 Reader's Guide

In the following, I elaborate on the structure of this thesis. Figure 1.1 shows an overview of the three main aspects of this thesis, clarifies their relations, and displays how the different parts of this thesis contribute to achieving the research aims of this interdisciplinary work.

This thesis is divided into four parts to elaborate on the challenges and benefits of digital learning tools in reading and spelling development.

Part I introduces the reader to the relevant fields of reading and spelling disorders as well as to the interaction design challenges in educational applications. I give an overview on the state of research of the causes and treatment of reading and spelling disorders in Chapter 2. In particular, I focus on the association between syllable stress awareness and reading and spelling skills. Further, I provide an overview on the benefits and disadvantages of therapeutic, computer-based, and digital game-based training programs. Lastly, to address the second aspect of this thesis, I discuss the current state of research on (touch) interaction styles for the use with children in Chapter 3, particularly with focus on mobile spelling games.

Part II describes our approach to support reading and spelling development of German primary school children with the use of state-of-the-art technology. In particular, I describe the rationale, design, and development of our digital game-based spelling training Prosodiya in Chapter 5. Further, I present our approach to automatically generate speech material and our approach of automatic visual input enhancement of reading material in Chapter 6, addressing the third aspect of this thesis.

Part III includes the results of two major studies. In particular, I present the results of our study comparing different touch interaction styles in a touch-based spelling game for children in Chapter 7. To address the major contribution of this thesis, I present the results of our randomized controlled field trial evaluating the feasibility, validity, and effectiveness of Prosodiya in Chapter 8.

Part IV concludes this thesis with summarizing the findings of the different studies involved in achieving the three aims of this thesis. In Chapter 9, I elaborate on the benefits of digital learning tools in reading and spelling development and derive implications for practical applications based on the empirical findings of each of the three main aspects of this thesis.

1.3 Notational Conventions

1.3.1 Pedagogical Boxes

In Section 5.2, every paragraph will start with blue pedagogical boxes to briefly summarize the educational content taught throughout Prosodiya.

Scope of the next paragraph

- Learn the notation rules used in this thesis to express different linguistic properties of written and spoken language.
- Learn how the state of research or scientific contributions of this thesis are summarized.

1.3.2 Linguistic Notations

In this thesis, we also need to distinguish between object language and linguistic property language. Table 1.1 lists the notation rules used in this thesis to express different linguistic properties of written and spoken language.

Table 1.1: Notation rules to express different linguistic properties of written and spoken language.

Notation rule	**Example**
German words are written in italics and their translation is provided in square brackets	*rennen* [to run]
The written separation of words into syllables (hyphenation) is marked by hyphens (-)	*ren-nen* [to run]
Stressed syllables are written in capital letters	The word *rennen* [to run] is stressed on the first syllable, while the second syllable is unstressed: *REN-nen*
Graphemes and phonemes representing specific linguistic features are underlined	The word *rennen* [to run] is spelled with an ambisyllabic consonant doubling *nn* that marks the short vowel of the stressed syllable: *RE<u>N-n</u>en*
The phonetic representation of sounds uses the notation of the International Phonetic Association (IPA)[1]	The phonetic transcription of *rennen* [to run] is /ˈʁɛnən/

[1] `https://internationalphoneticassociation.org`

1.3.3 Summary Boxes

Finally, every state of research or scientific contribution of this thesis ends with a red summary box.

Summary

- Blue pedagogical boxes are used to express educational content that will be practiced in Prosodiya.
- The thesis uses explicit notation rules to express different linguistic properties of written and spoken language.
- Red summary boxes are used to summarize the state of research or scientific contributions of this thesis.

Chapter 2

Causes and Treatments of Dyslexia

In this chapter, I reflect on the state of research on the causes and treatments of reading and spelling disorders that our digital game-based spelling training is founded on. First, I explain the relationship between literacy skills and phonological awareness in Section 2.1, with a specific focus on syllable stress awareness and German orthography. Then, I give an overview of the treatment of reading and spelling disorders in Section 2.2, highlighting the advantages and disadvantages of therapeutic, computer-based, and digital game-based interventions. I conclude this chapter with a summary.

2.1 Dyslexia, Phonological Awareness, and Syllable Stress

According to current research, dyslexia is not caused by a single factor, but rather is influenced by myriad factors, including genetic disposition, socioeconomic factors, cognitive functions, and the perception and processing of visual and acoustic information (Schulte-Körne & Remschmidt, 2003). In this regard, the phonological deficit theory is the most well-developed and evidence-based theory that sees a causal role of phonological skills in children's development of reading and spelling (cf. Ramus, 2003; Snowling, 2001) – children with good phonological skills become good readers and good spellers, while children with poor phonological skills progress more poorly (cf. Goswami, 1999). As such, a deficient phonological awareness is known as one major cause of dyslexia (Bradley & Bryant, 1983; Snowling, 1995).

Phonological awareness refers to the ability to deal with the sound system of a language and to detect, distinguish, and manipulate segments of a language (Klicpera, Schabmann, & Gasteiger-Klicpera, 2013). In the broader sense, phonological awareness refers to the ability to analyze and manipulate larger sound units, such as words, syllables, and rhymes, while the ability to analyze and manipulate single sounds (phonemes[1]) is referred to as phonological awareness in the narrower sense, or phonemic awareness.

2.1.1 Speech Rhythm and Syllable Stress

Phonological awareness also includes the perception and processing of prosodic features. A shortcoming in the perception of prosodic features is a strong predictor for dyslexia (Goswami et al., 2013; Leong, Hämäläinen, Soltész, & Goswami, 2011; Sauter, Heller, & Landerl, 2012). One of these features is syllable stress, an important characteristic of German speech rhythm. In spoken language, similarly to music in which the tone is the recurring element, the speech rhythm is generated by syllables. German, along with other languages including English, Russian, Germanic languages, and European Portuguese, belongs to the stress-timed languages category (Kohler, 1986). In stress-timed languages, speech rhythm is generated by the regular appearance of stressed syllables, and the intervals between stressed syllables tend to have a constant duration of approximately 500 milliseconds (Arvaniti, 2009; Pompino-Marschall, 2009).

Stressed syllables differ from unstressed syllables in intensity (loudness), duration (length), and pitch. Stressed syllables are on average louder, longer (Jessen, Marasek, Schneider, & Claßen, 1995), and oftentimes higher in pitch than unstressed syllables. Further, the rise time (the time required to reach peak signal intensity) of stressed syllables is shorter (Thomson & Jarmulowicz, 2016), i.e., the increase in amplitude of the speech envelope is steeper – the vowel sound of the stressed syllable gets loud faster (Pompino-Marschall, 2009). In contrast, unstressed syllables are compressed and reduced to fit the rhythm. In Figure 2.1, the speech signal of the word *BUT-ter* [butter] is displayed, with the peak signal intensity on the phoneme /u/ indicating that the word is stressed on the first syllable. The rhythm of stress-timed languages differs to, among others, syllable-typed languages, such as Italian, French, and Spanish, in which the duration of every syllable is approximately constant in time.

Recent empirical findings have shown that the perception of stress is impaired in dyslexic children (Goswami et al., 2013; Jiménez-Fernández, Gutiérrez-Palma, & Defior, 2015; Leong et al., 2011) and that syllable stress awareness highly correlates with reading and spelling skills (Sauter et al., 2012). Particularly, Sauter et al. (2012) investigated the relationship between stress awareness and reading and spelling skills in German third and fourth graders. In their study, they asked the children to select the one sentence out of three whose stress pattern matched the rhythm played on a piano. They found that the ability to correctly identify the matching stress pattern highly correlated with children's reading and spelling skills. They also investigated

[1] A phoneme refers to the smallest identifiable unit of sound. When speaking, phonemes are combined to form words. For example, the phonemic representation of the word *malen* [to paint] is /ˈmaːlən/.

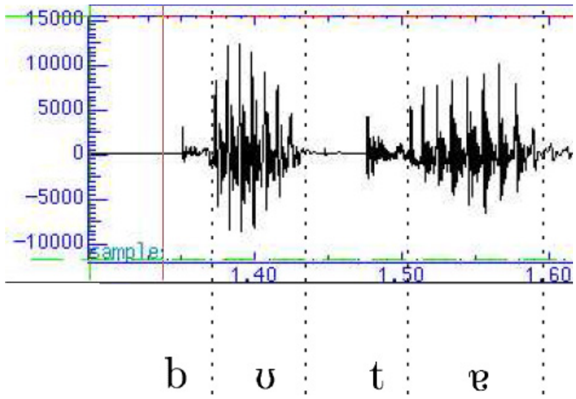


Figure 2.1: Oscillogram of the word *Butter* [butter] (Mayer, 2014, p. 95). The peak signal intensity on the phoneme /u/ indicates that the first syllable is stressed.

children's abilities to segment spoken sentences into syllables. For this, they asked the children to speak the sentences rhythmically and to simultaneously tap out the resulting rhythm. They evaluated whether the number of taps corresponded to the number of syllables of the target sentence. They found a strong association between children's syllable segmentation and spelling, but not with reading. In line with the latter finding, other studies have shown that rhythmic (singing) games only enhanced or correlated with spelling and not with reading (Overy, 2003; Overy, Nicolson, Fawcett, & Clarke, 2003). Sauter et al. (2012) argues that possibly no attention was paid to the stress pattern in rhythmic tasks that focused on segmental phonological skills, such as tapping out the rhythm of a song or clapping along to a sentence, and not on the suprasegmental skill of stress identification.[2] This argument is supported by studies that have shown positive training effects on reading and spelling skills of rhythmic interventions that included stress identification (Bhide, Power, & Goswami, 2013; Thomson, Leong, & Goswami, 2013).

2.1.2 Syllable Stress and Vowel Length Marking

For German dyslexics, one explanation for the relationship between poor spelling and impaired stress identification is thought to be found in the association between stress and German orthographic markers, also known as *"Dehnungs- und Doppelungszeichen"* [lengthening and doubling marks]. Orthographic markers, i.e., graphemes marking long and short vowels, generally occur in stressed syllables (markers for long vowels, such as the bigram *ie* in *BIE-ne* [bee]) or in conjunction with stressed syllables (markers for short vowels, such as the ambisyllabic consonant doubling *tt* in *Ge-WIT-ter* [thunderstorm]) (Staffeldt, 2010; Vennemann, 2011).

Mastering the complex orthographic rules to mark long and short vowels is a major

[2] Speech segmentation refers to the process of segmenting spoken language into smaller units, such as words, syllables, and phonemes. Suprasegmental characteristics are vocal effects that extend over more than one speech segment, such as pitch, stress, tone, or word juncture (Crystal, 2011).

difficulty for German children (Klicpera & Gasteiger Klicpera, 2000; Landerl, 2003). Short vowels are consistently marked by following two rules (cf. Ise & Schulte-Körne, 2010): (i) "If the short vowel phoneme is followed by only one consonant in the same morpheme,[3] then this consonant has to be doubled in the spelling (e.g., *rennen* [to run], and *Ball* [ball])", and (ii) "if the short vowel phoneme is followed by two or more consonant phonemes in the same morpheme, then these consonants are not doubled (e.g., *Felsen* [rock] and *Wald* [forest])".

In contrast, the marking of long vowels is more complex and less consistent (cf. Ise & Schulte-Körne, 2010). Long vowel phonemes can be marked (i) by doubling the vowel grapheme (e.g., *Haar* [hair]), (ii) by a diphthong[4] (e.g., *Daumen* [thumb]), by marking the long vowel *i* with the bigram *ie* (e.g., *Biene* [bee]), (iii) by adding a "silent h" (e.g., *fehlen* [to miss]), or (iv) simply by the absence of a consonant doubling (e.g., the grapheme *o* is a long vowel phoneme in *holen* [to fetch sth.] but a short vowel phoneme in *wollen* [to want sth.]). However, the rules of long vowel marking are more complex and have many exceptions. For example, marking of the long vowel phoneme *i* follows the rule that "if *i* is a long vowel phoneme, then it is spelled with the bigram *ie* (e.g., *Biene* [bee])", with the exception of words that are not of German origin (e.g., *Kino* [cinema]), words in which the long vowel *i* is not preceded by a consonant (e.g., *Igel* [hedgehog]), words that are untypical for German as they have more than two syllables (e.g., *Maschine* [machine] or *Mandarine* [tangerine]), pronouns (e.g., *mir*, *dir*, *wir* [mine, yours, we] and *ihr*, *ihm*, *ihn* [her, him, his]), and others (cf. Röber, 2012).

The same phenomenon of vowel length marking can also be explained on the syllable level. Short vowels are marked orthographically "if the phonological word features an ambisyllabic consonant, a so-called syllabic joint. Then, the grapheme, which phonographically corresponds to the ambisyllabic consonant, is doubled"(Eisenberg, 2013, p. 266). According to syllable rules, an ambisyllabic consonant can function as the final sound of the first stressed syllable or as the initial sound of the following unstressed syllable (Eisenberg, 1998). For example, the consonant *n* in the words *REN-nen* [to run], *KEN-nen* [to know sb. or sth.], or *NEN-nen* [to name sb. or sth.] is ambisyllabic. According to a syllable rule stating that stressed syllables with short vowels are always closed,[5] it functions as the final sound of the first stressed syllable. According to a syllable rule stating that simple consonants between two vowels always belong to the syllable of the second vowel, it functions as the initial

[3] A morpheme is the smallest meaningful unit in written language. For example, the root of a word is a morpheme and *renn* is the root of *rennen* [to run].

[4] Diphthongs are double sounds formed by the combination of two different vowels in a single syllable. Typical German diphthongs are *ei*/*au* (e.g., *weinen* /ˈvaɪ̯ən/ [to cry] and *Kaiser* /ˈkaɪ̯zə [emperor], *eu*/*äu* (e.g., *freuen* /ˈfʀɔɪ̯ən/ [to be pleased] and *Bäume* /ˈbɔʏmə/ [trees]), and *au* (e.g., *Daumen* /ˈdaʊ̯mən/ [thumb]).

[5] Syllables that end with a single or cluster of consonant phonemes (the coda) are called closed syllables, i.e., the syllable is closed by the consonant phoneme(s). In contrast, open syllables are coda-less and end with a vowel phoneme.

sound of the unstressed vowel (Eisenberg, 1998).

As such, vowel length markers express phonological characteristics that are generally connected to syllable stress (Eisenberg, 1998). They express a long and loud syllable rhyme that is typically filled by a stressed long vowel (e.g., the long vowel /ˈeː/ in *NEH-men* [to take]) or by a stressed short vowel which is connected with an ambisyllabic coda (e.g., the short vowel /ˈɛ/ + ambisyllabic coda /n/ → /ˈɛn/ in *REN-nen* [to run]). Thus, the phonological origin of orthographic markers is connected to syllable stress. However, this phonological origin can be superimposed by morphological processes. For example, the ambisyllabic consonant structure can vanish in inflected words (e.g., *RENNT* [he/she/it runs], or *ge-RANNT* [I/we/they/he/she/it ran]), or word formation processes can shift the primary stress to another, unmarked syllable (e.g., *AB-fall* [trash]). However, each of these orthographically marked words can be traced back to the basic form of the trochee – the German disyllabic standard word in which the first syllable is stressed and the second syllable is unstressed (e.g., *FAL-len* [to fall], *REN-nen* [to run], *FEL-sen* [rock], *SE-geln* [to sail]). The phonological origin of orthographic markers lies in this basic form that consists of a stressed and an unstressed syllable.

To conclude, besides morphological skills, lexical knowledge and knowledge of spelling rules (Galuschka & Schulte-Körne, 2016; Ise & Schulte-Körne, 2010; Schulte-Körne & Mathwig, 2013), syllable stress awareness may play a role in the orthographic stage of spelling acquisition, particularly in the spelling of long and short vowels (Sauter et al., 2012). Processing verbal stress adequately may thus help children to acquire the complex spelling rules that underlie vowel length spelling in German orthography.

Although complementing reading and spelling trainings with explicit focus on syllable stress seems reasonably to improve literacy skills, we are not aware of any empirically evaluated digital training that systematically teaches orthographic knowledge in combination with the awareness of syllable stress. What has been shown is that interventions including rhythmic exercises focusing on syllable segmentation evidently improve the spelling ability in German poor spellers (e.g., Reuter-Liehr, 1993; Tacke, Wörner, Schultheiss, & Brezing, 1993). Additionally, rhythmic trainings that contain exercises to match the correct syllabic stress pattern to words have been shown to be beneficial for the development of literacy and phonological awareness of English poor readers (Bhide et al., 2013; Thomson et al., 2013). Further, training material focusing on typical German trochees is widely used and accepted (e.g., Mildenberger Verlag, 2018; Röber, 2009). Yet, syllable stress awareness has not been included comprehensively in digital spelling trainings for German. This thesis aims to fill this gap by presenting and evaluating a novel digital game-based spelling training that focuses on improving the awareness of syllable stress and linking the stressed syllable's linguistic features to orthographic principles of German orthography.

Summary of syllable stress awareness in dyslexia research

- Syllable stress awareness, i.e., perceiving and identifying stress patterns of words or sentences, is highly related with reading and spelling skills.
- In the German orthographic system, the phonological origin of vowel length marker, i.e., the orthographic marking of long and short vowels, is seen in syllable stress of the typical German trochee – the German disyllabic standard word in which the first syllable is stressed and the second is unstressed.
- Processing verbal stress adequately might help children to acquire complex spelling rules that underlie vowel length spelling in German
- There is a gap in research of empirically evaluated approaches that focus on improving the awareness of syllable stress or that associate the linguistic characteristics of the stressed syllable to orthographic regularities of German orthography.

2.1.3 On the Matter of Different Orthographies

Interventions for dyslexic children differ in content depending on the target language. If a certain type of exercise has been shown to improve spelling in one language, it does not necessarily transfer to other languages.

Languages differ in the depth of their orthography with regard to reading, i.e., the degree to which the pronunciation of a word can be predicted from its spelling, and spelling, i.e., the degree to which the spelling of a word can be predicted from its pronunciation. With respect to reading, German is considered a shallow (or transparent) orthography with quite consistent grapheme-phoneme (or letter-to-sound) correspondences (Seymour et al., 2003), i.e., the relationship between each grapheme and phoneme is simple and predictable. In contrast, languages with a deep orthography, such as English, have many inconsistent and complex grapheme-phoneme correspondences. For example, the grapheme *a* corresponding to the vowel of the stressed syllable is always pronounced the same in the German words *Ball*, *Name*, *Arm*, and *Parade*, while it is pronounced differently in each of the English words ball, name, arm, and parade.

Further, German orthography, just like in English, closely adheres to the principle of morpheme consistency (Landerl & Reitsma, 2005), i.e. "the spelling of morphemes is preserved in different word forms (e.g., *fahren* [to drive], *Fahrer* [diver], *Gefährt* [vehicle])" (Landerl & Thaler, 2013, p. 136). The orthographic spelling rules are only applicable to the word stem, which is consequently spelled with high consistency. Thus, once the spelling of a certain word stem is stored, it can be applied to all word forms (Landerl & Reitsma, 2005). Moreover, with regard to word stress, German words usually adhere to stem stress (Bußmann, 2008, p. 22), i.e., the stress falls on the first syllable of the stem of the word.

The principle of morpheme consistency is probably the main reason why the translation of phonemes into graphemes is much less consistent in German (Landerl & Thaler, 2013). In spelling, one has to choose among various possible translations of a phoneme into a grapheme, which also explains the considerably large number of homophonic spellings (e.g., *malen-mahlen* [to paint-to mill], *Aale-Ahle* [eals-awl], *fiel-viel* [fell-many]). The orthographic vowel length marking, as explained in Section 2.1.2, is especially inconsistent. Thus, German is considered a rather deep orthography with respect to spelling (Landerl & Thaler, 2013).

The differences in orthographic consistencies between languages result in differences in reading and spelling difficulties of dyslexic children and consequently in different therapeutic approaches (Aro, 2004; Frith, Wimmer, & Landerl, 1998; Landerl, Wimmer, & Frith, 1997; Rau, Moll, Snowling, & Landerl, 2015). Hence, the remainder of this thesis focuses on the remediation of reading and spelling disorders in German dyslexic children.

2.2 Treatment of Reading & Spelling Disorders

In the following, I provide a brief overview on the remediation of reading and spelling disorders in German dyslexic children and reflect on the types of interventions that have been empirically proven to improve literacy acquisition. This section is primarily based on our article "How to train your syllable stress awareness – A digital game-based intervention for German dyslexic children" (Holz, Brandelik, et al., 2018).

According to the clinical guideline of the diagnosis and treatment of German dyslexic children (cf. Galuschka & Schulte-Körne, 2016), the treatment of reading and spelling disorders can be divided into different areas, corresponding to different stages in the progress of reading or spelling development. Recommendations are inferred for evidence-based treatments for each of these areas based on the results of meta-analyses of randomized controlled field trials (cf. Galuschka, Ise, Krick, & Schulte-Körne, 2014; Ise, Engel, & Schulte-Körne, 2012; McArthur et al., 2012). An overview is given in Table 2.1.

The areas of the treatment of reading disorders include: (i) syllabic and phonemic awareness, i.e., awareness of syllable and sounds, (ii) reading accuracy, (iii) reading fluency, and (iv) reading and text comprehension. For the first stage, they recommend trainings to identify, categorize, segment, delete, or discriminate syllables and sounds in words. The second stage includes systematic instructions of grapheme-phoneme (letter-to-sound) correspondences and exercises of phoneme synthesis. In this regard, derivational synthesis refers to blending (pulling together) individual parts of a language within words, e.g., blending individual sounds or syllables to words. Reading fluency should be trained with systematic exercises of phoneme, syllable, and morpheme synthesis. Lastly, reading comprehension includes interventions that feature tasks in which participants learn to extract textual information,

Table 2.1: Overview of evidence-based treatment of reading and spelling disorders (Galuschka et al., 2014, p. 284).

Intervention							
Treatment of reading disorder				**Treatment of spelling disorder**			
Awareness of syllables and sounds	**Reading accuracy**	**Reading fluency**	**Reading/text comprehension**	**Awareness of syllables and sounds**	**Phoneme-grapheme allocation**	**Memory retrievals**	**Knowledge of rules and morphemes**
Exercises aiming to identify, categorize, segment, delete, or discriminate syllables and sounds in words	Systematic instruction of grapheme-phoneme correspondence and exercises for phoneme synthesis	Systematic exercises for phoneme, syllable, and morpheme synthesis	**If reading accuracy or speed are low:** Instructions of grapheme-phoneme correspondences or systematic exercises for phoneme, syllable, and morpheme synthesis **As a result of speech/ language disorder:** Interventions to increase vocabulary repertoire and competencies in terms of syntax and grammar	Exercises aiming to identify, categorize, segment, delete, or discriminate syllables and sounds in words	Systematic instruction of phoneme-grapheme correspondences and exercises for phoneme analysis at the lexical and sublexical levels	Systematic exercises for storing/ remembering graphemes	Acquiring orthographic and morphemic regularities

summarize it, and relate it to existing knowledge (Galuschka et al., 2014).

The areas of the treatment of spelling disorders are: (i) syllabic and phonemic awareness, (ii) phoneme-grapheme correspondence, (iii) grapheme memory entries, and (iv) knowledge of orthographic rules and morphemes. The first stage of spelling acquisition parallels the first stage of reading promotion and thus maintains the same interventional recommendations. For the second stage, phonics instruction (systematic instructions of phoneme-grapheme [sound-to-letter] correspondences) and exercises for phoneme analysis at the lexical and sublexical levels[6] are recommended. For the latter, analysis refers to segmenting words into respective parts, e.g., phonemes or syllables. For the third stage, systematic exercises for storing and remembering frequent sequences of graphemes are recommended. Finally, exercises to acquire orthographic and morphemic regularities are recommended for the fourth stage of the treatment of spelling disorders.

The conclusions drawn from Galuschka and Schulte-Körne (2016) are summarized on the next page.

[6] The lexical level refers to the level of (whole) words, while the units of sublexical levels are sounds, syllables, and morphemes.

Summary of effective treatment approaches for reading and spelling disorders

Reading

- Reading skills can most effectively be improved with systematic instruction of grapheme-phoneme correspondences and phoneme, syllable, and morpheme synthesis.
  - In this regard, derivational synthesis refers to blending (pulling together) individual parts of a language within words, e.g., blending individual sounds or syllables to words.

Spelling

- Spelling performance can most effectively be improved by using systematic instructions of phoneme-grapheme correspondences, exercises analyzing sounds, syllables, and morphemes, as well as trainings enabling the acquisition and generalization of orthographic regularities.

2.3 From Therapeutic to Digital Game-Based Interventions

Evidence-based treatments of reading and spelling disorders can be applied in different ways, i.e., in learning facilities and/or at home, with a human tutor and/or a digital device. In the following, I briefly summarize the advantages and disadvantages of therapeutic, computer-based, and digital game-based interventions in order to motivate the game-based pedagogical approach of our spelling training.

2.3.1 Therapeutic Interventions

Commonly, reading and spelling disorders are treated in therapeutic interventions administered by trained practitioners, such as teachers or learning therapists, in weekly individual or group sessions over several months. Therapeutic interventions are recommended treatments for dyslexic children (Galuschka & Schulte-Körne, 2016) and can reliably improve reading (e.g., Groth, Hasko, Bruder, Kunze, & Schulte-Körne, 2013; Klicpera, Weiss, & Gasteiger-Klicpera, 2013; Tacke, 2005) and spelling (e.g., Ise & Schulte-Körne, 2010; Reuter-Liehr, 1993). Although significant improvements in reading or spelling can be observed after several weeks or months, children should continue to receive support until their ability to read and spell reaches a level that enables them to participate in public life in an age-appropriate way (Galuschka & Schulte-Körne, 2016). This usually results in several years of intense support and treatment. However, this may often not be provided if the healthcare system has no provision for funding (Galuschka & Schulte-Körne, 2016) – as is the case in Germany. In Germany, learning therapy is not covered by health insurance but by the youth welfare office. The application processes for financing or reimbursement of learning therapy can be complicated and tedious, creating a disadvantage for families who cannot afford to pay for learning therapy privately. As a result, affected children may not receive appropriate treatment timely, sustainably, or long enough. Importantly, therapeutic interventions should be implemented by experts in reading and spelling development and its promotion (Galuschka & Schulte-Körne, 2016) rather than by peers, parents, or university students – which may appeal to affected families as a cost-effective alternative but whose effectiveness could not be confirmed unequivocally (Galuschka et al., 2014; Ise et al., 2012).

To conclude, therapeutic interventions are reliable and recommended treatments when administered by experts, but they are cost-intensive, time- and location-dependent, and may not be available timely or long enough.

2.3.2 Computer-Based Interventions

In addressing the disadvantages of therapeutic interventions and offering new ways to engage young learners, it has been shown in recent years that computer-based interventions and the use of information and communication technology (ICT) successfully complement traditional teaching and learning therapy in improving reading and spelling in German dyslexic children (e.g., Kargl, Purgstaller, Weiss, & Fink, 2008; Kast, Baschera, Gross, Jäncke, & Meyer, 2011; Klatte, Bergström, Steinbrink, Konerding, & Lachmann, 2018) and, more generally, facilitates literacy acquisition in dyslexic children (e.g., Cidrim & Madeiro, 2017; Drigas & Batziaka, 2016). Moreover, computer-based interventions are independent of time and place. They can be designed in such ways that children can use them autonomously for homework, complementary to school and learning therapy without a real tutor, e.g., by using children-friendly instructions and feedback. In fact, while computer instructions may be equally effective as human tutors (e.g., in handwriting and spelling cf. Berninger, Nagy, Tanimoto, Thompson, & Abbott, 2015), children have shown to concentrate better while engaged with computer-based interventions than in traditional school tasks (Ronimus, Kujala, Tolvanen, & Lyytinen, 2014). In addition, interactive experiences in computer-based interventions can motivate young learners and help to attenuate their daily struggles in literacy acquisition (Cidrim & Madeiro, 2017). Furthermore, computer-based interventions are able to automatically adapt the learning content to the specific needs of individual children. This is necessary for dyslexics who have heterogeneous difficulties in different levels of literacy acquisition (Rose, 2009). They also may offer continuous and more frequent assessment of proficiencies and knowledge (e.g., Ninaus, Kiili, Mcmullen, & Moeller, 2017; Sense, Behrens, Meijer, & van Rijn, 2015) to get an overview of (individual) learning outcomes with less effort or to gain further insights in learning processes.

Finally, gamification, i.e., "the use of game design elements in non-game contexts" (Deterding et al., 2011, p. 2), plays a major role in computer-based interventions. For example, computer-based interventions have been enriched with, among others, points, badges, or reward systems. The role of gamification is primarily to invoke the same psychological experiences as games generally do (Huotari & Hamari, 2012). In educational and learning contexts, gamification mostly positively affects learning and increases motivation, engagement in, and enjoyment of learning tasks (Hamari, Koivisto, & Sarsa, 2014).

2.3.3 Digital Game-Based Interventions

One step further, digital game-based interventions are the top tier of digital interventions for children with learning disorders and special needs. Digital game-based learning is also often referred to as serious games, educational games, or edutainment. Such digital game-based approaches extend gamified computer-based interventions by addressing the lack of fully utilizing the engaging and motivational potential of digital games.

In this regard, it is crucial to differentiate between gamified computer-based interventions and digital game-based approaches. While gamified computer-based interventions merely incorporate elements of games to existing tasks that may be unengaging, tedious, or boring (Plass, Homer, & Kinzer, 2015), game-based interventions are designed as full-fledged games for educational purposes (Deterding et al., 2011) that focus on designing activities as playful tasks (Plass et al., 2015). As defined by Wouters and van Oostendorp (2013), digital games have to be interactive, based on a set of agreed rules and constraints, directed toward a clear goal that is often set by a challenge, and constantly provide feedback either as a score or changes in the game world to enable self-monitoring of progress towards the goal.

At this point, it is important to note that gamified computer-based interventions are often advertised as actual educational *games*, which can result in negative consequences due to the expectations posed to educational games: Parents and children see enjoyment as one of the central principles in educational software that is used in the home environment (Kerawalla & Crook, 2005) and sometimes even prioritize enjoyment over educational benefits. Thus, digital game-based interventions are expected to be engaging and fun and thus naturally motivating (Kerawalla & Crook, 2005; Ronimus & Lyytinen, 2015). Further, the motivational design of interventions used in the home environment is particularly important, whereas learning effectiveness is referred to the most essential aspect at school (Ronimus & Lyytinen, 2015). Hence, the mere use of game design elements in computer-based interventions and then selling them as games may result in failure to live up to these expectations.

The benefits of game playing as a learning process are widely acknowledged (Gee, 2003; Prensky, 2003) and research on digital game-based learning has become more popular in recent years (for an overview see Boyle et al., 2016 and Hainey et al., 2016). Digital game-based learning has been shown to be effective or even outperform conventional instruction methods in terms of learning and retention, such as

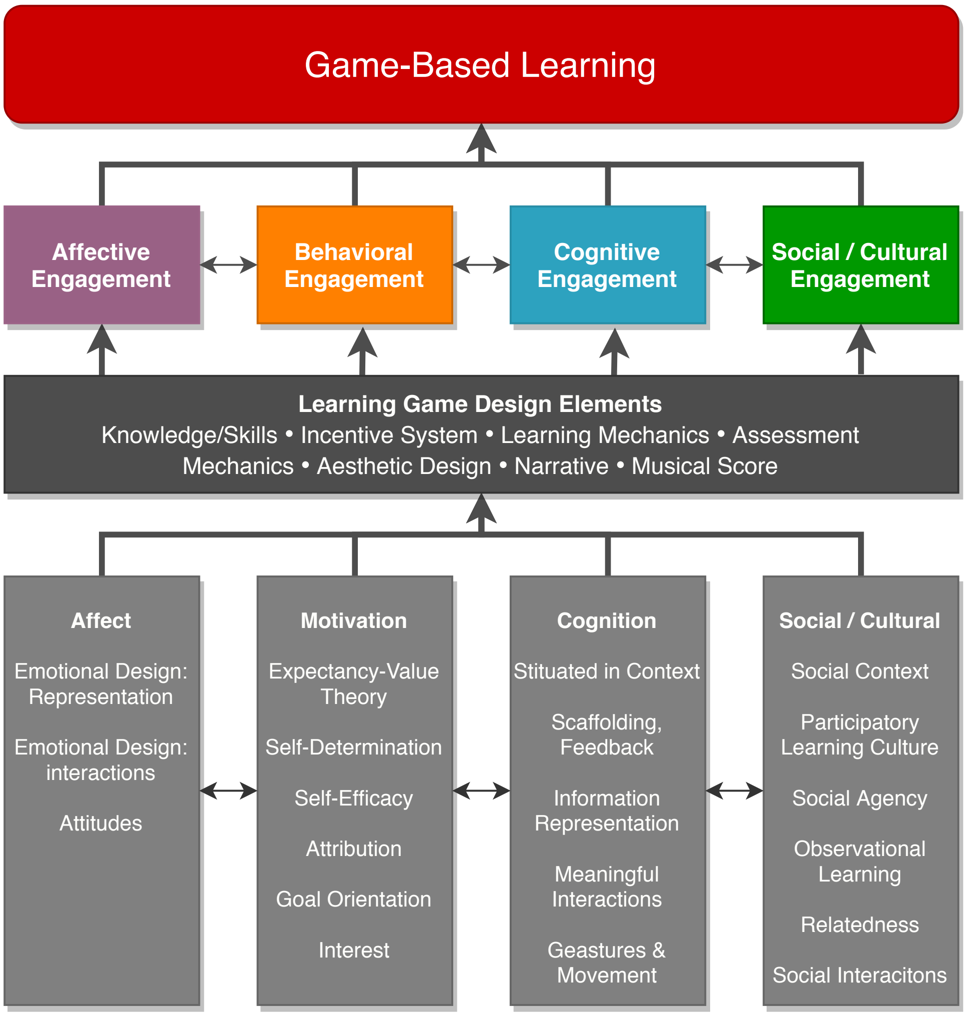


Figure 2.2: The integrated design framework of game-based and playful learning (Plass et al., 2015, p. 263).

lectures, reading, drill and practice, or hypertext learning environments (Wouters & van Oostendorp, 2013). In fact, this is particularly the case for language learning as highlighted by the meta analysis of Wouters and van Oostendorp (2013). Specifically for learning disorders, educational games have proven to support children with dyslexia or dyscalculia (e.g., Abrami, Borohkovski, & Lysenko, 2015; Ninaus, Kiili, McMullen, & Moeller, 2016), and, most importantly for this thesis, the acquisition of reading and spelling in German dyslexics (e.g., Berkling, 2017; Görgen, Huemer, Schulte-Körne, & Moll, 2020; Lenhard & Lenhard, 2016).

Game elements embedded in digital game-based interventions, such as feedback, reward, or narratives, influence learning positively (Wouters & van Oostendorp, 2013) and play a crucial role in achieving learning goals (Boyle et al., 2016). They address negative feelings, such as frustration, demotivation, or boredom (Deterding et al., 2011), promote engagement and learning for children with special needs (Ke

& Abras, 2013), boost children's engagement with literacy activities, foster skill-reinforcement, and enhance the perception of reading progress (Holmes, 2011). Game-based learning may even reengage learners who disengaged from learning, i.e., learners who lost interest, motivation, and engagement in learning and who cannot be engaged with other methods (Griffiths, 2002; Squire, 2008).

In addition, digital game-based interventions are also especially suited to foster learning through embodied cognition, i.e., mapping of gestures or movement to key features of the content to be learned (Plass et al., 2015), as discussed in more detail in our articles Holz, Beuttler, and Kirsch (2017) and Holz, Brandelik, et al. (2018). For example, a Kinect-based literacy game, using gestures and movements in in-game activities, had more impact on children's literacy outcomes compared to an intervention without embodied activities (Homer et al., 2014).

However, to ensure that defined learning outcomes are still in focus, special attention must be payed to balancing educational effectiveness and quality of learning with game play – a corollary to the design process of digital game-based interventions (Arnab et al., 2015; Ke & Abras, 2013; Plass et al., 2015; Quinn, 2005). If the educational content prevails the game experience, learner's motivation may decrease. On the other hand, if too much emphasis is placed on game experience and fun, it may undermine learning (Gros, 2017). If the right balance between seriousness and gaming is found, digital game-based interventions "are able to engage learners on an affective, behavioral, cognitive, and sociocultural level in ways few other learning environments are able to do" (Plass et al., 2015, p. 270) – as emphasized by the integrated design framework of game-based and playful learning of Plass et al. (2015), see Figure 2.2.

2.3.4 Digitization is not a Blanket Solution

It is important to state that digital approaches should be used as a supplementary tool in or outside of class in assisting dyslexic children – but may not replace traditional teaching and learning strategies or therapeutic interventions (Kyle, Kujala, Richardson, Lyytinen, & Goswami, 2013; Reid, Strnadová, & Cumming, 2013). Commonly, single digital programs do not cover all processes of reading and/or spelling development.

At this point, it is also important to emphasize that digital learning and teaching approaches are not the panacea to cure learning disorders. For example, if exercises or games are not adapted to the actual skill and language levels of the children, they may fail to understand or solve the task independently. This may lead the children to skim an app without engaging with it, to adopt trial-and-error learning such as random guessing, or to misuse the app to engage in more fun activities than the actual learning goal (Falloon, 2013). However, this may ultimately result in frustration or demotivation of the children or, more generally, to senseless waste of time that teachers need to compensate for.

2.3.5 State of Research on the Effectiveness of Treatment Approaches for Spelling Disorders

Research on the spelling remediation in German dyslexics was mainly done with weekly therapeutic interventions administered by experts (e.g., Ise & Schulte-Körne, 2010; Reuter-Liehr, 1993; Schulte-Körne & Mathwig, 2013; Tacke, 2005) or instructed parents (e.g., Schulte-Körne, Deimel, & Remschmidt, 1998; Schulte-Körne, Schäfer, Deimel, & Remschmidt, 1997), or with digital (game-based) interventions in daily to weekly supervised training sessions during school lessons (Kargl et al., 2008; Klatte et al., 2018) or after school (e.g., Berkling, 2017), sometimes with additional training at home (Kargl et al., 2008).

However, randomized controlled field trials (RCFT) on the effectiveness of digital (game-based) treatment approaches – when the training is carried out under "real-world conditions" in the home environment without adult help – are missing in clinical practical guidelines (Galuschka & Schulte-Körne, 2016) and meta reviews (Galuschka et al., 2014; Ise et al., 2012; McArthur et al., 2012). While Görgen et al. (2020) were recently able to show in a RCFT that a digital game-based reading training carried out independently in the home environment can significantly improve reading abilities for trained word material in German children with reading disorders, I am not aware of such RCFTs on spelling trainings for German primary school children.

To support literacy acquisition in children in the home environment, those interventions must specifically address two main aspects. First, the pedagogical approach implemented in the intervention must be linguistically sound and based on empirical findings to ensure educational efficacy – the effect on learning under optimal conditions. Second, the intervention has to ensure its ecologically valid impact on literacy acquisition under real-world conditions. This means that children must be able to work with and complete the training without adult help, for example with the use of age-appropriate instructions and feedback, and that they are motivated and engaged over a longer periods of time to maximize learning outcomes.

The summary of the different treatment forms for dyslexia is given on the next page.

Summary of treatment forms for dyslexia

- Therapeutic, computer-based, and digital game-based interventions are effective approaches in supporting dyslexic children to improve their reading and spelling abilities.
- Therapeutic interventions are cost-intensive and may not be implemented timely or long enough due to tedious and complicated application processes for financing or reimbursement of learning therapy.
- Digital interventions are independent of time, location, and real tutors and can automatically adapt the learning content to the performance levels of individual children.
- Digital game-based approaches may be as effective as therapeutic interventions while increasing, among others, motivation, engagement in, and enjoyment of learning tasks with the use of game design elements.
- There is a research gap in the effectiveness of computer-based and digital game-based interventions for German dyslexic children in the home environment.

2.4 Summary

To conclude the background on syllable stress awareness in dyslexia research and effective treatment approaches for dyslexia, I briefly summarize the key facts that are most important to follow the remainder of this thesis.

Summary of syllable stress and treatment forms in dyslexia research

Syllable stress in dyslexia research

- A deficient phonological awareness – the ability to deal with the sound system of a language and to detect, distinguish, and manipulate segments of a language, such as words, syllables, rhymes, or single sounds – is one major cause of dyslexia.
- The perception of (syllable) stress and other prosodic features, which belong to phonological awareness in the broader sense, is highly correlated with reading and spelling abilities and a deficiency in stress identification is a strong predictor of dyslexia.
- Syllable stress is closely related to the vowel length marking of German orthography, i.e., the orthographic marking of long and short vowels. The phonological origin of vowel length marking is seen in syllable stress of the typical German trochee – the German disyllabic standard word in which the first syllable is stressed and second is unstressed.
- There is a lack of treatment approaches (and the empirical evaluation thereof) that focus on syllable stress awareness and on linking the linguistic characteristics of stressed syllables to orthographic regularities of German orthography.

Treatment forms for dyslexia

- Reading and spelling in dyslexic children can successfully be improved with therapeutic, computer-based, and digital game-based interventions.
- Game-based approaches may be as effective as therapeutic interventions while increasing, among others, motivation, engagement in, and enjoyment of learning tasks.
- There is a research gap in the effectiveness of computer-based and digital game-based interventions for German dyslexic children carried out in the home environment.

Chapter 3

Interaction Design Challenges in Educational Applications

In this chapter, I elaborate on one of the major challenges of interaction design in educational applications, namely the choice of the most appropriate interaction style. For this, I first reflect on the state of research of interaction styles with children in Section 3.1. Then, I provide an overview of mobile spelling exercises for touch devices in Section 3.2. I conclude this chapter with a summary.

As previously mentioned, game mechanics belong to the building blocks of game-based learning. Thus, when developing educational applications for children, one of the central design decisions is how actions are performed to complete a task. For example, in a simple educational spelling game in which children write single words by selecting letter objects, the question arises concerning which interaction style should be implemented to perform the action of moving and arranging the letters. The term "interaction style" is commonly used to refer to the different ways users can interact with an application or a game (Soegaard, 2015). For the given example, the question is whether to implement a drag-and-drop, a point-and-click, or a click interaction style to arrange the letters in the spelling line.

The design decision impacts performance (Donker & Reitsma, 2007a; Girard & Johnson, 2009; Hamza & Salivia, 2015; Inkpen, 2001; Joiner, Messer, Light, & Littleton, 1998; Ward, 2014), user experience (e.g., Barendregt & Bekker, 2011; Girard & Johnson, 2009; Inkpen, 2001), and educational effectiveness (Schwartz & Plass, 2014). For young learners, it is especially challenging to implement appropriate interaction styles as it not only depends on their physical and mental development (e.g., Hamza & Salivia, 2015; Vatavu, Cramariuc, & Schipor, 2015), but also on their previous experience with technology (Barendregt, 2015).

In the present thesis, I zoom into the interaction design of one central game mechanic of Prosodiya, namely the implementation of a touch-based spelling game. The aim is to investigate the appropriateness of different interaction styles to move and interact with letters to write words. I am doing so in our article "Interaction Styles in Context: Comparing Drag-and-drop, Point-and-touch, and Touch in a Mobile Spelling Game" (Holz & Meurers, submitted). This part of the thesis is greatly based on that article.

When consulting current research and guidelines on interaction styles, the conflicting results and recommendations leave the designer puzzled (Hourcade, 2008). Often point-and-click is said to be more appropriate than drag-and-drop (Chiasson & Gutwin, 2005; Gelderblom & Kotzé, 2009; Girard & Johnson, 2009; Inkpen, 2001; Joiner et al., 1998; Roman, 2015; Soni, Aloba, Morga, Wisniewski, & Anthony, 2019; Ward, 2014) – other times the opposite (Barendregt, 2015; Barendregt & Bekker, 2011; Donker & Reitsma, 2007a; Hamza & Salivia, 2015). As we argue in the next section, the conclusions seem to depend on the input modality (mouse vs. touch), the type of task, the age of the children, and the typical interaction experience of children with technology at the time the study was conducted.

As highlighted in Section 2.2, interventions that enable the acquisition and generalization of orthographic regularities can most effectively improve spelling in dyslexic children. This is also true for computer-based spelling trainings for German dyslexic children (e.g., Baschera, 2011; Berkling, 2017; Kargl et al., 2008; Kast et al., 2011; Klatte et al., 2018) – and spelling games for mobile touch devices are reportedly more engaging than paper-pencil exercises are (e.g., Rello, Bayarri, & Gòrriz, 2013; Rello, Bayarri, Otal, & Pielot, 2014). Hence, the interaction design decisions made when developing spelling exercises could greatly influence the learning gain in dyslexic children. However, research is lacking on the appropriateness of different interaction styles in (mobile) spelling games. In the present thesis, I fill this research gap by investigating the three most commonly used interaction styles in a mobile touch-based spelling game. For this, we systematically compare the drag-and-drop, point-and-touch, and touch interaction styles with regard to subjectively perceived workload, user experience, and processing times in a lab experiment with twenty-five German children aged 8–11.

In the following, I discuss the current state of research on interaction styles for children and mobile spelling games. The methods and results of our experiment are reported in Chapter 7.

3.1 Interaction Styles and Children

In this section, I summarize the empirical findings on interaction styles for young children and the published design guidelines based on those findings.

Joiner et al. (1998) compared point-and-click and drag-and-drop using a computer mouse with children aged 7–12. In their first study, 7-year-old children were found to be faster and more accurate with pointing compared to dragging. The second study compared the performance of children from three age groups (5–6, 8–9, and 11–12 years). Older children were found to be faster and made fewer errors than younger children, independent of the interaction style. The youngest children were slower and made more errors with dragging than with pointing, while no such performance differences arose for the other age groups. Therefore, they concluded that point-and-click is the most appropriate interaction style for young children.

Inkpen (2001) found in a study with 68 children aged 9–13 that point-and-click was faster, led to fewer errors, and was preferred to drag-and-drop. The examined task was a puzzle-solving game using a computer mouse. With point-and-click, objects were moved by clicking on them, then another click on the point to where they should move, and a final click to make the move. With drag-and-drop, the mouse had to be kept down on the object while dragging it to the desired position. One of the reasons why children preferred point-and-click was that their fingers became tired of holding the mouse button down. Again, point-and-click was concluded to be more appropriate. In line with these findings, in their design principles for children's technology, Chiasson and Gutwin (2005) recommend that drag-and-drop should be replaced with point-and-click.

Donker and Reitsma (2007b) found the opposite results when comparing drag-and-drop and point-and-click in a "moving objects" task with 107 Dutch children aged 6–7 years. Children were asked to move one letter that was falsely written in a word into a trash bin. Drag-and-drop was found to be faster and to cause fewer interaction errors than point-and-click. They conclude that drag-and-drop is the most appropriate for educational software. They discuss whether error-handling differences may have influenced previous results favoring point-and-click. E.g., when children made an error in drag-and-drop in the study by Inkpen, Booth, and Klawe (1996), they had to perform the entire trial again, whereas for point-and-click they only had to redo the incorrect click. Children thus may perform drag-and-drop more accurately and slowly to avoid time-consuming rectifications (Donker & Reitsma, 2007b).

Girard and Johnson (2009) compared drag-and-drop with point-and-click in multiplication tasks with children aged 7–9 using a computer mouse. Children computed the result by first selecting the column of the digit (units, tens, hundreds) and then selecting numbers by point-and-click or using sliders with drag-and-drop. They found that point-and-click was more effective in terms of achievement, interaction error, speed and accuracy of answer, and was preferred over drag-and-drop. Gelderblom and Kotzé (2009) concur in their lessons learned in the design of technology for children aged 6–8.

Barendregt and Bekker (2011) and Barendregt (2015) performed two studies to investigate children's spontaneous use of drag-and-drop and point-and-click interaction styles in two educational math games on a computer using a computer mouse and internal touch pads. In the first game, focusing on division, children had to draw lines to equally divide candy among four babies. In the second game, children had to position numbers on a number line. In the main experiment of the first study (Barendregt & Bekker, 2011), in which both games only supported point-and-click, they tested 12 Dutch children aged 7–12 years and found that many children tried to apply drag-and-drop as a first spontaneous reaction and struggled with the point-and-click interaction style. In the second study (Barendregt, 2015), the games supported both interaction styles. Twenty-six Dutch children aged 4–6 years were asked to play without previously being informed about interaction styles. When the children played a game without watching a demo showing an interaction style, children spontaneously used drag-and-drop instead of point-and-click for both games, independent of how often they had used the different interaction styles before. The spontaneous use of drag-and-drop was especially clear for actions with a natural mapping to keep the mouse button pressed, such as cutting or drawing lines, as in the cutting game. Point-and-click was more easily adopted in the moving game. The authors assume that the children developed habits from their experiences with drag-and-drop-like functionality in desktop and mobile phone interfaces. The results suggest that drag-and-drop may be appropriate for use in educational applications, even for very young children, and that it is important to not only consider performance measures but also take habits and spontaneous use of interaction styles into account.

Ward (2014) reported on the HCI requirements of a Computer Assisted Language Learning tool for 4–5 year old Irish primary school students. The tool used keyboard and mouse input and included an exercise in which jumbled letters had to be rearranged to spell the word correctly. Two interaction styles were considered: selecting a letter and dragging it to the spelling line, or placing any letter on the spelling line that is clicked on. Both options were tested. Drag-and-drop was reported as problematic, requiring more dexterity for the movement, so the click option was adopted in the final system.

Roman (2015) conducted a usability study with 4–6 year-olds for an online streaming service. Children either used a mouse or a laptop touchpad. She concludes that clicking is favored over drag-and-drop. The same year, Hamza and Salivia (2015) investigated the performance of, among others, point-and-touch and drag-and-drop of 4–5-year-olds with an iPad application. They report that the children tried to perform drag-and-drop in the point-and-touch tasks as it apparently was more familiar, easier, and less confusing to them. They also observed that the children had better motor control in the tasks using drag-and-drop. They showed an effect of age, with 5-year-olds outperforming the younger ones. In other research (Azah, Syuhada, Batmaz, Stone, & Wai, 2014; Soliman & Nathan-Roberts, 2018), toddlers and infants were shown to be capable of using various gestures with mobile touch devices, with the drag-and-drop gesture being acquired at around the age of three. Interestingly, the touchscreen interaction design recommendations for children by Soni et al. (2019) still recommend avoiding drag-and-drop.

In sum, the varied findings summarized here support no simple conclusion as to the most appropriate interaction style for educational applications for children. Instead, we argue that it is necessary to consider the interaction modality (touch vs. computer mouse), the particular implementation of the interaction styles, the context and type of task, and the ages and interaction experiences of the participating children. While all of the discussed interaction styles are used in published spelling games, we are not aware of any research that systematically compares different interaction styles in spelling games for children in which single or multiple words are written.

3.2 Spelling Exercises for Mobile Touch Devices

Spelling exercises available for mobile touch devices differ, among other features, how letters are arranged to write words and in the interaction styles used to select or move the letters (cf. Holz & Meurers, submitted).

As foundation for a concise overview, I downloaded exemplary apps for Android devices from the Google Play Store that contained spelling games for German primary school children. Four exemplary apps are shown in Figure 3.1. In the following, I briefly describe the different types of spelling games.

Writing words using a complete keyboard: Words are written using the the device's default or a custom keyboard, see Figure 3.1a. Touching a letter inserts it into or appends it to the end of the word, depending on the cursor's position.

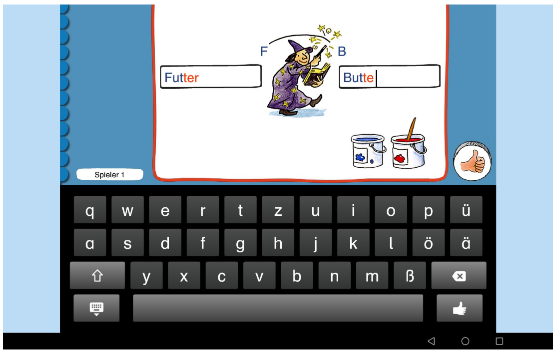
Writing words using a restricted keyboard: Words are written using a partially displayed keyboard. The letter colors in Figure 3.1b refer to different keyboard parts. Other parts become visible by swiping left/right, or by pressing arrows keys.

Writing words using a predefined set of letters: Words are written using a set of letters, see Figure 3.1c. Optionally, distracting letters are included to increase task difficulty. Depending on the app, a letter can be used once or multiple times. In the former case, the number of available letters equals the word length (plus optionally distracting letters).

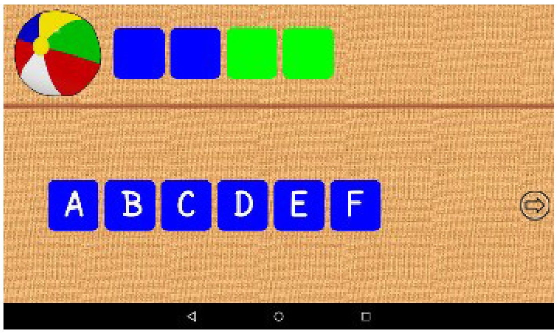
Correcting or writing words with limited interactions: An unfinished or misspelled word is presented and needs to be completed or corrected by inserting, deleting, or replacing letters, see Figure 3.1d. The focus is on major spelling challenges derived by typical dyslexic errors.

Touch-based spelling games (apart from those with a complete keyboard) use three different interaction styles to select and arrange letters in order to write a word: drag-and-drop, point-and-touch, and touch. As I only refer to touch devices in this thesis, I renamed *click* to *touch*.

Prosodiya includes a spelling game in which words are written using a predefined set of letters (cf. Section 5.2.5 on page 53). As highlighted in our literature review, the decision regarding which interaction style should be implemented in the spelling game cannot be made conclusively based on previous research. To determine the

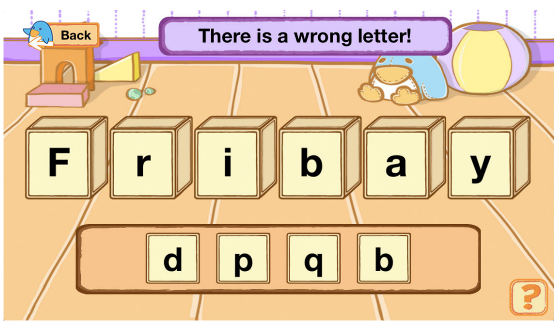


(a) *ABC der Tiere 1* (Animal Alphabet 1; Mildenberger Verlag GmbH, 2016). A complete keyboard is used for writing words. Target word in the right gap is *Butter* [butter].



(b) *ABC Deutsch Lernen Grundschule* (ABC – Learning German in Primary School; Sonnenwald Apps, 2013). Only parts of a keyboard are visible at once, other letters become available by swiping or tapping on arrows. Target word is *Ball* [ball].

(c) *Kinder lernen deutsche Wörter – lesen und schreiben* (Children Learn German Words – Reading and Spelling; Brainy Ape Studio, LLP, 2017). Only pre-defined letters are available to write words. Target word is *Mais* [corn].



(d) *Dyseggxia* (Rauschenberger et al., 2015; Rello et. al, 2014). A misspelled word needs to be corrected using insertion, deletion, or replacing of letters. Target word is *Friday.*

Figure 3.1: Exemplary mobile spelling games.

most appropriate interaction style for the given context, we systematically compare the drag-and-drop, point-and-touch, and touch interaction styles with regard to subjectively perceived workload, user experience, and processing times. The results of our experiment are reported in Chapter 7.

3.3 Summary

Summary of the interaction design challenges in educational applications

- Drag-and-drop, point-and-click (point-and-touch), and click (touch) interaction styles are commonly used in educational spelling applications to interact with letter objects to write words.
- The current state of research is contradictory when it comes to defining the most appropriate interaction style for children. For example, point-and-click is often reported to be more appropriate than drag-and-drop – other times the opposite.
- Various factors may influence the results reported in scientific studies, such as
  - interaction modality: touch vs. computer mouse,
  - particular implementation of an interaction style,
  - context and type of task,
  - age and interaction experience of participating children.
- The interaction styles have not been compared systematically in educational spelling games for children, although they are all frequently used in published spelling games.

Contributions of this thesis

- Systematic literature review on the appropriateness of the interaction styles for the use with children.
- Systematic investigation of drag-and-drop, point-and-touch, and touch interaction style with regard to subjectively perceived workload and user experience in a touch-based spelling game. Results of our experiment are reported in Chapter 7.

Chapter 4

Aims and Scope of This Thesis

This thesis addresses three main aspects of the challenges and benefits of digital learning tools to support literacy acquisition in German primary school children.

1.) Development and Evaluation of Prosodiya (Chapters 5 and 8)

Computer- and game-based interventions have shown great promise in the treatment of reading and spelling disorders. The major contribution of this thesis is the development and evaluation of "Prosodiya" – a mobile game-based training to improve reading and spelling abilities of German primary school children in the home environment. The development and evaluation of Prosodiya follows three major principles to achieve the goal of effectively improving children's literacy skills: feasibility, validity, and educational effectiveness.

1.1.) User-Centered and Evidence-Based Development (Chapter 5)

For digital game-based learning to be effective, the right balance between game experience and learning must be found. In particular, I address Prosodiya's pedagogical content and game design.

A) Evidence-Based Pedagogical Approach (Section 5.2). The aim is to develop linguistically sound exercises based on empirical findings on the causes and effective treatment of reading and spelling disorders. In particular, I address exercises to improve syllable stress awareness, syllable segmentation, vowel length distinction, identification of orthographic markers, and spelling.

B) Game Design (Section 5.3). For Prosodiya to be feasible in the home environment, it has to be easy to understand and easy to use by primary school children. Further, it has to deliver good game experience to engage and motivate children with poor reading or spelling skills over several months to maximize learning gains. In particular, I address the design rationales of individual game elements.

1.2.) Empirical Evaluation (Chapter 8)

I investigate the feasibility, validity, and educational effectiveness of Prosodiya in a randomized controlled-field trial. In particular, the evaluation addresses the following research questions:

A) Feasibility (Section 8.2). Is Prosodiya feasible for use in the home environment, i.e., can it be used by primary school children without adult help and does it continue to engage and motivate over several months? How do children perceive individual game elements of Prosodiya?

B) Effectiveness (Section 8.3). What is the effect of Prosodiya on syllable stress awareness and reading and spelling skills?

C) Validity (Section 8.4). Are the exercises included in Prosodiya theoretically sound and effectively implemented? To what extent are syllable stress awareness and reading and spelling skills related? To what extent are these literacy skills related to the children's training performances of individual exercises included in Prosodiya?

2.) Comparison of Interaction Styles in Educational Applications

A) Chapter 3 and Chapter 7. The current state of research is puzzling when it comes to determining the most appropriate (touch) interaction style for children. That is, the question of whether to implement a drag-and-drop, point-and-click, or click interaction style to move and interact with objects in educational applications is challenging. The aim of this work is to determine the most appropriate interaction style for a tablet-based spelling game for primary school children. For this, I compare drag-and-drop, point-and-touch, and touch with regard to mental workload, user experience, and writing time in a lab experiment.

3.) Automatic Generation and Enhancement of Learning Material (Chapter 6)

The creation of appropriate learning content includes, besides the selection of appropriate task items, the provision of such material in audio or written formats, the latter often with visual text enhancement. The provision of content is costly and thus limits the available material to the budget and design space of pedagogical experts of a product. However, the published material does not cover the needs of all children, which would be necessary with regard to internal differentiation. I address these issues by exploring the limitations and potential of automatic generation of language learning material in two specific domains.

A) Automatic Generation of Minimal Pairs (Section 6.1). Minimal pairs, i.e., words that differ in only one sound (e.g., *bin* and *pin*), are commonly used in speech and language therapy. To reduce the effort in providing large audio-dictionaries, I explore the use of text-to-speech tools to automatically generate such minimal pairs. In particular, I present our new method to improve the pronunciation of synthetically generated pseudowords. I address the discrimination of synthetically generated minimal pairs and their pronunciation quality in an online study.

B) **Automatic Input Enhancement of Reading Texts (Section 6.2).** Visually enhanced texts with custom spacing and syllables alternately displayed in different font colors are commonly used in teaching and learning therapy to support literacy acquisition. However, age-appropriate enhanced reading material for older children is scarce and the range of tools that provide automatic visual input enhancement is limited and lack full customization. To address this, I present COAST, a web-based system to automatically enhance syllable structure, word stress, and spacing in texts. The system's feasibility is investigated in a usability study.

Part II

Our Approach

Chapter 5

Prosodiya

As outlined in the background chapter on the causes and treatments of dyslexia, digital training programs are recognized approaches to improve reading and spelling in children with special learning needs. In this chapter, I describe our approach to effectively and in a playful way support reading and spelling development in German primary school children. This chapter is structured as follows: In Section 5.1, I introduce the reader to Prosodiya – our novel digital game-based spelling training, in Section 5.2, I describe the pedagogical content of Prosodiya and explain the different curriculum units, and in Section 5.3, I describe the game design of Prosodiya and highlight the most important game elements.

5.1 Introduction to Prosodiya

Prosodiya is our novel approach to support German primary school children to independently improve their reading and spelling skills in the home environment. Prosodiya primarily aims at improving syllable stress awareness, the awareness of linguistic features related to syllable stress, and ultimately spelling abilities. The training program, implemented as a serious game for mobile touch devices, is based on recent empirical findings and is, to some extent, similar to evidence-based rule-based spelling trainings (e.g., Ise & Schulte-Körne, 2010; Reuter-Liehr, 1993).

Prosodiya differs from similar empirically evaluated approaches in that it focuses on syllable stress awareness and on linking the linguistic features related to syllable stress to orthographic regularities of German orthography, such as vowel length markers. These abilities, as explained in Section 2.1.2, play a special role in literacy acquisition and are specifically impaired in dyslexic children. This is where Prosodiya comes in. The training program shifts the children's attention to relevant areas of words to clarify the association between syllable stress and orthographic marking of long and short vowels and teaches the children how such syllables are spelled. In doing so, it ultimately leads to a rule-based orthographic spelling training.

This part of the thesis is primarily based on our articles “How to train your syllable stress awareness – A digital game-based approach for German dyslexic children” (Holz, Brandelik, et al., 2018) and “Design Rationales of a Mobile Game-Based Intervention for German Dyslexic Children” (Holz, Beuttler, & Ninaus, 2018), to which I refer the reader for further information on the design rationales of its game elements and its pedagogical content. The version of the game described in the following refers to the currently publicly available version (`https://app.prosodiya.de`, as of 29/01/2019) that was continuously improved based on the feedback received in pilot (Holz, Brandelik, et al., 2017) and effectiveness studies (cf. Holz, Beuttler, & Ninaus, 2018; Holz et al., unpublished; Holz, Ninaus, Meurers, & Kirsch, 2018).

The development and evaluation of Prosodiya are the major contributions of this thesis to extend the current state of research in the field of digital (game-based) training programs. This chapter is limited to the description of the design, development, and implementation of Prosodiya. I report the empirical evaluation with regard to feasibility, training and game experience, educational effectiveness, and validity in Chapter 8.

5.1.1 Requirements of Prosodiya

Prosodiya targets German dyslexic children aged 5–12. Due to this target group, specific requirements, which were derived from interviews with game experts, practitioners, and learning therapists, must be met.

Hence, Prosodiya was designed following three major principles: feasibility, validity, and educational effectiveness.

Feasibility. The main purpose of Prosodiya is to improve reading and spelling in primary school children in addition to classroom and learning therapy, primarily in the home environment. For this, Prosodiya must be feasible to be used at home by children without adult help. Thus, Prosodiya has to be easy to understand and easy to use by primary school children and deliver good game experience to engage and motivate children with poor reading and/or spelling skills over several months to maximize learning gains. Feasibility and training experience can be assessed by investigating children’s training behavior and feedback collected in questionnaires covering questions on the game’s usability, game experience, perceived self-efficacy, perceived usefulness, and other metrics.

Validity. For Prosodiya to be an effective treatment, the pedagogical content must be linguistically sound and based on empirical findings on the causes and effective treatment of reading and spelling disorders. Importantly, it is not enough that the pedagogical approach is theoretically sound, but we also have to ensure that it is effectively implemented in the game. Validity of the pedagogical approach can be assessed by investigating the relationships between syllable stress awareness and reading and spelling abilities. Validity of the implementation can be assessed by investigating the relationship between these literacy skills and children's training performances.

Effectiveness. Finally, to improve reading and spelling in the home environment – to be educational effective – is the ultimate goal of Prosodiya. To prove the educational effectiveness of our treatment approach, we must not only demonstrate its educational efficacy, i.e., showing that it works under optimal conditions, but also demonstrate that Prosodiya has a significant effect on reading and spelling in the real world, under real-world conditions. The training's effectiveness can be assessed by evaluating changes in syllable stress awareness and reading and spelling abilities over the course of a field trial in which children practice at home under real-world conditions.

5.1.2 Iterative Children-Centered Game Design

Digital game-based interventions need to deliver high user and gaming experience while being properly designed in terms of educational effectiveness and learning (Arnab et al., 2015; Ke & Abras, 2013; Plass et al., 2015; Quinn, 2005). Although children have their own preferences and needs and may not be seen as "just short adults" (Druin, 2002), developers often approach parents or teachers to ask about their children's needs rather than asking the children directly (Druin, 1999). However, if design decisions are not adapted to the target group, it may pose barriers and make the game less accessible, particularly to children with special educational needs (Durkin, Boyle, Hunter, & Conti-Ramsden, 2013). Thus, including dyslexic primary school children throughout the whole design and development process was necessary. According to Druin (2002), children can take the role of the user, tester, informant, and design partner. Therefore, we propose and utilized a design and development approach called iterative children-centered game design (ICCGD) to always focus on the target group. The ICCGD combines the two familiar and successful approaches of user-centered design (Abras, Maloney-Krichmar, & Preece, 2004; Norman & Draper, 1986, UCD) and iterative game design (Fullerton, 2008, IGD). Accordingly, we followed the ICCGD for each part of the game.

The ICCGD consists of an initial requirement and context analysis based on expert interviews and observations of paper-based prototypes tested in learning therapy. The result is a first concept of the game. After that, two main phases in different development stages are constantly repeated until the game reaches release. These two main phases again consist of three constantly repeating sub-main phases – playtesting, evaluation, and refinement. In the first main phase, the prototype is tested and evaluated internally with team members and experts following the IGD ("internal iteration"). The second main phase is referred to as "user-centered iteration," in which the internally refined prototype is tested by the children in a session

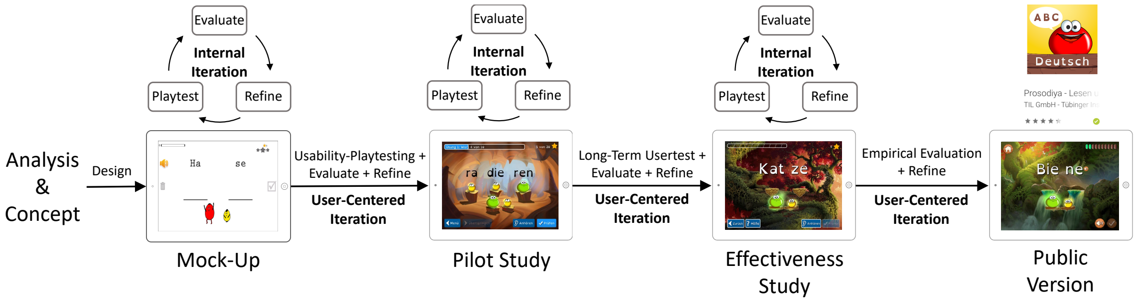


Figure 5.1: Exemplary Iterative Children-Centered Game Design Process for the game "Stress pattern".

of usability-playtesting or unsupervised user tests over a longer period of time. An example of the ICCGD can be seen in Figure 5.1.

5.1.3 Defining Prosodiya as Game-Based Learning

According to the definition of Plass et al. (2015) that digital game-based learning includes games for educational purposes focusing on designing activities as playful tasks, I define Prosodiya as a digital game-based spelling training. That is, we carefully crafted the balance of the design of learning objectives and game play. Plass et al. (2015) sees the most important arguments for digital game-based learning in motivation, engagement, adaptivity, and graceful failure. To address these arguments, researchers mostly agree that *content and skills*, *game mechanics*, *visual aesthetics*, *narrative*, *incentives*, and *musical score* are the building blocks of digital game-based learning (Plass et al., 2015, see Figure 2.2 on page 20).

Accordingly, I address these building blocks to describe Prosodiya. I direct the reader to Holz, Brandelik, et al. (2018) for a closer look on Prosodiya's pedagogical objectives (game mechanics and content and skills) and to Holz, Beuttler, and Ninaus (2018) for detailed analysis and evaluation of Prosodiya's visual aesthetic design, narrative and player progress, reward system, and other game elements. Videos demonstrating Prosodiya and highlighting different aspects can be accessed at `https://youtube.prosodiya.de`.

In the next sections, I first describe the educational content implemented in Prosodiya, referring to the building blocks *game mechanics* and *content and skills*. Then, I explain the design of individual game elements.

Summary of the introduction to Prosodiya

Prosodiya

- Prosodiya aims at improving syllable stress awareness, the awareness of linguistic features related to syllable stress, and spelling.
- Prosodiya differs from similar approaches in that it focuses on syllable stress awareness and on linking the linguistic features related to syllable stress to orthographic regularities of German orthography, such as vowel length markers.
- We define Prosodiya as a digital game-based spelling training.

Major design principles of Prosodiya

- **Feasibility**: Prosodiya should be usable by children unassisted at home and engage and motivate over a longer periods of time.
- **Validity**: The pedagogical approach must be based on empirical findings on the causes and effective treatment of reading and spelling disorders. Further, the content must be implemented adequately.
- **Effectiveness**: The training intends to improve children's language skills when used at home under real-world conditions.
- The development follows an iterative children-centered game design process to find the right balance between game experience and learning.

Contributions of this thesis

- Design and development of Prosodiya. The educational content is described in Section 5.2 and the game design in Section 5.3.
- Evaluation of the feasibility, effectiveness, and validity of Prosodiya in a randomized controlled field trial in order to fill the research gap of empirically evaluated digital game-based spelling interventions for the home environment. The evaluation of Prosodiya is reported in Chapter 8.

5.2 Educational Content

In this section, I explain the educational content of Prosodiya, which refers to the building blocks *game mechanics* and *content and skills* of game-based learning (cf. Section 2.3.3). To date, the first module of Prosodiya has been published that focuses on syllable stress awareness, syllable segmentation, vowel length distinction, orthographic marking of long and short vowels, and spelling. Further modules that focus on, among others, morphological skills (e.g., identifying word stems), are subject to development.

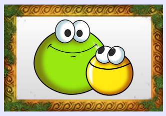
In the following, I first briefly explain the structure of Prosodiya's curriculum and difficulty adjustment. Then, I explain in detail the educational content that is covered in five curriculum units.

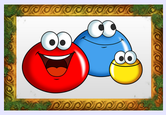
5.2.1 Curriculum and Difficulty Adjustment

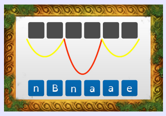
Scope of the First Module of "Prosodiya"

The five curriculum units of Prosodiya's first module

I. Identification of syllable stress and syllable boundaries.
II. Vowel length distinction, i.e., identifying open (long vowels) and closed (short vowels) syllables.
III. Identification of orthographic markers of open and closed syllables.
IV. Spelling of words.
V. Consolidation and automation of acquired knowledge.







The curriculum of Prosodiya is divided into five curriculum units and includes four individual levels whose difficulties increase at different rates throughout the game, see Figure 5.2. The difficulty addresses task-specific characteristics, i.e., changing the complexity of a task and the orthographic complexity of words.

At the top level, different linguistic or orthographic skills are covered in individual units. These skills range from syllable stress awareness to vowel length distinction, identification of orthographic markers for long and short vowels, and, finally, applying spelling rules.

On the second level, units consist of one or more chapters, depending on the scope of the unit. For example, the third unit "Orthographic markers" is split into two chapters, whereas the first chapter deals with the orthographic marking of open syllables (long vowels) and the second chapter deals with the orthographic marking of closed syllables (short vowels).

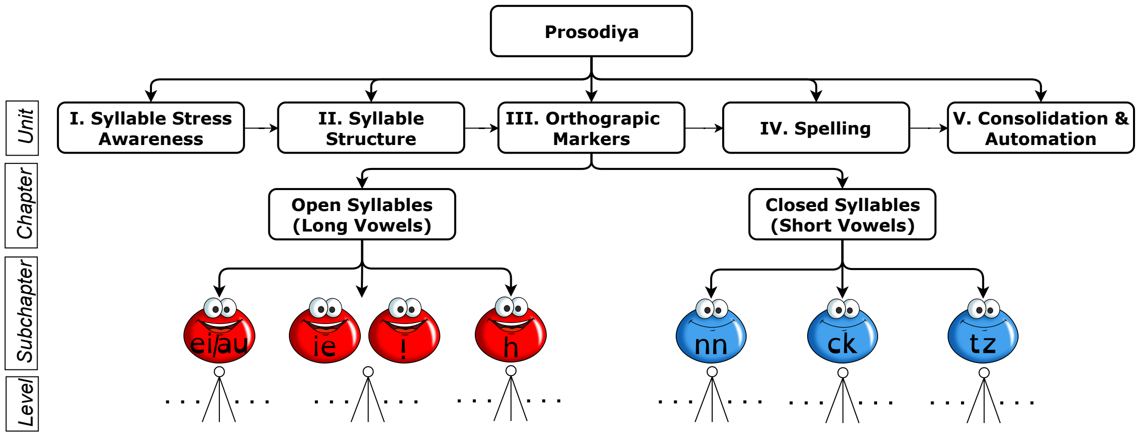


Figure 5.2: Overview of the pedagogical structure of the present version of Prosodiya. The game increases in complexity and difficulty on four levels at individual rates: units, chapters, subchapters, and levels.

At mid-level, subchapters within a chapter deal with different linguistic or orthographic sub-competencies. For example, the chapter on the orthographic marking of long vowels first deals with diphthongs, then with the spelling of the long *i* (i.e, the bigram *ie* and exceptions), and finally with the "silent h".

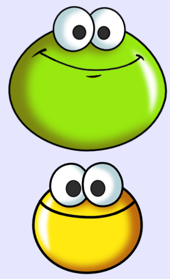
Lastly, levels within a subchapter increase in difficulty of the words' structures and complexities as well as in task complexity. For example, the orthographic complexity of words increases as follows: First, phonetically accurate words are trained, i.e., words that are spelled exactly how you hear them (each letter represents one spoken sound). Then, word length and number of syllables increases. Third, words with consonant clusters are practiced, and lastly words with vowel length markers are covered. On the other hand, task complexity increases by decreasing hints and support provided to the children. For example, the game "Stress pattern" starts displaying target words syllabified and reveals the number of syllables to the children. Later, the written word is replaced by a corresponding image and/or the number of syllables is not revealed to the children, which results in tasks that also include syllable segmentation.

The word selection as well as unlocking of new content adapts to the individual proficiency level of each child.

5.2.2 Unit I: “Syllable Stress Awareness”

Scope of Unit I “Syllable Stress Awareness”

- Learn that words are formed of syllables.
- Learn that not all syllables are equal: Stressed syllables are louder, longer, and oftentimes higher in pitch than unstressed syllables.
- Identify the stress pattern of frequent German words.
- Segment words into syllables.



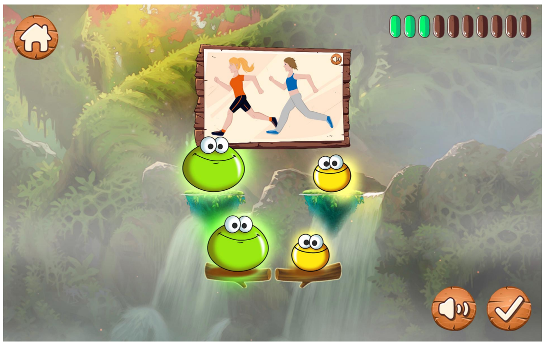
In the first unit of Prosodiya, children train their syllable stress awareness by producing stress patterns of given words. They do so by dragging and dropping cartoon characters onto platforms. The characters are referred to as *“Kugellichter”* [“spherical lights”] and represent the pedagogical agents of Prosodiya, see Section 5.3.3. A big green Kugellicht is used for stressed syllables and a small yellow Kugellicht for unstressed syllables, see Figure 5.3. In case of the word *REN-nen* [to run], the first syllable *ren* is stressed, while the second syllable *nen* is unstressed, following the typical German trochee. We provide three different sound files for each word that increase with regard to the intensity of the intonation. If children request help or submit a wrong answer, the word is spoken in the next stronger intonation level to give scaffolding feedback.

This unit continuously increases in difficulty in that the word length and complexity of the orthographic structures of the target words increases and less frequent stress pattern are practiced. Additionally, the number of syllables is not always revealed to the children (see Figure 5.3b) and the displayed written word may be replaced by a corresponding image (see Figure 5.3c).

As we received feedback in our effectiveness study that children wished for more variety in the tasks during the first unit of the game, we also implemented a task of syllable counting for the public version, see Figure 5.4. Additionally, easy spelling games (cf. Section 5.2.5) are also introduced in the first unit of the game.

While other digital approaches already focus implicitly on the stressed syllable by teaching children to detect the “king of the syllable” (e.g., Ernst Klett Verlag, 2015), for which the children are asked to mark the vowel of the first syllable of disyllabic trochaic words, Prosodiya is the first digital approach to explicitly focus on identifying syllable stress.

(a) Low difficulty: syllabification as well as number of syllables are revealed to the learner.



(b) Medium difficulty: words are presented by images and no longer in written form, but the number of syllables is revealed.

(c) High difficulty: neither the number of syllables nor syllabification are revealed to the learner, additionally requiring syllable segmentation to solve the task.

Figure 5.3: Game 1: "Stress pattern". Children identify stress patterns by placing the Kugellichter on respective platforms. The big green Kugellicht is used for stressed syllables, the small yellow Kugellicht for unstressed syllables. Three different levels of difficulty are shown in Figure 5.3a, Figure 5.3b, and Figure 5.3c. The disyllabic target word *REN-nen* [to run] is stressed on the first syllable.

Figure 5.4: Game 5: "Syllable counting". Children count the number of syllables by pressing the "+" and "–" buttons. The trisyllabic target word is *er-IN-nern* [to remember].

5.2.3 Unit II: "Syllable Structure"

Scope of Unit II "Syllable Structure"

- Learn that the vowel phoneme of stressed syllables can be short (closed syllable) or long (open syllable).
- Open syllables end with a vowel. The long vowel can be lengthened, keeping the mouth open, without the word sounding weird.
- Closed syllables end with a consonant. The "stopper" squeezes the vowel and the mouth is closed at the lips, the teeth, or by the tongue.


In the second unit, children work on perceiving and distinguishing the length of the vowel phoneme of the stressed syllable. For this, we implemented a novel variant of the commonly used vowel length distinction task that builds upon the competence of stress pattern recognition. In addition to identifying syllable stress, the children have to decide whether the stressed syllable is open (the syllable ends with a long vowel, big red Kugellicht with its mouth open) or closed (the vowel is closed by a consonant, big blue Kugellicht with closed mouth), see Figure 5.5. Again, due to the feedback received in our effectiveness study to add more variety to Prosodiya, we implemented an additional simplified version of this game in which children only need to identify the vowel length, without rebuilding the stress pattern, see Figure 5.5c. The difficulty of this unit increases similarly to the first unit.

We provide sound files of minimal pairs for each word to support the learner when they require help or submit wrong answers. The minimal pairs consist of the correct pronunciation of the word and a pseudoword counterpart for which the vowel length of the stressed syllable was changed to the contrary.

In this unit, we also address mouth motor activities by teaching the children that at the end of open syllables, they can continuously lengthen the vowel, which keeps the mouth open. At the end of closed syllables, however, the consonant is "stopping" and "squeezing" the vowel and the mouth is closed at the lips, the teeth, or by the tongue. The wording of *open* and *closed* is also reflected in the features of the mouth of the Kugellichter. Children with dyslexia have difficulties permeating the sound level of a language in order to improve letter-sound correspondence on the segmental level (Moll, Wallner, & Landerl, 2012). Mouth motor activity, which is intact in dyslexic children (Schulte-Körne & Remschmidt, 2003), can be used to facilitate learning of letter-sound correspondence (Boyer & Ehri, 2011). The approach of including mouth motor activity is novel to computer-based interventions.

Processing the structure of the stressed syllable is a necessity for acquiring the complex spelling rules that underlie spelling of long and short vowels in German orthography (see also Section 2.1.2), which is covered in the next unit of the game.

(a) Advanced stress pattern game for the word *REN-nen* [to run], whose stressed syllable contains the short vowel *e*.

(b) Advanced stress pattern game for the word *Ba-NA-ne* [banana], whose stressed syllable contains the long vowel *a*.

(c) Vowel length distinction game for the word *SCHLIT-ten* [sleigh], whose stressed syllable contains the short vowel *i*.

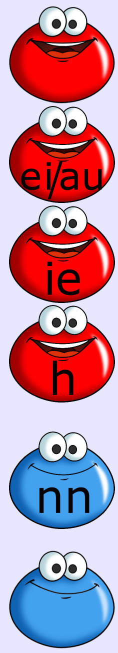
(d) Vowel length distinction game for the target word *FRIE-ren* [to freeze], whose stressed syllable contains the long vowel *i*.

Figure 5.5: Game 2: "Open and closed syllables" – or "Vowel length distinction". Children learn to distinguish whether the stressed syllable is open (the syllable ends with a long vowel; big red Kugellicht with its mouth open) or closed (the syllable ends with a consonant, resulting in a short vowel; blue Kugellicht with closed mouth). Figure 5.5a and Figure 5.5b show an advanced version of the game "Stress pattern", in which children have to identify the stress pattern of words with extra focus on the stressed syllable's vowel length. Figure 5.5c and Figure 5.5d show simple vowel length distinction games.

5.2.4 Unit III: "Orthographic Markers"

Scope of Unit III "Orthographic Markers"

- Learn how open syllables are spelled. The long vowel phoneme can be
  - unmarked (e.g., *RE-den* [to talk]),
  - marked with a diphthong (double vowel, e.g., *LAU-fen* [to walk]),
  - marked with the bigram *ie* in case of the German long *i* (e.g., *FLIE-ge* [fly]),
  - marked with a silent h (e.g., *NEH-men* [to take]).
- Learn how closed syllables are spelled.
  - If the short vowel phoneme is followed by only one consonant, the consonant has to be doubled in the spelling (e.g., *KEN-nen* [to know sb. or sth.]).
    - ∗ Instead of doubling *z* and *k*, the graphemes *tz* and *ck* are used in spelling.
  - If the short vowel phoneme is followed by two or more consonants, the consonant closing the syllable is not doubled in the spelling (e.g., *BLIN-ken* [to flash]).



After the children acquired the knowledge about syllable stress and the structure of the stressed syllables, they learn the rules that underlie the spelling of open and closed syllables Prosodiya's third unit. This unit includes two different game types in which children first learn to recognize the orthographic marker that belongs to the vowel of the stressed syllables, see Figure 5.6 and Figure 5.7 on page 52, and then spell out the word in a simplified spelling game, see Figure 5.8 on page 54. I refer to Section 2.1.2 for the detailed explanation of the marking of long and short vowels in German orthography.

First, children learn about the orthographic marking of long vowels and later about the marking of short vowels. They learn that long vowels can be (i) not marked orthographically (e.g., *RA-ten* [to guess]), (ii) marked with a diphthong (double vowel, e.g., *DAU-men* [thumb]), (iii) marked with the bigram *ie* in case the vowel is a long *i* (e.g., *BIE-ne* [bee]), or by adding a "silent h" (e.g., *FEH-len* [to miss]). In case for the long *i*, unmarked exceptions are also taught (e.g., *TI-ger* [tiger] or *Man-da-RINE* [tangerine]). Words that are marked by adding a silent h are exceptions that do not follow explicit rules and must be memorized and learned by heart with memos such as "*Das stumme h, das ist nicht schwer, steht meist vor l, m, n, und r*" [the silent h precedes mainly but not necessarily the letters l, m, n, and r after a long vowel phoneme]. For the children to better memorize words with a silent h, all words that are marked with a silent h that will be practiced in a level (e.g., *KOH-le* [coal], *FOH-len* [foal], and *FAH-ren* [to drive]) are shown

(a) Simple orthographic marker identification: children can only choose between unmarked long (red) and short (blue) vowels and consonant doubling (blue with *nn*) for the word *REN-nen* [to run].

(b) Consolidation task: children have to identify the correct orthographic marking for the word *SCHLIT-ten* [sleigh] from all the markers taught in the game.

Figure 5.6: Game 3: "Orthographic markers". Children need to identify the correct orthographic marker for the vowel of the stressed syllable. Children first need to recognize if the vowel of the stressed syllable is long (red) or short (blue) and then whether the vowel is marked orthographically. Figure 5.6a shows a more simple variation of the game, Figure 5.6b a more advanced variation.

and read out successively at the very beginning of the level, before the first word is practiced.

In the second part of this unit, they learn about the two rules that underlie the spelling of closed syllables. They learn that (i) "if the short vowel phoneme of the stressed syllable is followed by two or more consonants, the "stopper" (the consonant closing the syllable) is not doubled in the spelling (e.g., *FEL-sen* [rock])", and (ii) "if the short vowel phoneme of the stressed syllable is followed by only one consonant phoneme, then the stopper has to be doubled in the spelling as well" (e.g., *REN-nen* [to run]). The ambisyllabic consonant doubling has two special cases that are also trained: *ck* is written instead of *kk* (e.g., *HA-cke* [pick]) and *tz* is written instead of *zz* (e.g., *HIT-ze* [heat]). The orthographic marking of short vowels is taught using the phonetic rules that originate in the stressed syllable of typical German trochees (see Section 2.1.2) and is explained children-friendly as follows: "if you can hear no other consonant after the stopper of the closed syllable before you hear the next vowel, then the stopper must be doubled! For example, in the word *REN-nen* [to run], you can only hear one consonant after the vowel of the closed syllable, the stopper. You can hear a vowel directly after the stopper! In such cases, you can pronounce the stopper twice. If you can pronounce the stopper twice, then you also have to spell it twice!".

The difficulty increases in the phonetic similarity of choices. For example, the chapter about the long vowel *i* starts with comparing words that have an unmarked long vowel with words whose long *i* is marked by the bigram *ie*. Later on, exception words with a long vowel *i* that are not marked orthographically (e.g., *TI-ger* [tiger])

(a) The long vowel phoneme *i* of the word *BIE-ne* [bee] is marked with the bigram *ie*.

(b) The vowel phoneme *i* in the word *WIN-ter* [winter] is short (blue).

Figure 5.7: Game 3 "Orthographic markers" to teach the spelling of the long vowel phoneme *i*, which is typically marked with the bigram *ie*, see Figure 5.7a. In the given levels, children need to identify whether the *i* is long (red) or short (blue, e.g., Figure 5.7b). Exceptions in which the long vowel *i* is not marked orthographically are also practiced (e.g., *TI-ger* [tiger]).

and words with a short vowel *i* (e.g., *WIN-ter* [winter]) are added to the pool of words, see Figure 5.7.

In the course of these chapters, the two games "Orthographic markers" and "Spelling" are used alternately so that the children first learn about the respective orthographic marker and then foster their knowledge by spelling out the words. In this unit of Prosodiya, the "Spelling" game only offers the exact letters of a target word to spell it, resulting in a letter arrangement task.

The different orthographic markers and their linguistic characteristics are introduced in individual tutorials. For example, ambisyllabic consonant doubling (e.g., *nn*, *ck*, *tz*) is explained as follows: "if you can hear no other consonant after the stopper of the closed syllable before you hear the next vowel, then the stopper must be doubled! For example, in the word *REN-nen* [to run], you can only hear one consonant after the vowel of the closed syllable, the stopper. You can hear a vowel directly after the stopper! In such cases, you can pronounce the stopper twice. If you can pronounce the stopper twice, then you also have to spell it twice!".

This unit of Prosodiya is particularly important as the training to recognize orthographic markers is crucial for spelling acquisition (Galuschka et al., 2014; Landerl, 2003), and the inclusion of algorithms of spelling rules to detect and apply orthographic marking has been successfully shown to improve spelling (e.g., Ise & Schulte-Körne, 2010; Kargl & Purgstaller, 2010) and is recommended by clinical practical guidelines (Galuschka & Schulte-Körne, 2016). However, the algorithms to determine orthographic marking of vowel length have not been related to syllable stress in other computer-based interventions before.

5.2.5 Unit IV: "Spelling"

Scope of Unit IV "Spelling"

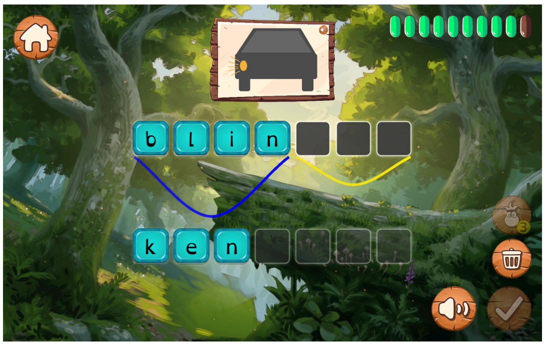
- Foster the previously acquired metalinguistic knowledge of the spelling of long and short vowels by writing the words. Distracting letters vary the difficulty of this game unit.
  - No distracting letters → letter arrangement task
  - Easy distracting letters that share no phonological similarities to actual letters of the word → letter discrimination and arrangement task
  - Difficult distracting letters that may lead to phonologically very similar or homophonic misspellings

The fourth unit primarily focuses on spelling words to foster children's previously acquired knowledge. In spelling games, children pick letters from the letter area and arrange them in the spelling line, see Figure 5.8. The letter area contains a predefined set of letters that each can be used once to write the word.

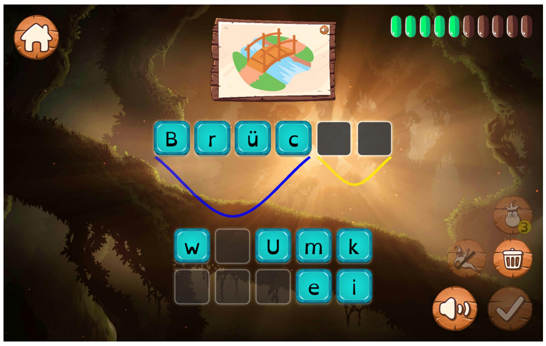
Easy Spelling Game. In easy spelling games, no distracting letters are used, resulting in a letter arrangement task. In addition, syllable arcs are drawn underneath the spelling line in some conditions to help link the awareness of orthographic markers to the stressed syllables and to help in syllable segmentation. The colors of the syllable arcs refer to syllable stress and vowel length: yellow for unstressed syllables, red for open stressed syllables, and blue for closed stressed syllables.

Difficult Spelling Game. In comparison to the spelling games practiced earlier, this chapter increases the difficulty by adding distracting letters to the set of available letters. These distracting letters are either not part of the written word or duplicates of present letters. This unit of the game increases the difficulty of the spelling game in terms of adjusting the phonological similarity of distracting letters to actual letters of the word. First, distracting letters that do not share phonological similarities to any letter of the word are used, resulting in a letter discrimination task, see Figure 5.8b. Later on, distracting letters that can lead to phonologically very similar or even homophonic misspellings are used, see Figure 5.8c. Homophonic words sound alike but are misspelled or have a different meaning. For example, the letters {*ä*, *h*, *l*, *m*} are added to the word *FEL-sen* [rock] that may lead to homophonic misspellings, such as *FEL-lsen* or *FÄL-sen*, or to phonologically very similar misspellings such as *FEL-sem* or *FEH-lsen*. To make the chapter more varied, the other games are also practiced.

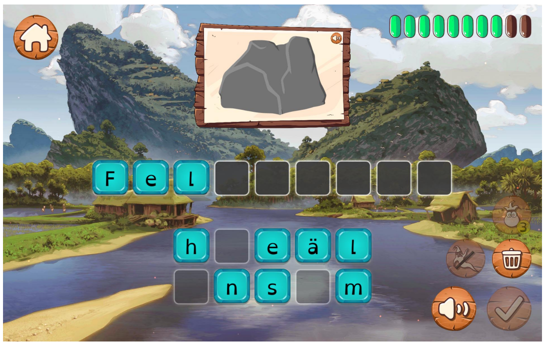
To support scaffolding feedback, individual letters can be solved or distracting letters can be deleted after the children entered a misspelled word.



(a) Easy spelling game for the word *blinken* [to flash]: the available set of letters does not include distractors, resulting in a letter arrangement task.



(b) Medium difficult spelling game for the word *Brücke* [bridge]: distracting letters that do not share phonological similarities to actual letters of the word are added to the pool of letters, resulting in a letter discrimination task.



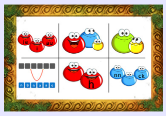
(c) Difficult spelling game for the word *Felsen* [rock]: distracting letters that may lead to phonologically similar or even homophonic misspellings are added to the pool of letters.

Figure 5.8: Game 4: “Spelling”. Children arrange letters in the spelling line to write single words. The game comes in various difficulty levels. Figure 5.8a displays an easy version without distracting letters, Figure 5.8b a medium difficult version with distracting letters that do not share phonological similarities to actual letters of the word, and Figure 5.8c displays a difficult version with distracting letters that may lead to homophonic or phonologically similar misspellings.

5.2.6 Unit V: "Consolidation and Automation"

Scope of Unit V "Consolidation and Automation"

- Consolidate previously acquired linguistic knowledge and automate reading and spelling processes by practicing with all games.



In the fifth unit of Prosodiya's first module, children consolidate their previously acquired linguistic knowledge about German orthography. For this, all games of the previous units are practiced in medium or hard difficulties to automate reading and spelling processes.

To conclude, the educational content of Prosodiya's first module targets at improving children's (i) syllable stress awareness and syllable segmentation, (ii) vowel length distinction, (iii) identification of orthographic markers, and (iv) spelling.

5.2.7 Word Material

The trained word material of the experimental version consists in total of 399 words taken from the *Grundwortschatz GUT1* (Basic Vocabulary GUT1; Grund, n.d.), the *Marburger Rechtschreibtraining* (Marburg Spelling Training; Schulte-Körne & Mathwig, 2013), the *Kieler Leseaufbau* (Kiel Reading Training; Dummer-Smoch & Hackethal, 2011), and the *childLex* (Schroeder, Würzner, Heister, Geyken, & Kliegl, 2015).

As the orthographic regularities trained in Prosodiya generally apply to the trochaic word form,[1] the experimental version only included words in their base forms and non-compound nouns. Plural is used in case of monosyllabic nouns (e.g., the plural form *Bäu-me* [trees] is trained instead of *Baum* [tree]). Morphological inflection, i.e., conjugation and declension, is not yet covered. Exercises to deduce the orthographic marking of inflected words, such as to learn that the inflected word form *rennt* [he/she/it runs] is spelled with an ambisyllabic consonant doubling as it is derived from the orthographically marked base form *rennen* [to run], are currently being developed.

[1] This also includes trisyllabic words with an unstressed prefix, such as *ver-LIE-ren* /fɛɐ̯.ˈliː.ʁən/ [to loose]

5.2.8 Summary

Summary of the educational content of Prosodiya

Curriculum and difficulty adjustment

- Prosodiya's first module is divided in five curriculum units that focus on syllable stress awareness, syllable segmentation, vowel length distinction, orthographic vowel length marking, and spelling.
- The game progresses on four individual levels whose difficulties increase at different rates: (i) units: linguistic and orthographic skills, (ii) chapters and (iii) subchapters: linguistic and orthographic subcompetencies, and (iv) levels: word structure and complexity.

Unit I

- In the first unit, children learn to identify syllable stress and to apply syllable segmentation by practicing with the game "Stress pattern", which increases in complexity, and with the game "Syllable counting".

Unit II

- In the second unit, children learn to distinguish vowel lengths in stressed syllables (long vs. short) and to differentiate open (end with a vowel) from closed syllables (end with a consonant) by practicing with the games "Open and closed syllables" and "Vowel length distinction".

Unit III

- In the third unit, children learn about the orthographic marking of long and short vowels in different versions of the game "Orthographic markers" and in easy "Spelling" games.
- Marking of long vowels includes (i) unmarked long vowels, (ii) dipthongs, (iii) the long vowel phoneme *i* that is marked with the bigram *ie* and its exceptions, and (iv) the "silent h".
- The orthographic marking of short vowels (i.e., ambisyllabic consonant doubling and consonant clusters) is taught using the phonetic rules that originate in the stressed syllable of typical German trochees.

Unit IV

- In the fourth unit, children foster their previously acquired knowledge in spelling games. The "Spelling" game increases in complexity by adding distractor letters of increasing phonological similarity.

Unit V

- In the fifth unit, children consolidate and automate their spelling processes by practicing with all games.

Word material

- Prosodiya currently includes a total of 399 words taken from different basic vocabularies. As trained orthographic regularities generally apply to the trochaic word form, the present version only includes words in their base forms and non-compound nouns with at least two syllables.

5.3 Game Design Elements

The educational content explained in Section 5.2 determines the validity and efficacy of our training program. However, feasibility and training experience also play a major role in the effectiveness of such treatment approaches. On one hand, the usage of Prosodiya is restricted if children cannot play independently. On the other hand, if children have a bad game experience, it is more likely that they quit playing before the training is completed, putting the efficacy at risk. As explained in Section 2.3.3, game design elements are used in learning environments to positively engage the learner and to invoke position emotions in order to positively affect learning (Hamari et al., 2016; Plass, Heidig, Hayward, Homer, & Um, 2014) and to increase motivation, satisfaction, and perception towards the learning material (Um, Plass, Hayward, & Homer, 2012). In the following, I describe our approach to keep children engaged with the game and to enable the training to be used unassisted. First, I describe Prosodiya's visual design, followed by the description of its narrative, environment, and game progress. Then, I introduce the pedagogical agents, followed by the explanation of the game's tutorials. Then, I explain the implemented feedback. Finally, I describe the implemented rewards and incentives.

5.3.1 Visual Aesthetic Design

As concluded that "Learning is fun if it increases without boundaries in difficulty. As long as the graphics are good." (Berkling, Faller, & Piertzik, 2017, p. 7), we aimed to make the game's graphical appearance appealing, consistent, and simple. To address the requirement of having high quality graphics, we collaborate with a renowned comic artist and licensed images from the comic "The Wormworld Saga" (`https://wormworldsaga.com`) and adjusted them for our needs to fit the story, atmosphere, and mechanics of our game. The characters (cf. Section 5.3.3) and interface elements are designed by different artists. The pictures representing words (e.g., Figure 5.3a on page 47) were also designed iteratively to fit in the game's overall look and feel and to describe the words as best as possible. Further, we use OpenDyslexic (Gonzalez, 2014) as the font, which is designed specifically for dyslexics.

To keep it simple and not overload the children, we limit the game screen to game elements that are required by respective activities and forego additional elements that may distract the learner or hinder learning.

5.3.2 Narrative, Environment, and Game Progress

Prosodiya is embedded in a fantasy-themed setting, which has been shown to be beneficial for motivation, involvement, and learning (Cordova & Lepper, 1996; Parker, Lepper, Bartholomew, Cordova, & Mayer, 1992). The eponymous fantasy world is haunted by a mysterious fog that covers all the peaceful land, causing the inhab-

Figure 5.9: In-game map. All regions except for the final chapter – the *Magic Forest* – have been successfully freed from the mysterious fog that is haunting Prosodiya.

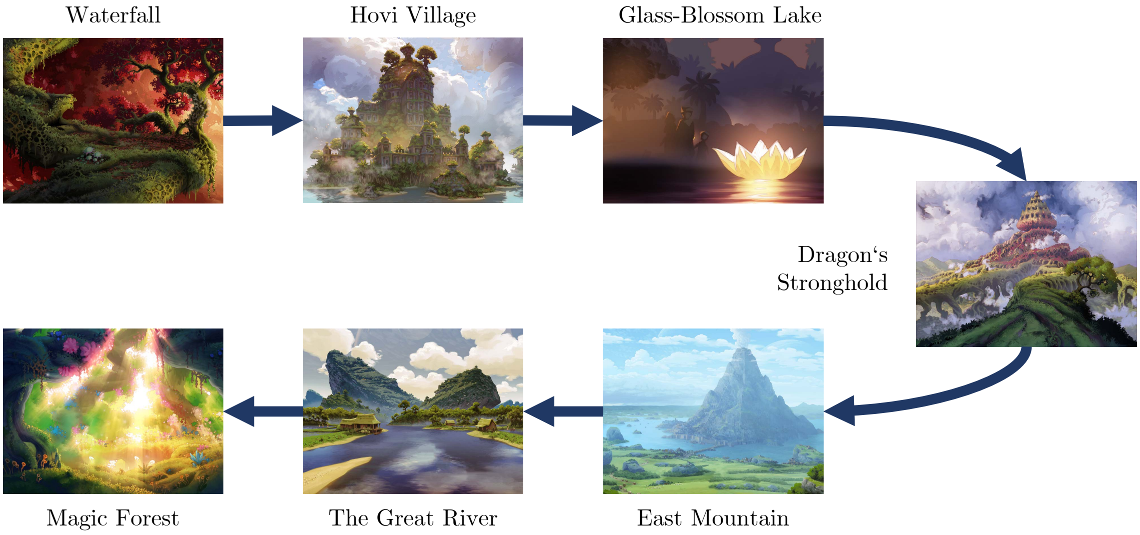


Figure 5.10: Journey through the world of Prosodiya as reflected in the game's narrative. Each image represents an exemplary background for its eponymous chapter.

itants to live in worries and sorrow, see Figure 5.9. The little inhabitants called Kugellichter are the game's protagonists and pedagogical agents, seek the children's help as they themselves are too weak to help their homeland. Only the children, guided by the Kugellichter through the world of syllables and orthography, can disperse the suppressing fog and free the land from its dreadful destiny. In order to decipher the mysteries of German orthography and obtain the *"wisdom of words"*, they need to understand and use the *"power of the stressed syllable"*.

The story relates to worries and needs of families of children with dyslexia. Affected children often experience the difficulty of literacy acquisition as an impenetrable fog – they feel like there is "no land in sight". The game aims at helping children to clear their blurred visions and to feel comfortable in the world of reading and writing. Our aim was for the story to relate to real-life struggles of affected children in order to positively affect their lives beyond the game's world.

The narrative, environment, and game progress matches the progression of the three lower levels of Prosodiya's curriculum and difficulty system, i .e., chapters, subchapters, and levels (cf. Section 5.2.1). Each chapter is embedded in a unique environment, and has an eponymous landmark that needs to be freed from the fog, which is reflected by the map and level-based environments of subchapters, see Figure 5.9 and Figure 5.10. The children's journey starts at the *Waterfall* – the source of the

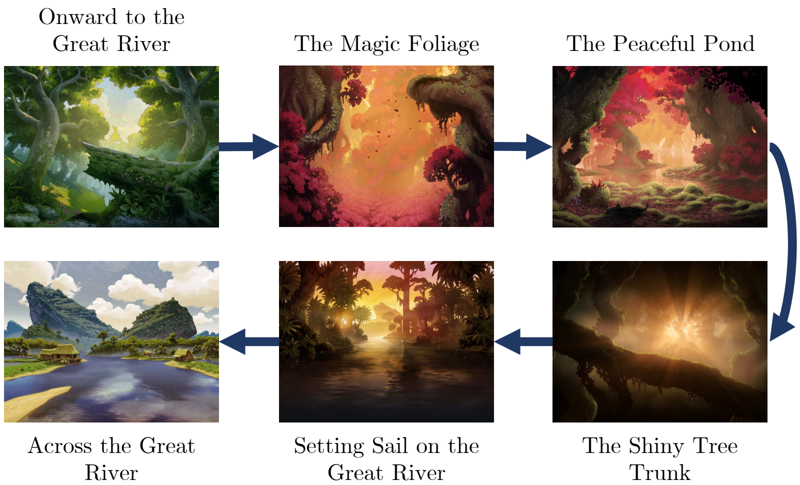


Figure 5.11: Progressing through the chapter *The Great River*. Each subchapter has a unique background environment.

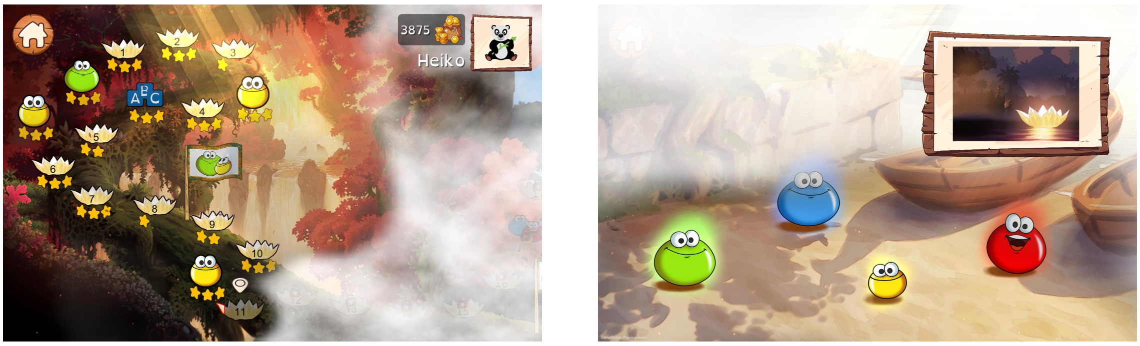
stressed syllable's power – before it takes them through the *Hovi-Village* to rescue its inhabitants, all the way to the *Glass-Blossom Lake* for its purification. Subsequently, the *Dragon's Stronghold* leads them to higher grounds, past the *East Mountain* and across *The Great River*, before the journey ends in the *Magic Forest*.

We use three game elements embedded in a weekly and daily progression system to convey the progress of the game: the world map, cutscenes, and change of background environments and atmosphere. While the story is explicitly told in cutscenes narrated by the Kugellichter (see Figure 5.12b), the deliverance of regions is also reflected on the map (see Figure 5.12a) and in changes of background environments used in levels (see Figure 5.10 and Figure 5.11). We implemented this multilevel progress, which also implicitly tells the story by progressing through the levels' backgrounds, to increase the children's self-perception of progression, their perception of positive affect and immersion, and to maintain motivation over longer periods of time. In the following, I explain each of these game elements in more detail.

5.3.2.1 Map

We designed the in-game map of Prosodiya as the "main scene" of the game from the children's perspective, see Figure 5.12a. Each time children progress through the game, corresponding regions on the map are redeemed from the fog and adjacent areas call for their help, awaiting them with new challenges.

On the map, children can either play new levels to make progress and unlock new content, or play old levels to beat their previous high scores and gain stars. We used glass blossoms as level symbols, the yellow Kugellicht to indicate cutscenes, and individual icons for each tutorial. Additionally, flags corresponding to the game's chapters indicate the linguistic challenges that are practiced in the area.



(a) In-game map. Glass blossoms are used as level symbols. To progress in Prosodiya, the level closest to the fog needs to be played, indicated by an animated landmark.

(b) Cutscene “At the shores of the Glass-Blossom Lake”. Kugellichter narrate the story and tell the children about the secrets of the glass blossoms.

Figure 5.12: In-game map (Figure 5.12a) and exemplary cutscene (Figure 5.12b).

5.3.2.2 Cutscenes

In cutscenes, the Kugellichter continue the narration of the story. To support the storytelling, corresponding images are displayed in a wooden frame. For example, in the cutscene displayed in Figure 5.12b, children made their way from the *Hovi-Village* and arrived at the shores of the *Glass-Blossom Lake*. After the path has been cleared, they are asked to dispel the fog from the lake so that the inhabitants of Prosodiya can dive for glass blossoms to regain their power and strength that was lost due to the fog.

In our effectiveness study (cf. Holz et al., unpublished), we received feedback that cutscenes explicitly telling the game’s story and progression are very motivating and were missed in the study version. In the study version, only a prologue of the story was implemented to raise the children’s interest. In the current version, each chapter provides multiple cutscenes.

5.3.2.3 Change of Environment and Atmosphere

The world’s exploration is also reflected in different environments used in levels. The background environment of a level is determined by its corresponding chapter. Chapters are considered as the game’s milestones, represented by unique landmarks, see Figure 5.10. In a pilot version (cf. Holz, Brandelik, et al., 2017), the game only had one background image for each chapter. Many children and their parents reported that the change of setting had a strong positive effect on their motivation and self-awareness of progression. Thus, we added more background environments to the game, which also follows the advice of Berkling et al. (2017) to frequently introduce new elements to the game. In the current version, each chapter provides unique background environments for each of its subchapter that continue the journey through a chapter, see Figure 5.11 on the previous page. That is, the current version of the game provides more than 40 unique background environments. The frequent change of environments makes the game more varied and intends to increase the children's self-perception of progression.

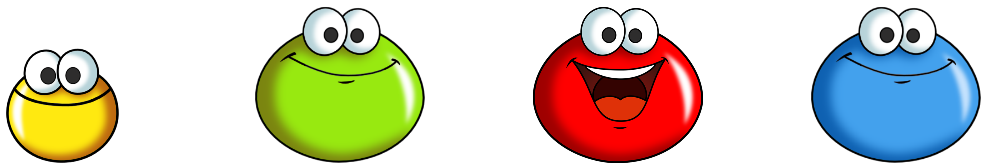
As children will not master a subchapter each day, we included a daily progression. We embedded a fog and lighting system within levels (cf. `http://prosodiya.com/level-run`). For each task solved, the fog lightens up, glowing inhabitants of Prosodiya show up (e.g., fireflies), and other light sources like torches, lanterns, or sun rays appear to brighten the atmosphere. This low-level progress is also emphasized by a progress bar, see Figure 5.3 on page 47.

5.3.3 Pedagogical Agents

A commonly used approach to engage children in games are social interactions via companions (Lim & Reeves, 2010). They can help players to understand game mechanics, teach game interactions, and resemble story-driven elements. In serious games, those companions are often referred to as pedagogical agents. Despite companions, agents are more focused on giving contextual feedback and explanations customized for the children in need (Goldberg & Cannon-Bowers, 2015; Nwana, 1990). In Prosodiya, the Kugellichter are the pedagogical agents and serve different purposes in order to support and accompany the player.

Firstly, the agents introduce themselves as inhabitants of Prosodiya and establish a connection between the player and the world, explaining where they are and what is happening. They take the role of typical companions by the side of the children. In cutscenes, the Kugellichter narrate the story and continue the storyline. Secondly, the pedagogical agents take over the role as linguistic tutors. In each chapter, children get introduced to upcoming linguistic challenges and receive rule-based explanations by the agents on how to solve those challenges, see Section 5.3.4. Thirdly, the agents are also responsible for tutoring linguistic knowledge and provide scaffolding feedback to the input of the user, see Section 5.3.5.

The pedagogical agents were designed with a close feedback loop including learning therapists, children, and game designers. They are designed to both fit into the game mechanics as well as to make them attractive to children. To satisfy all needs, we designed the Kugellichter to be unisex (as advices by Koivisto & Hamari, 2014), child-friendly regarding color and shape, and to facilitate learning. We chose round shape as this may induce positive emotions that facilitate learning and improve comprehension (Plass et al., 2014). To ensure that color blindness or other color impairments do not affect learning, color was not the only unique feature of the pedagogical agents. The agents also differ in their sizes and in the shapes of their mouths. Each agent provides a unique feature that links to its linguistic characteristics and supports clear distinction, see Figure 5.13. There are four types of agents representing different linguistic characteristics. As mentioned before and communicated in the first tutorial, the yellow Kugellicht is smaller in size and represents unstressed syllables, see Figure 5.13a. Its counterparts are the bigger green



(a) The small yellow Kugellicht represents unstressed syllables.

(b) The big green Kugellicht represents stressed syllables.

(c) The big red Kugellicht with open mouth represents open syllables.

(d) The big blue Kugellicht with closed mouth represents closed syllables.

Figure 5.13: The *"Kugellichter"* ["spherical lights"] are the pedagogical agents of Prosodiya. They take over the role of companions, narrators, and linguistic tutors.

(Figure 5.13b), blue (Figure 5.13d), and red (Figure 5.13c) Kugellichter who represent different stressed syllables. The green and blue Kugellicht look alike when disregarding their colors. However, they do not occur simultaneously and, thus, green-blue deficiencies do not adversely affect game play.

5.3.4 Tutorials and Tooltips

In order for Prosodiya to be used by primary school children without adult help, two crucial design elements are instruction and feedback.

We implemented interactive tutorials for each featured game or linguistic characteristic. The Kugellichter function as pedagogical agents and narrators and introduce and explain game mechanics, linguistic competencies and challenges, and impart linguistic knowledge. This instructional support, particularly when focusing on learning new skills and selecting relevant (new) information, has been shown to improve learning (Wouters & van Oostendorp, 2013).

As the tutorials also start new sections of the game, they can be seen as on-boarding phases. The benefits of successful on-boarding phases are crucial for a long-term success of the game (cf. Gabe Zichermann and Christopher Cunningham, 2011) and thus are of special importance.

We kept the tutorials short, simple, and fun and aimed to ensure that children understand the game mechanics as well as the linguistic background. To proceed within a tutorial, children are frequently asked to actively solve the current step, following the instructions of the pedagogical agents, see Figure 5.14a. Besides the instructional support, the tutorials also continue the story in addition to the cutscenes.

We focused on a high level of interactivity to increase the children's participation and to ensure that they understand new game mechanics and linguistic principles. Multiple gameplay iterations showed that, in addition to linguistic explanations, it is important to teach and practice the interactions in the game before new content is practiced.

Based on observations in pilot studies, one detailed and comprehensive tutorial in

(a) Tutorial on the use of *ck*: The yellow Kugellicht explains with the example word *Wecker* [alarm clock] that, instead of doubling the letter *k*, the grapheme *ck* is used in spelling. It then asks the children to move the *ck*-Kugellicht onto the leaf.

(b) Exemplary tooltip that briefly explains game mechanics and, in this case, the use of the consonant doubling *ck*.

Figure 5.14: Tutorials and tooltips teach children the game mechanics and linguistic backgrounds so they can practice without adult help. Figure 5.14a shows the tutorial for a specific version of the game "Orthographic markers", in which the special case *ck* of consonant doubling is explained. Figure 5.14b shows the corresponding tooltip.

the beginning of a chapter is not enough. Children may forget about the objective of the game, its game mechanics, or about linguistic and orthographic characteristics, especially when they take a longer break from the game. Hence, we also implemented short and spot-on task explanations, so-called tooltips, that appear at the start of each level and that can be accessed manually during play, see Figure 5.14b. The spot-on content consists of a spoken explanation with the voice of the yellow Kugellicht and a simple image of the level's objective and challenges. Depending on the degree of difficulty, the children may also get additional hints on what has changed in the gameplay or what to pay attention to.

5.3.5 Feedback

Besides instructions, feedback is crucial for knowledge improvement and skill acquisition and may affect the motivation of learners (cf. Shute, 2008). Prosodiya uses scaffolding and knowledge of correct response (KRC) feedback. Scaffolding feedback may help dyslexic children to solve exercises faster (Kazakou & Soulis, 2015) and KCR feedback has been shown to support memorization and deeper learning (e.g., Corbalan, Kester, & J.G. van Merriënboer, 2009; Erhel & Jamet, 2013).

The feedback depends on the children's answers and is as follows: if the answer is correct, a positive sound is played, stars are collected and added to the current score, the progress bar is adjusted, and game elements respond positively, e.g., Kugellichter happily bounce up and down. A different, more sophisticated sound is played if the task is solved at the first go. In the case of wrong answers, children are encouraged to try again. Affective encouragement may also positively affect their performances (e.g., Schmitt, Hurwitz, Duel, Linebarger, & Nichols Linebarger, 2018).

Scaffolding Feedback. In addition, scaffolding feedback facilitates children solving a task when they fail to do so. In this regard, scaffolding feedback is defined as hints or information on areas that exceed the children's current knowledge that enable them to solve a task they can not complete without extra help (Wood, Bruner, & Ross, 1976). For example, words are replayed with increasingly emphasized intonation when children fail to identify the stress patterns. Or, in the case of spelling exercises, children may delete distracting letters, i.e., letters not found in the target word, or get individual letters solved automatically.

If children are not able to solve a word within three trials, the solution is displayed. When present, the pedagogical agents give spoken feedback as their empathetic responses may positively impact learning (Plass et al., 2015).

5.3.6 Rewards and Incentives

We designed different rewards for Prosodiya. Children can collect points when answering correctly. They get more points if they solve a task at the first go to avoid trial-and-error behavior. Upon finishing a level, children are rewarded with a summary, see Figure 5.15. Depending on their performances, the level might have been successfully mastered, unlocking subsequent game content. To account for poorer-performing children and to avoid frustration, subsequent content is also unlocked after dynamically adapted number of level repetitions. To provide a high replay value and to increase training effects, we use a 1-3 star rating (i.e., more stars for higher performance) for each level, displayed underneath the level symbol on the world map, see Figure 5.12a on page 60. In the current version, collected points cannot be redeemed and only reflect in-game achievement.

Figure 5.15: Exemplary summary of a level.

5.3.7 Summary

Summary of the game design elements of Prosodiya

Visual aesthetic design

- Prosodiya's graphical appearance intends to be appealing, consistent, and simple. Among others, images from the comic "The Wormworld Saga", are licensed and adapted to deliver high quality graphics.
- OpenDyslexic, a font specifically designed for dyslexics, is used.

Narrative, environment, and game progress

- Prosodiya is embedded in a fantasy-themed setting. The story is about the eponymous world that is haunted by a mysterious fog. The inhabitants seek the children's help to dispel the fog as they themselves are too weak to help their homeland.
- The narrative, environment, and game progress is designed to match Prosodiya's curriculum and difficulty adjustment. The progress is reflected in an in-game world map, cutscenes, and change of background images and atmosphere.

Pedagogical agents

- Kugellichter are the game's pedagogical agents. They take the roles of companions, linguistic tutors and narrators and are responsible to provide (scaffolding) feedback.
- The design of the agents is unisex and children friendly regarding color and shape. Each agent provides a unique feature that links to linguistic characteristics to facilitate learning.

Tutorials and tooltips

- Highly interactive tutorials are used to introduce game mechanics and explain linguistic characteristics and knowledge.
- Tooltips are short explanations appearing at the start of a level that describe main game mechanics and linguistic challenges of the level.

Feedback

- Scaffolding and knowledge of correct response feedback is used.
- Scaffolding feedback helps children to solve tasks that they would not be able to solve without extra help. E.g., the intonation of audio files increases if children fail to identify the stress pattern of a word.

Rewards and incentives

- Children collect points when solving tasks. Children get more points if they solve tasks at the first go.
- At the end of a level, children are rewarded with a summary and – depending on their performances – unlocked game content.
- A 1–3 star rating (more stars = higher score) displayed underneath mastered levels on the map is used to provide high replay value.

Chapter 6

Automatic Generation and Enhancement of Learning Materials

As previously outlined, the development of digital game-based interventions includes the appropriate implementation of pedagogically sound and playful tasks, the embedding of game design elements, and the user-centered interaction design geared towards the needs of the target group and the use of current technology. What I haven't highlighted so far is the creation of appropriate learning content. Content creation includes the selection of appropriate task items (e.g., words, sentences, texts, questions, etc.) and the provision of the material in terms of audio recordings or texts, the latter often with specific visual enhancement. However, the provision of content is costly in terms of effort, time, and money. This means that the content is limited to the budget of a product and the knowledge and design space of the pedagogical experts working on it.

This may, however, result in the lack of appropriate materials for individual children in specific learning domains, which, in turn, makes the implementation of internal differentiation in primary education harder. Importantly, internal differentiation was found to positively affect language performances in primary education, especially when computer-assisted learning environments were used as differentiation tools (cf. Deunk, Smale-Jacobse, de Boer, Doolaard, & Bosker, 2018). As I argue in the next two sections, the automatic generation of language learning material can serve as a differentiation tool and contribute to the provision of individually adapted learning material of children with special learning needs. In Section 6.1, I elaborate on the use of text-to-speech systems in order to automatically generate minimal pairs. In Section 6.2, I present our web-based application that provides automatic and highly customizable enhancement of reading material.

6.1 Automatic Generation of Minimal Pairs with TTS Tools

As explained in Section 5.2.3, Prosodiya uses minimal pairs consisting of a lexical word and a pseudoword counterpart to facilitate the task of vowel length distinction. For example, to distinguish whether the vowel phoneme *i* in *Biene* [bee] is short or long, the program reads out the minimal pair consisting of the lexical word *Bi̲e̲ne* (/ˈbi̲ː̲.nə/) and its pseudoword counterpart *Bi̲nne* (/bɪ̲n.ə/). For this purpose, all minimal pairs were recorded by a professional speech therapist, a costly and time-consuming process. In our article "Optimizing the Quality of Synthetically Generated Pseudowords for the Task of Minimal-Pair Distinction" (Holz, Chinkina, & Vetter, 2018), we investigated the use of text-to-speech tools (TTS) to automatically generate such minimal pairs. This section is mainly based on that article.

In general, so-called minimal pair therapy is used to teach the ability to distinguish speech sounds through the use of pairs of words that differ by a single phoneme, such as *pin* and *bin* (Barlow & Gierut, 2002). As stated in Section 2.1.2, the ability to distinguish long and short vowels is necessary for their correct orthographic markings in the written word and, thus, is crucial for successful literacy development (Klicpera, Gasteiger-Klicpera, & Schabmann, 1993; Landerl, 2003). For example, for the word *rennen* [to run] (/ˈʁɛnən/), the pseudoword counterpart in which the short vowel phoneme /ɛ/ is changed to the long vowel phoneme /e/ is /ˈʁeːnən/, which could be orthographically represented as *rehnen* or *renen*.

For an auditory training of such minimal pairs, trained educators, such as speech and learning therapists, or audio recordings are required. This is, however, a costly step in content provision. In the case of computer-based interventions, this also limits the number of minimal pairs the children can practice with. In order to reduce the effort and increase the size of the provided audio-dictionary, the audio files can be generated automatically with TTS tools. However, the quality of pronunciation of freely available online TTS tools is not always optimal, especially for pseudowords, as we could demonstrate in Holz, Chinkina, and Vetter (2018).

TTS systems are nowadays part of everyday life and have already been encountered in educational contexts. More specifically for this thesis, TTS can be used in language learning systems (e.g., Berkling, Pflaumer, & Lavalley, 2015; Mitkov, 2004) or in the therapy of phonological awareness, which has numerous benefits. For example, Tallal et al. (1996) could show that the use of synthetically slowed down language material can be a predictor for better reading and writing performance. Ptok and Meisen (2008) postulate that the ability to recite minimal pairs also correlates with reading and writing performance. A form of a minimal pair training to learn spelling patterns was also fashioned by Berkling (2017) in the language learning program "Phontasia". Phontasia is a digital learning application that utilizes an Apple TTS system and works on the basis of the phonics method widely used in England. As research on Phontasia yielded positive results, the use of TTS systems for learning of spelling is confirmed as a promising approach.

As for the evaluation of TTS tools, Handley (2009) argues that TTS systems are not sufficiently studied for use in language learning programs. Indeed, a clear research focus of TTS systems to date has been placed on the intelligibility of the speech output, which is not sufficient in the context of computer-assisted language learning. Handley suggests that further investigations with regard to accuracy, naturalness, and expressiveness have to be conducted in order to be able to exploit the full potential of TTS in learning applications. In addition to that, the performance of TTS tools has only been systematically examined with lexical words as input, and not with pseudowords.

With these limitations of the current research in mind, we conducted a study to investigate whether a consistent modification of the input into TTS systems can optimize the speech output for lexical and pseudowords.

6.1.1 Summary of our Findings

To investigate the pronunciation quality of synthetically generated lexical words and their pseudoword counterparts, we generated minimal pairs that covered the categories *vowel length* and *plosives*. For the category *vowel length*, the pseudoword counterpart was generated by changing the length of the vowel phoneme of the stressed syllable from long to short and vice versa (e.g., long to short: *BIE-ne* /ˈbiː.nə/ [bee] → *BIN-ne* /ˈbɪ.nə/ and short to long: *WIN-ter* /ˈvɪn.tɐ/ [winter] → *WIEHN-ter* /ˈviːn.tɐ/). Considering the category *plosives*, the pseudoword counterpart was generated by changing the initial plosive sound of a word from voiced (p,t,k) to voiceless (b,d,g) and vice versa (e.g., voiced to voiceless: *TUR-nen* /ˈtʊʁ.nən/ [to do gymnastics] → *DUR-nen* /ˈdʊʁ.nən/ and voiceless to voiced: *BIE-ne* /ˈbiː.nə/ [bee] → *PIE-ne* /ˈpiː.nə/).

6.1.1.1 Pronunciation Quality of Lexical and Pseudowords Entered as Plaintext

First, we investigated in a quantitative analysis the pronunciation performance of lexical words and their pseudoword counterparts generated by freely accessible TTS tools when using plaintext as input. We included the following TTS tools: GSpeech (Creative-Solutions, 2018), iSpeech (ISpeech, Inc., 2018), MARY TTS (Schröder & Trouvain, 2003), MWS Reader (DirectINNOVATION, UG, 2018), Oddcast (Oddcast, Inc., 2018), Amazon Polly (Amazon Webservices, Inc., 2018), and IBM Watson TTS (IBM, 2018). The results show that all freely accessible TTS tools pronounced lexical words with sufficient to good quality (percentage of correctly pronounced lexical words ranges between [84%, 100%]). However, all tested TTS tools – except for Watson TTS (96%) – pronounce pseudowords entered in plaintext very poorly ([54%, 87%]) and thus can't be used in educational applications. We assume that this is caused by the fact that TTS systems perform lookups

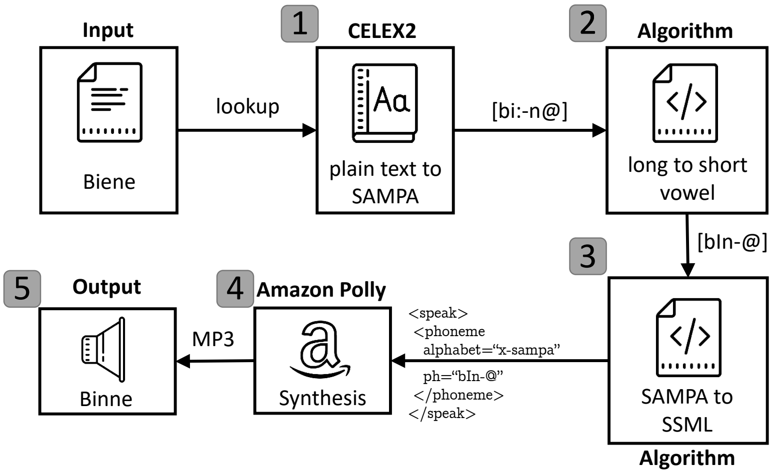


Figure 6.1: Our proposed algorithm to synthetically generated pseudoword counterparts of a lexical word. As an example, for the word *Biene* (['bi:-n@]; bee) with a long vowel *i*, the pseudoword counterpart *Binne* ['bI-n@] with a short vowel *i* is generated.

for lexical words, and if the word is not found (as in the case for pseudowords), the synthesis of pseudowords is based on models trained with lexical words, thus, resulting in outputs pronounced very similar or equal to their lexical counterparts.

The evaluation revealed two drawbacks of using plaintext as input: First, the pronunciation of pseudowords is often unpredictable and insufficient. E.g., some TTS tools pronounced the pseudoword *Wiehnter* /ˈwiːər/ too similar to its lexical counterpart *Winter* /ˈwɪ̠n.tə/ [winter]. Second, using plaintext excludes some German words whose pseudoword counterpart cannot be expressed in plaintext. For example, words in which the vowel is succeeded by the phoneme /x/ as in *Kuchen* /ˈkuːxn̩/ [cake], the pseudoword counterpart /ˈkʊxn̩/ has no plaintext representation.

6.1.1.2 Our Approach to Synthetically Generate Minimal Pairs

Based on the previous findings, we propose a new algorithm to improve the pronunciation of synthetically generated pseudowords. The algorithm uses X-SAMPA[1] instead of plaintext as input and SSML (Speech Synthesis Markup Language) to adjust prosodic features in order to increase the distinction of the minimal pair. We opted for Amazon Polly as the TTS engine as Watson TTS did not offer support for X-SAMPA at the time of the study. The algorithm is shown in Figure 6.1.

We investigated the discrimination of synthetically generated minimal pairs and the naturalness and clarity of their pronunciation in a crowdsourcing experiment with adult participants. We could show that the task of selecting the correct word from a minimal pair of a pseudoword and lexical word was completed equally successfully

[1] X-SAMPA (Extended Speech Assessment Methods Phonetic Alphabet) is a computer-readable mapping of the IPA. X-SAMPA notation is given in square brackets, e.g., the phonetic transcription of *Biene* [bee] in X-SAMPA is [bˈiː-n@].

when a pseudoword was generated by our algorithm or pronounced by a human. Interestingly, participants perceived the words generated by the TTS as significantly more natural and the human-produced words as significantly clearer. The lower rating of naturalness and higher rating of clarity for words produced by humans may result from the fact that these audio files were taken from Prosodiya whose pronunciation focuses on clarity, distinction, and articulation.

We may conclude that by modifying the input of a TTS tool, the pronunciation of synthetically generated pseudowords can reach a quality level comparable to their lexical counterparts. In the future, we aim to replicate the experiment with German (dyslexic) primary school children to shed more light on the suitability of synthesized speech in the language learning context. Upon positive results, the vocabulary of language learning applications, such as Prosodiya, can be extended automatically with less effort.

Summary of automatic generation of minimal pairs with text-to-speech tools

- Distinction of minimal pairs, i.e., pairs of words that only differ by a single phoneme, is commonly used to improve phonological awareness. For example, minimal pairs that only differ in vowel length (e.g., *Biene* /ˈbiː.nə/ [bee] vs. *Binne* /bɪn.ə/) are used in vowel length distinction tasks.
- The content creation of minimal pairs in form of audio files is costly in terms of effort, time, and money.
- Text-to-speech (TTS) tools have shown great promise for the use in language learning but are not sufficiently studied.
- TTS tools have not been examined systematically with pseudowords.

Contributions of this thesis

- Systematic evaluation of freely accessible TTS tools to synthetically generate minimal pairs consisting of a lexical word and its pseudoword counterpart for the use in language learning applications.

Results (Holz, Chinkina, & Vetter, 2018)

- TTS tools pronounce lexical words with sufficient quality when plaintext is used as input.
- TTS tools perform poorly on the pronunciation of pseudowords when plaintext is used as input (with the exception of Watson TTS).
- Using X-SAMPA transcription instead of plaintext and enhancing the input with SSML (Speech Synthesis Markup Language) can greatly improve the pronunciation of synthetically generated pseudowords.
- In our crowdsourcing experiment, the task of selecting the correct word from a minimal pair of a lexical word and its pseudoword counterpart was completed equally successfully when the minimal pair was generated by our algorithm and when pronounced by a human.

6.2 Automatic Visual Input Enhancement of Reading Material

Besides automatically generating audio files for language learning, digital technology can also facilitate the provision of age- and skill-appropriate reading material. For example, texts could be automatically visually enhanced with custom spacing and different coloring of syllables to facilitate reading development (see Figure 6.2). This visual text enhancement can be automatized using natural language processing (NLP) tools and applied on any text. In our article "COAST – Customizable Online Syllable Enhancement in Texts: A flexible framework for automatically enhancing reading materials" (Holz, Weiss, Brehm, & Meurers, 2018), we present a web-based application to easily and automatically enhance syllable structure, word stress, and spacing in reading materials. This section is chiefly based on that article.

As explained in Section 2.2, syllable synthesis and syllable analysis are essential components of evidence-based reading and spelling trainings (Galuschka & Schulte-Körne, 2016). Scheerer-Neumann (1981) has shown that specific training of segmenting words into syllables can improve reading accuracy of reading impaired German primary school children significantly. Additionally, computer-based programs for primary school children that sequentially speak and highlight syllables can facilitate the learning process of reading (Jiménez et al., 2007; Olson & Wise, 1992).

Figure 6.2: Reading text enhanced according to the *Silbenmethode* (Syllable Method; Mildenberger Verlag, 2018, p. 12). Marking phonetic syllables in alternate colors helps children to find the meaning of words faster.

Based on these empirical findings, visually enhanced texts with custom spacing and syllables alternately displayed in different font colors are commonly used in teaching and learning therapy to support the acquisition of reading and writing. This so-called *"Silbenmethode"* (Syllable Method; Mildenberger Verlag, 2018; Röber, 2009) teaches children to focus and understand syllables and their structures rather than single characters and is commonly used in Germany, which is reflected by popular reading materials, such as *ABC der Tiere* (Animal Alphabet; e.g., Handt, Kuhn, & Mrowka-Nienstedt, 2019, see Figure 6.2) and *Leselöwen* (Reading Lions; Loewe Verlag, 2020).

While there is plenty of such reading material available in analog (books) or digital formats (mobile or desktop app, web-apps) that may even include read-aloud function, the published material does not cover all the needs of all children. In fact, a mother contacted me during my studies and asked if I knew of more visually enhanced reading materials for her reading-impaired child. She realized that these reading materials had a great positive impact on her son's literacy skills. However, her child was not interested in reading the stories since he was already eleven years old and the stories target younger children in pre- or primary school. She expressed the urgent need of age-appropriate reading material for her son.

Although there are already tools for automatic syllable enhancement for German, namely *celeco Druckstation* (celeco Printing Station; Klische, 2007)[2] and *ABC Silbengenerator* (ABC Syllable Generator; Müller, 2013),[3] these tools are platform-dependent (only available for Windows) and lack in customization.

In response to this, we developed COAST.[4] COAST is a web application for easy and automatic visual enhancement of syllable structure, word stress, and spacing in texts. A brief comparison of the aforementioned tools and COAST is given in Table 6.1 and a more detailed description in Holz, Weiss, et al. (2018).

6.2.1 The COAST System

The primary focus of COAST is on functionality and practicability. In terms of functionality, COAST offers a high degree of customization for text enhancement, supports management of annotation schemes, and includes syllable stress. We extend the approach of text enhancement that is provided by state of the art tools to make syllable structures and stress more salient for German native (dyslexic) speakers using NLP resources. Enhancing the text with such additional linguistic information may boost children's abilities to segment words into relevant components and may help them to learn to focus on relevant areas of words, for example to learn the orthographic marking of long and short vowels, as explained in Section 2.1.2. To account for practicability, we implement this functionality by collaborating closely

[2] `www.celeco.de/`

[3] `www.abc-der-tiere.de/index.php?id=388`

[4] `www.sfs.uni-tuebingen.de/coast/`

System Feature	Silbengenerator	celeco Druckstation	COAST
Platform Independent	✗	✗	✓
Web-Based	✗	✗	✓
Freely Available	(✓)	✗	✓
Free Text Input	✓	✓	✓
Text Box	✗	✓	✓
Basic Text Layout Customization	✓	✓	✓
Additional Text Layout Customization	✗	(✓)	✓
Customizable Syllable Enhancement	✗	✓	✓
Configuration Templates	✗	n.a.	✓
Stress Annotation	✗	✗	✓
Syllable Arcs	✗	✓	✗
Customizable Analysis	(✓)	(✓)	✓
Crowd-Sourcing	✗	✗	✓
Exercise Generation	✓	✓	✗

Table 6.1: Comparison of *ABC Silbengenerator* (ABC Syllable Generator; Müller, 2013), *celeco Druckstation* (celeco Printing Station; Klische, 2007), and COAST (Holz, Weiss, et al., 2018).

with prospective users, in particular teaching practitioners, to meet real-life demands.

The system overview of COAST is displayed in Figure 6.3 and is briefly described in the following. I refer to Holz, Weiss, et al. (2018) for detailed information on the conception and implementation of COAST. Texts are analyzed in the back-end using NLP tools, database lookups, and manual annotation. For this, we use spaCy (Honnibal & Johnson, 2015) for initial parsing, tokenization, and part-of-speech (PoS) tagging. With this information, COAST queries a database initialized with the CELEX2 language corpus (Baayen, Piepenbrock, & Gulikers, 1995) to infer primary word stress and syllable structure. To account for the sparsity of linguistic resources and wrong database entries, we allow manual annotation of unknown or wrongly annotated words. These manual annotations are stored in user-specific database entries that can be validated in a newly proposed crowdsourcing mechanism and later be inserted into the global database. The front-end offers a high degree of customization in the visual representation of analyzed texts as well as generating, saving, and applying layout templates to a text. The customization offers to modify colors and font of different syllable types as well as making syllable boundaries more salient by using a syllable delimiter character. Finally, COAST allows to customize the text layout independent of syllable enhancement, such as font size, line spacing, and spacing of words, syllables, and letters. The results of user tests indicate that COAST can be used intuitively and time efficiently. An example of COAST's view for text analysis and enhancement is given in Figure 6.4. Currently, COAST supports German and English.

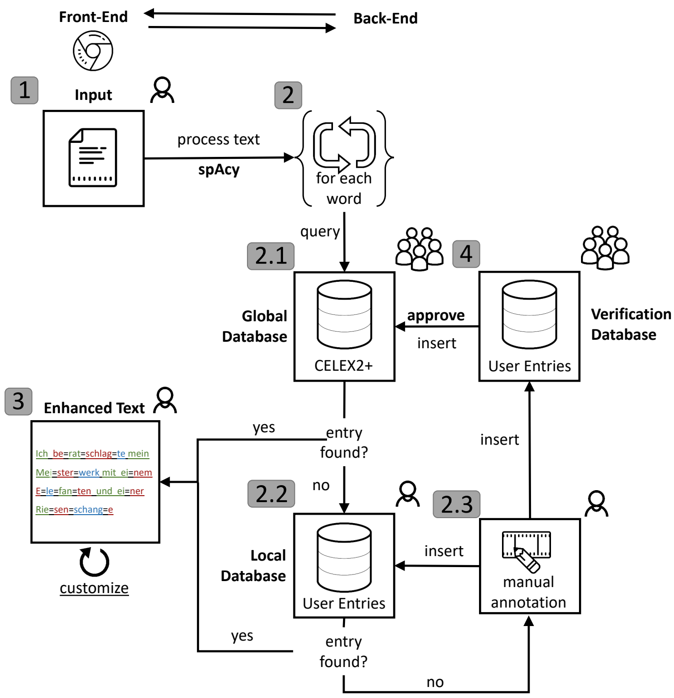


Figure 6.3: System overview of COAST. Texts are processed in the back-end using NLP tools, database lookups, and manual annotation in case of unknown words. The front-end offers rich customization to visually enhance the preprocessed text.

Summary of automatic visual input enhancement of reading material

- Visually enhancing syllables in words, e.g., by alternating the font color of syllables, helps children in the process of literacy acquisition.
- Published material with such syllable enhancement is available in analog and digital formats. However, the published material is limited and does not cover all the needs of young learners.
- Tools that support automatic syllable enhancement in texts are only available for Windows and leave room for improvement in terms of customization.

Contributions of this thesis: COAST (Customizable Online Syllable Enhancement in Texts; Holz, Weiss, et al., 2018)

- COAST is a web app for easy and automatic enhancement of syllable structure, word stress, and spacing in texts using NLP resources.
- COAST features a highly customizable text enhancement and template functionality and introduces a novel crowd-sourcing mechanism to address the data sparsity in language resources.
- COAST's feasibility and usability was demonstrated in user tests.

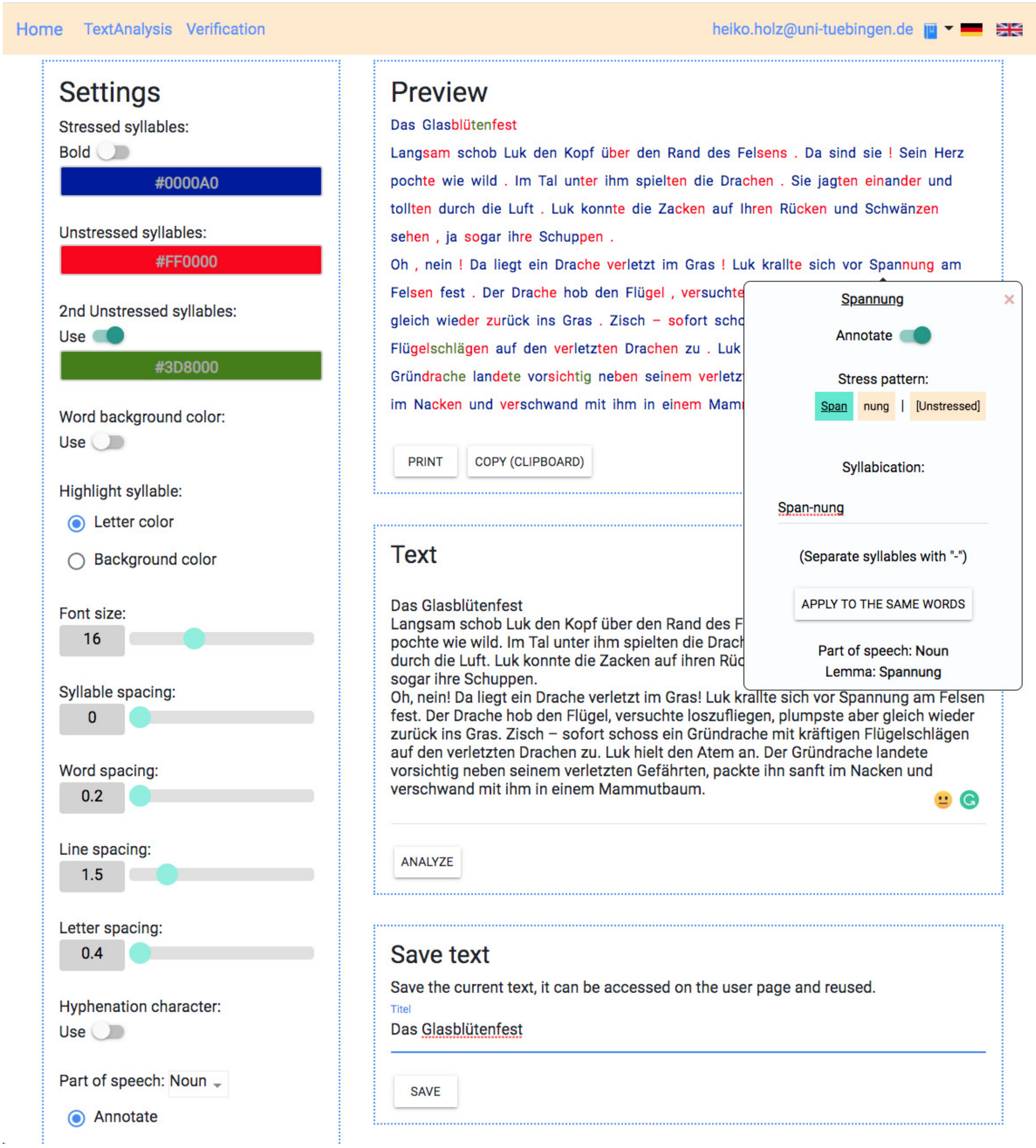


Figure 6.4: View of text analysis and visual input enhancement of COAST. Users can insert or edit text in the lower text box. The preview of text enhancement is given in the upper box. On the left side, users can edit settings regarding syllable annotation. In the preview view, the user clicked on the word *Spannung* [tension], which results in a popup offering custom annotation of the clicked word.

Part III

Major Results and Discussion

Chapter 7

Comparison of Touch Interaction Styles in a Mobile Spelling Game

As shown in the evaluation of Prosodiya (cf. Chapter 8), children reported that the “Spelling” game was their favorite exercise in the training program. Moreover, children reported to perceive a positive influence of Prosodiya on their spelling skills. This highlights the importance of spelling exercises in digital game-based spelling trainings. For the effectiveness study, we implemented a drag-and-drop interaction style to move and arrange the letters to write words in the spelling line.

However, some children reported that this interaction style was physically more demanding than typing on their parent’s smartphones or tablets, particularly if they practiced multiple spelling exercises in one training session. Further, the current research and guidelines on interaction styles with children are puzzling and contradictory (cf. Chapter 3). Thus, without further research, it is not possible to make an evidence-based decision about the appropriateness of different interaction styles to maximize children’s playing experiences.

Therefore, we systematically compare the drag-and-drop, point-and-touch, and touch interaction styles with regard to workload, user experience, and writing times in order to determine the most appropriate solution for touch-based spelling games. This chapter is greatly based on our article “Interaction Styles in Context: Comparing Drag-and-Drop, Point-and-Touch, and Touch in a Mobile Spelling Game” (Holz & Meurers, submitted).

In the following, I first briefly describe the methods and procedure of this study. Then, I report the results on children’s subjectively perceived workload, user experiences, and writing times. I conclude this chapter by summarizing our findings.

7.1 Methods

7.1.1 Participants

Twenty-five German primary school children (14 boys and 11 girls) from third (12 children) and fourth grade (13 children), aged 8–11 years ($M = 9.9$, $SD = 0.57$), participated in the study. Eight of the children were diagnosed with developmental dyslexia (seven boys and one girl).

7.1.2 Spelling Game

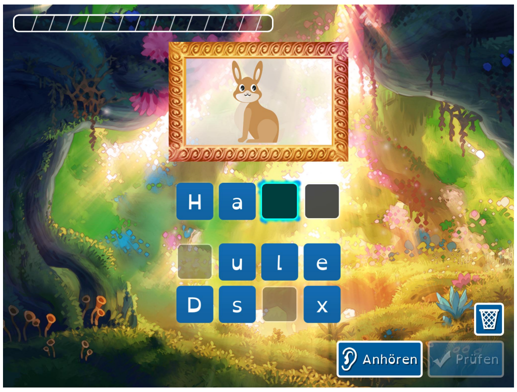
The spelling game of Prosodiya (cf. Section 5.2.5) was used for the study. In order to focus on the interaction style and not on the difficulty of spelling, we decided to set the capitalization automatically, to expose the number of letters of the target word, and to deploy an easy set of distracting letters that may not result in homophonic misspellings. For example, the children were asked to write the word *Hase* [bunny] by arranging the respective letters from the letter area containing $\{H\ u\ l\ e\ D\ s\ a\ x\}$ in the four slots of the spelling line, see Figure 7.1a. We implemented drag-and-drop, point-and-touch, and touch interaction styles to move and arrange the letters. I briefly explain the interaction styles in the following.

Drag-and-Drop. In the drag-and-drop condition, letters are moved and placed by touching the letter object and dragging it to the desired position before it can be placed by releasing the touch, see Figure 7.1b. Drag-and-drop interaction offers the most options: a letter can either be dropped on an empty slot in the spelling line, be inserted between two already placed letters, or replace an already placed letter. As recommended to reduce errors caused by drag-and-drop (cf. Donker & Reitsma, 2007a), we visually indicate where dragged letters can be dropped: for (re-)placement and insertion, the dragged letter and the letter/empty slots in the spelling line wiggle when their responsive zones intersect.

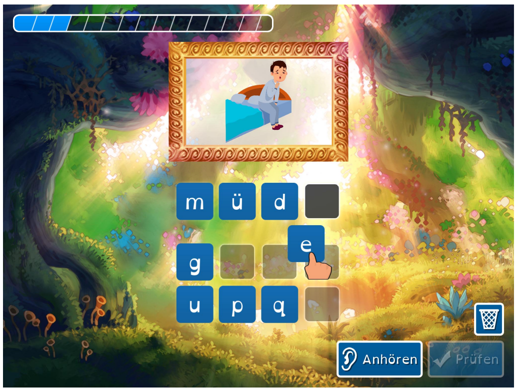
Point-and-Touch. To spell words with the point-and-touch interaction style, children first touch a letter and then touch at the desired position they want to move the letter, see Figure 7.1c. Point-and-touch offers the possibilities to place, delete (touching a letter twice), and swap letters. All interactions can be canceled by touching anywhere outside the spelling line.

Touch. To write words using the touch interaction style, children simply touch on a letter that they want to insert into or delete from the spelling line, see Figure 7.1a. The first free slot of the spelling line has a luminous border and pulses lightly to indicate where the next letter will be placed.

Tutorials. Each interaction style is explained in an interactive tutorial in which children are step-wise introduced to the individual features of the interaction style.



(a) “Touch”. Target word is *Hase* [bunny]. The next touched letter will be inserted at the spelling line’s third position.



(b) “Drag-and-drop”. Target word is *müde* [tired]. The letter “e” is currently being dragged from the letter area to the spelling line.

(c) “Point-and-touch”. Target word is *Banane* [banana]. The letter “a” is currently selected in the letter area. While selected, touching a spelling line position will move the letter there.

Figure 7.1: Interaction styles implemented in the spelling game. Hand symbols are used in animated instructions of the tutorials to explain the interaction mechanics.

To ensure that children understand everything, we used animated audio instead of textual instruction (cf. McKnight & Fitton, 2010). Further, the system waits upon successful execution of requested actions before the tutorial continues, i.e., placement, deletion, swapping, and insertion of letters, and using the buttons.

7.1.3 Measures

To compare the interaction styles, we assessed subjectively perceived workload, user experience, writing times, and direct rankings of the interaction styles.

Figure 7.2: Digital slider for touch devices to collect children's workload ratings for use with adapted versions of the NASA-TLX. The mental demand subscale is displayed. The arrow's position corresponds to values between 0 to 100 in steps of 5.

Workload Questionnaire. To measure perceived workload for each interaction style, we used an adapted version of the highly regarded, multidimensional workload scale NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). For this study, we used the subscales *mental demand*, *physical demand*, *frustration*, and *effort* (leaving out the subscales *temporal demand* and *overall performance*). Laurie-Rose, Frey, Ennis, and Zamary (2014) adapted the NASA-TLX to measure perceived workload in children. For this study, we used and further adapted the kids version of the NASA-TLX to use with touch devices. In the kids version of the NASA-TLX, children are asked for each subscale to indicate their subjective experience on a slider. Drawings are used at the slider's border to reflect the subscales' endpoints (e.g., *very low mental demand*, and *very high mental demand*), see Figure 7.2. The slider's value range from 0 to 100 with increments of 5. For the present study, we developed a digital version of the slider for use with touch devices, designed new drawings in comic style of each subscale endpoint, and added a simplified title of each subscale above the slider. The adapted NASA-TLX for children is listed in Appendix A.4.

Ease of Use, Fun, and Speed. In addition to the workload subscales of the NASA-TLX, we used self-designed questions to assess the subscales *ease of use*, *speed*, and *fun*. For this, we used 5-point Smileyometer (cf. Read, 2008) combined with a 5-point word scale, see Figure 7.3. The answer options for the subscales ease of use, speed, and fun ranged from *very hard* to *very easy*, *very slow* to *very fast*, and *not fun at all* to *very fun*, respectively.

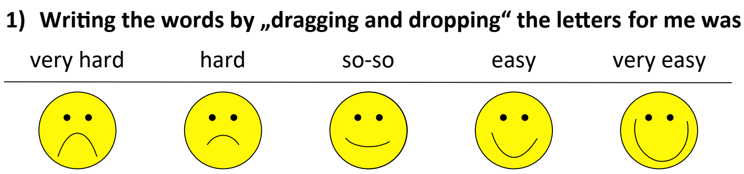


Figure 7.3: Smileyometer for the subscales ease of use (shown), speed, and fun.

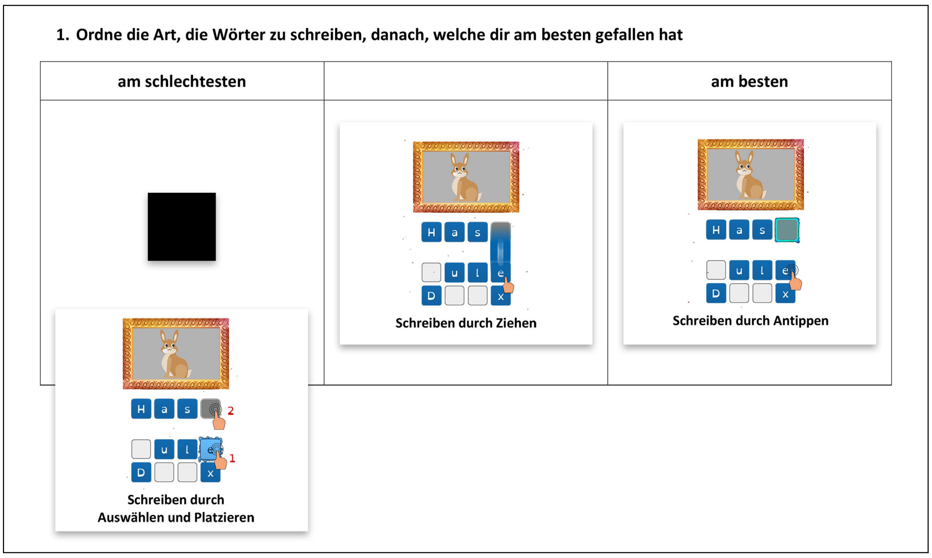


Figure 7.4: Fun-Sorter for the subscales like (shown), ease of use, fun, effort, speed, and mental demand. Interaction styles are represented by magnetic cards which are placed in individual cells. Order of the cells from left to right: worst, mediocre, best.

Fun-Sorter. Children do not necessarily respond honestly to rating scales (Hall, Hume, & Tazzyman, 2016). This may be caused, inter alia, by the desirability bias, i.e., children may not accurately respond regarding socially desirable characteristics in order to appear more appealing to the researcher (Evans et al., 2007), or by the acquiescence bias, i.e., the tendency to agree or respond positively to not tell the researcher that their product is not great (Danner, Aichholzer, & Rammstedt, 2015). This could cause an absent differentiation between the interaction styles although the children subjectively perceive the interaction styles differently. To take this into account, we implemented the Fun-Sorter (cf. Read, 2008; Read, Macfarlane, & Casey, 2002) to explicitly compare the three interaction styles for the subscales *like*, *ease of use*, *fun*, *speed*, *effort*, and *mental demand*. The Fun-Sorter is answered at the end of the experiment and requires the children to rank the three interaction styles in order of preference of the respective subscale. We printed a representative picture of each interaction style on a magnetic card that could be placed in a magnetic cell of the Fun-Sorter table, see Figure 7.4.

7.1.4 Procedure

We used a within-subject design with balanced order of interaction styles and word lists as independent variables. That is, each child wrote words using each interaction style in a pseudo-random order.

Children first wrote 15 words using the first interaction style on a tablet. Upon completion, they answered the subscales of the adapted NASA-TLX on the tablet. Afterwards, the Smileyometer questionnaire was answered on paper. Lastly, we asked children what they specifically liked or disliked about the interaction style. The second and third interaction styles were carried out the same way. In total, investigating one interaction style took about 15 minutes. Short breaks of five minutes were taken between two interaction styles. After finishing with the third interaction style, the children answered the final questionnaire including the subscales of the Fun-Sorter and the question about their favorite interaction style if they can only choose one to write words with. The experiment lasted about fifty to sixty minutes. Children were rewarded with a gift voucher from a local book store.

7.2 Results and Discussion

We performed linear mixed-model regression analyses in order to contrast the influence of interaction style (touch, drag-and-drop, point-and-touch), spelling proficiency (dyslexic [DD], typically developing [TD]), and grade (third, fourth) on the rating of the workload subscales of the adapted NASA-TLX, on the subscales of the Smileyometer questionnaire, and on writing times.

Descriptive results of children's perceived workload and user experiences as well as writing times are summarized in Figure 7.5. The ranking results of the Fun-Sorter and children's interaction style of choice are listed in Figure 7.6. For inferential statistics, I refer to our article "Interaction Styles in Context: Comparing Drag-and-Drop, Point-and-Touch, and Touch in a Mobile Spelling Game" (Holz & Meurers, submitted).

Our results suggest that simple touch interaction is the most appropriate interaction style for children in the proposed spelling game when compared to point-and-touch and drag-and-drop interaction.

7.2.1 Touch as the Preferred Interaction Style

First, we found that children needed significantly less time writing words with touch than with drag-and-drop or point-and-touch. This is confirmed by the subjective ratings of perceived speed, writing times, and free comments.

Looking at the repeated subjective ratings of workload and user experience subscales, touch was reported to be less physically demanding, less effortful, easier to use, more fun, and – compared to point-and-touch mainly – less frustrating and less mentally demanding. While typically developing children also reported touch as being less mentally demanding and less frustrating than drag-and-drop, this was not observable as such for dyslexic children. This is possibly due to the low sample size of only eight dyslexic children or due to the fact that they primarily struggle with correctly spelling the words (e.g., they need more corrections to spell the words), which may attenuate a distinctive subjective perception of mental workload that can be attributed to the interaction style.

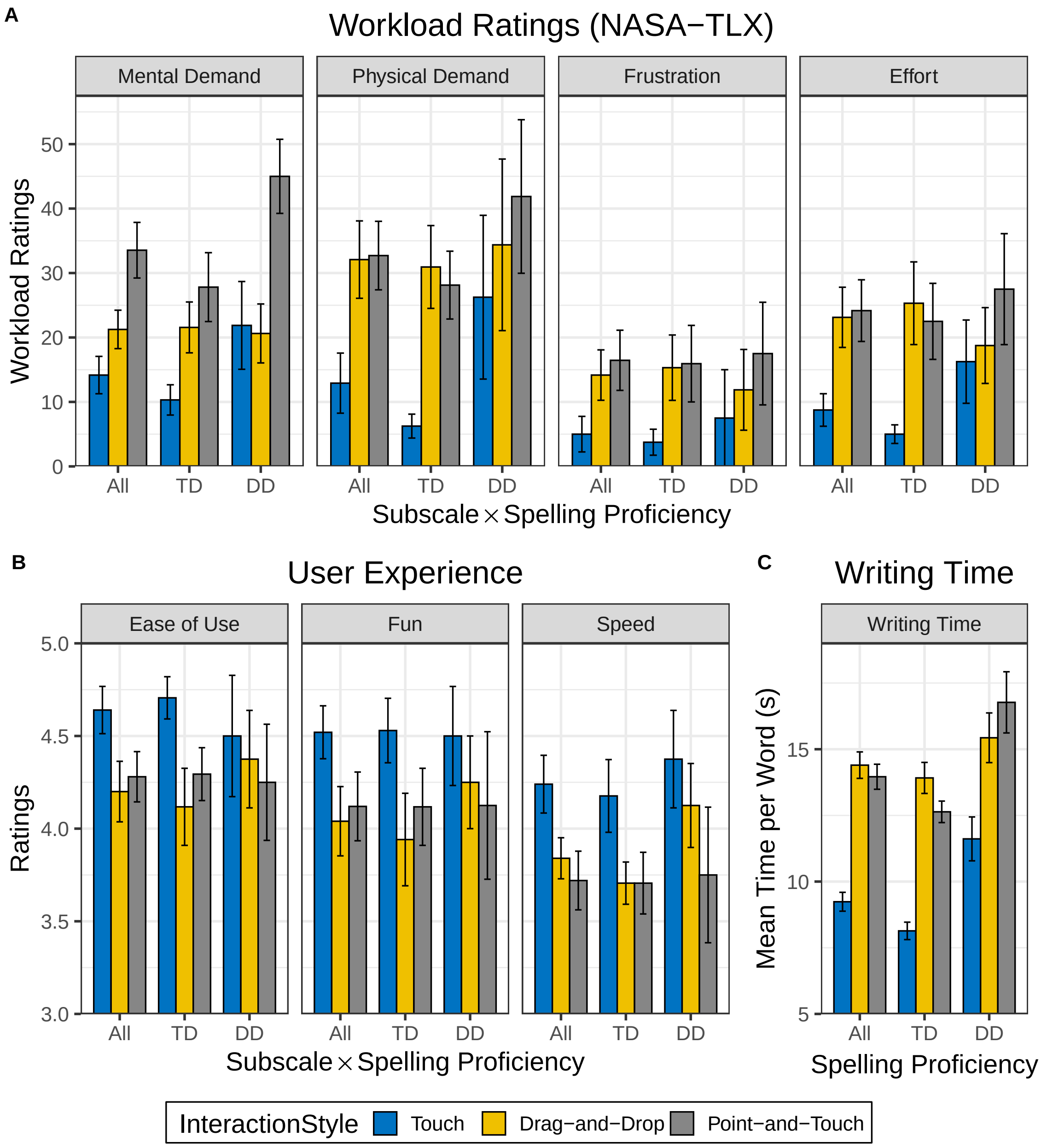


Figure 7.5: Children's mean ratings and writing time by subscale (facets), interaction style (color), and spelling proficiency (all = all children; TD = typically developing; DD = dyslexic). Bars represent the standard error of the mean. **A**: Children's workload ratings of the adapted NASA-TLX; **B**: Children's user experience ratings; **C**: Children's average writing times per word.

But the appropriateness of touch becomes systematically visible when looking at the results of direct rankings: touch was ranked better than drag-and-drop and point-and-touch by typically developing and dyslexic children in all subscales, i.e., like, ease of use, fun, effort, speed, and mental demand. Finally, touch was selected most often by 17 out of 25 children as their interaction style of choice.

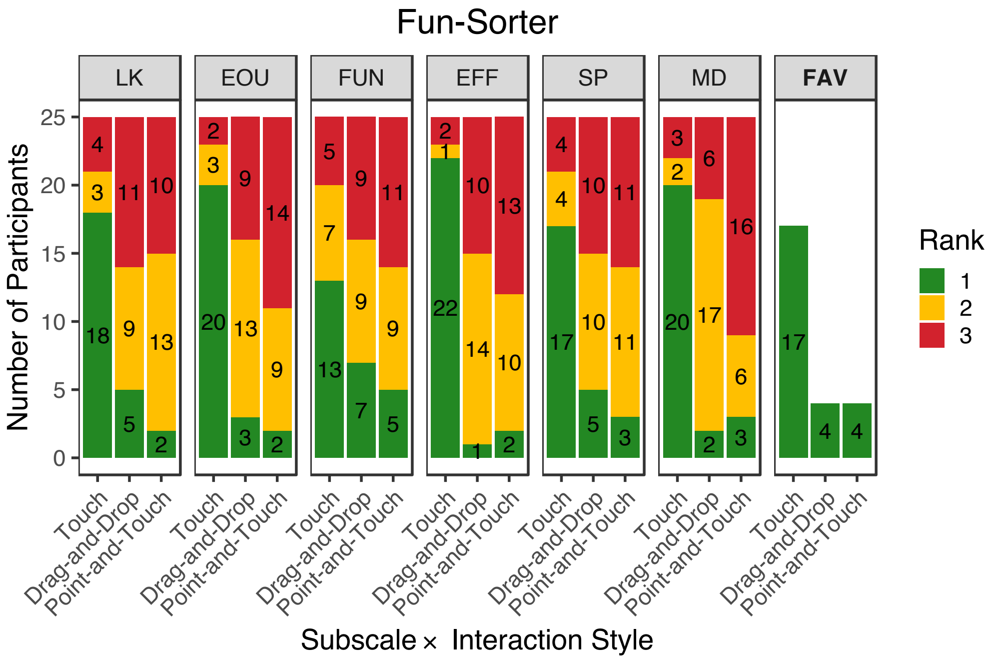


Figure 7.6: Children's Fun-Sorter rankings of the interaction styles by subscale (facets). LK = like; EOU = ease of use; FUN = fun; EFF = effort; SP = speed; MD = mental demand; FAV = favorite.

The findings indicating that touch should be favored over drag-and-drop and point-and-touch in a mobile spelling exercise may be caused by the fact that touching requires less interactions and hand/finger movements if words are written without correction, that children are getting increasingly familiar with mobile touch devices, or that click or simple touch is used in (on-screen) keyboards. This is supported by comments stating that touch felt like writing on a computer keyboard or on children's (or their parent's) mobile devices.

7.2.2 Point-and-Touch vs. Drag-and-Drop

Considering point-and-touch and drag-and-drop, we did not observe any difficulties in children using these interaction styles. Furthermore, children seemed to be already experienced in dragging and dropping, concurring recent findings (Barendregt, 2015; Hamza & Salivia, 2015). We observed that children successfully used all features of point-and-touch and drag-and-drop, i.e., deletion, swap, and insertion of letters. Two children exclaimed that they specifically liked that drag-and-drop – compared to the other interaction styles – offers the possibility of directly inserting letters between two already written letters. This was confirmed when we asked the children how we could improve the spelling game: some of them responded that they really liked swapping and inserting letters and suggested to implement these features also for the touch interaction style.

Interestingly, our findings contradict earlier results and design recommendations concluding that point-and-click or point-and-touch is more appropriate than drag-and-drop (Chiasson & Gutwin, 2005; Gelderblom & Kotzé, 2009; Girard & Johnson, 2009; Inkpen, 2001; Joiner et al., 1998; Roman, 2015; Soni et al., 2019; Ward, 2014). We found small advantages of drag-and-drop over point-and-touch in reported mental demand, i.e., that drag-and-drop was reported as less mentally demanding than point-and-touch; subjective ratings of other workload and user experience subscales and writing times did not differ. The difference in subjectively perceived mental demand may result from drag-and-drop interaction making it possible to free cognitive resources through offloading (Antle, 2013), i.e., by dragging letters in a playful way before dropping them. Although drag-and-drop was only rated to be more physically demanding and taking more effort than touch, but not than point-and-touch, a few children commented that drag-and-drop is generally "more exhausting for the fingers and the arm". Further advantages become more visible in ranking responses when children are forced to directly compare the two interaction styles, providing a more fine-grained picture. The results of the ranking responses showed that drag-and-drop was ranked better than point-and-touch regarding ease of use, fun, effort, speed, and mental demand, see Figure 7.6. The subscale like yields no clear difference between drag-and-drop and point-and-touch and seemed to polarize the children. While drag-and-drop was liked best more often, it was also liked least more often, and point-and-touch was mostly rated as the second best interaction style. Thus, our results concur more with reports on drag-and-drop to be unproblematic or to be even more appropriate than point-and-click (Barendregt, 2015; Barendregt & Bekker, 2011; Donker & Reitsma, 2007a), particularly for touch-based devices (Hamza & Salivia, 2015). We assume contradictory results arise from the type of task children performed, the input modality (mouse vs. touch), children's ages, and the previous experience of children interacting with technology at the time the studies were conducted. In the current era of mobile touch devices, swiping and dragging becomes more and more part of children's everyday lives. Children may expect the same interaction styles they have been using in other applications. Also, whether the performed actions have a more natural mapping to either of the interaction styles is an important factor to consider.

The implications of the comparison of drag-and-drop and point-and-touch are twofold. First, it seems that dyslexic and typically developing children perceived and handled these interaction styles generally somewhat alike, with a slight tendency towards the advantage of drag-and-drop. This is also reflected in that drag-and-drop and point-and-touch were selected by four children each as their interaction style of choice. Second, it highlights the importance of direct rankings in user studies to get deeper insights.

7.2.3 Relevance of Implementation Specifics

Respective point-and-touch, we observed five children trying to swap letters between the letter area and the spelling line by first selecting the letter in the spelling line and then touching a letter in the letter area. The study version only supported to swap letters the other way round. We suggest to implement swap both ways when opting for point-and-touch.

In pilot tests, we observed that children using the drag-and-drop interaction styles tried to drag letters very accurately to destined locations before dropping them, taking more time and effort than necessary. We thus enhanced the tutorial for drag-and-drop by clarifying the sufficiency to drag the letter just close to the destined location. We assume this additional explanation prevented ratings and writing times that would erroneously have led to the disadvantage of drag-and-drop. Thus, besides sufficiently big responsive zones and allowance of out-of-bound touches, it is of utmost importance to inform the children *precisely* how the interaction style works in order to make inferences on performance and other metrics.

7.2.4 Impact of Child Differences

Impact of Spelling Proficiency. Children's spelling proficiencies were expectantly reflected in writing times and reported mental demand: dyslexic children needed significantly more time to write words and tended to report higher mental demand than typically developing children. Apart from these main effects, we found no significant interaction effects between spelling proficiency and interaction style on the subjective workload and user experience ratings. Thus, the indications on the different interaction styles drawn above apply to typically developing and dyslexic children alike.

Impact of Grade. The influence of children's physical and mental development was reflected in that third graders tended to report higher physical demand and needed significantly more time to write words than fourth graders, concurring with recent studies (e.g., Hamza & Salivia, 2015; Vatavu et al., 2015). Apart from these main effects, we found no significant interaction effects between grade and interaction style on the subjective workload and user experience ratings.

7.3 Conclusion and Outlook

The current state of research is puzzling when it comes to determining the most appropriate (touch) interaction style for children (Barendregt, 2015; Donker & Reitsma, 2007a; Hourcade, 2008). To determine the most appropriate interaction style in a tabled-based spelling game, and to disentangle currently reported contradictions, we compared drag-and-drop, point-and-touch, and touch in a lab experiment.

We asked twenty-five German children aged 8–11 years, eight of whom were dyslexic, to arrange letters in a spelling line to write single words.

We were able to demonstrate that children aged 8–11 years can use drag-and-drop, point-and-touch, and touch without problems in the proposed spelling game. Furthermore, we observed that children are aware of and use unique features that constitute each interaction style, e.g., swapping letters using point-and-touch or inserting letters using drag-and-drop.

Our result suggest that touch is the least mentally and physically demanding, the least effortful and frustrating, the easiest to use, the most fun, and the fastest interaction style among the three. Additionally, touch was favored to drag-and-drop and point-and-touch in direct rankings with regard to liking, ease of use, fun, effort, speed, and mental demand. Finally, touch was selected as the interaction style of choice by 17 out of 25 children, whereas four children each chose drag-and-drop or point-and-touch as their favorite. Based on our results, touch seems to be the most appropriate interaction style in an educational touch-based spelling game – independent of spelling proficiency and grade.

Possibly, a hybrid interaction style combining features of various interaction styles would be even more appropriate. This was also proposed by the children who reported to like the possibility to swap and insert letters and wanted to have the same features in the touch interaction style. For example, the interaction style starts in the touch mode and long pressing or movements of a fingertip length (cf. McKnight & Fitton, 2010) switches into drag-and-drop or point-and-touch, enabling swapping of letters or inserting between already written letters in the spelling line without prior deletion.

Regarding the different groups of spelling proficiency, i.e., dyslexic and typically developing children, we found no significant interaction effects in the analyses of the three interaction styles. But dyslexic children needed systematically more time, they tended to report higher mental demand when writing the words, and their ratings did not differ as strongly between touch and drag-and-drop on some subscales. Thus, the conclusions drawn above for the different interaction styles in general hold for both, typically developing and dyslexic third and fourth graders.

Further research is still required to compare touch-based interaction styles in contexts that specifically address not only performance metrics, but also take into account workload, user experience, and children's habits and preferences. This is specifically the case with the emerging use of educational touch-based applications. We advice researchers and designers to select the interaction style carefully with regard to age-group, input modality, and context – and to not rely on results that do not address the requirements of their application. Furthermore, future studies on the design of touch-based spelling exercises could also investigate implications on learning. While one interaction style is found to be more appropriate according to workload and user experience, learning outcomes may tell a different story.

As for Prosodiya, the results suggest that we should change from the currently implemented drag-and-drop interaction style to touch or a hybrid interaction style as described above. Changing the interaction style could further improve children's positive training experiences reported in the next chapter.

7.4 Summary

Summary of the comparison of touch interaction styles in a spelling game

In this chapter, I assessed which touch interaction style is the most appropriate in a tablet-based spelling game for primary school children to interact with letter objects to write words. That is, we systematically compared the interaction styles drag-and-drop, point-and-touch, and simple touch in a within-subject experiment with 25 German primary school children (aged 8–11 years) with regard to subjectively perceived workload, user experience, and writing time. In the experiment, children wrote words using each of the interaction styles in a pseudo-random order. The results are as follows:

Touch is most appropriate in a game-based spelling exercise

- Touch was reported to be the least mentally and physically demanding, the least effortful and frustrating, the easiest to use, the most fun, and the fastest interaction style.
- Touch was favored to drag-and-drop and point-and-touch in direct rankings with regard to liking, ease of use, fun, effort, speed, and mental demand.
- Touch was selected as the favorite interaction style of choice by 17 out of 25 children.

Drag-and-drop scored slightly better than point-and-touch

- Drag-and-drop was reported to be less mentally demanding than point-and-touch, while the children's ratings of the other workload and user experience scales did not differ.
- Drag-and-drop was preferred over point-and-touch in direct rankings regarding ease of use, fun, effort, speed, and mental demand.

→ Our results suggest that touch is the most appropriate interaction style in a touch-based spelling game.

Chapter 8

Evaluation of Prosodiya

The design and development of our novel approach to improve children's literacy skills provide contributions to extend currently available digital spelling trainings for German dyslexic primary school children for the use at home. The chapters on the scientific background and on the user-centered development of Prosodiya aimed at convincing the reader that our approach has the potential to effectively and enjoyably improve children's literacy skills.

In this chapter, I evaluate the feasibility, effectiveness, and validity of Prosodiya – the main contribution of this thesis. I report the results of a randomized controlled field trial conducted in 2018 with 116 German primary school children. This chapter is mainly based on our articles "Design Rationales of a Mobile Game-Based Intervention for German Dyslexic Children" (Holz, Beuttler, & Ninaus, 2018), "Validity and Player Experience of a Mobile Game for German Dyslexic Children" (Holz, Ninaus, et al., 2018), and "A Digital Game-Based Training Improves Spelling in German Primary School Children – A Randomized Controlled Field Trial" (Holz et al., unpublished).

This chapter is structured as follows: In Section 8.1, I first describe the design of the study and the recruitment and demographics of participants. In Section 8.2, I evaluate the feasibility with regard to training behavior, game experience, usability, and children's perceived influence of the training on reading and spelling skills. In Section 8.3, I evaluate the training effect of Prosodiya on syllable stress awareness and reading and spelling skills. In Section 8.4, I evaluate the validity of Prosodiya by investigating the relationship between syllable stress awareness and reading and spelling skills as well as the relationship between these literacy skills and training performance metrics computed from game logs. In Section 8.5, I conclude the evaluation of Prosodiya by summarizing our findings.

8.1 Methods

8.1.1 Design

A two-period, wait-list controlled crossover treatment design was used to evaluate Prosodiya, with participants randomized to the immediate treatment group (ITG) or to the delayed treatment group (DTG). Pretests were conducted in February 2018 (T1) after which participants from the immediate treatment group trained 9–10 weeks at home with Prosodiya. Midtests were conducted in May 2018 (T2) after which the training of the immediate treatment group was discontinued and participants from the delayed treatment group were crossed to the active training and practiced 9–10 weeks at home. Posttests were conducted in July 2018 (T3).

Test sessions were administered in classrooms of participating schools and learning institutions or in rooms of the university. A test session was as follows: First, classroom tests of spelling and reading fluency were administered in groups, followed by individually administered assessments of syllable stress awareness and word reading. At T2 and T3, training experience questionnaires were answered after the spelling tests by children from the active training group. At the end of a test session, children were rewarded with toy dinosaurs, flexible pencils, erasers, or stamps.

8.1.2 Participants

We recruited primary school children from second to fourth grade at the age of 7–11 years via learning institutions, the youth welfare office, newspaper advertisement, and eight public primary schools in the area of Tübingen, Baden-Wuerttemberg, Germany. Flyer were sent to the institutions and we asked learning therapists, teachers, and employees of the youth welfare office to pass these to parents of poor spellers. In total, 137 families responded to the flyer of which eight dropped out before the study had started. Of the remaining 129 participants, we excluded thirteen children in the efficacy analyses, eleven children who received concurrent reading or spelling remediation and two children due to technical issues during training.

The final sample for efficacy analyses is listed in Table 8.1 and includes 116 children (65 boys), aged between 7–11 years (M = 8.85, SD = 0.93). Of the eligible 116 children, 58 children were assigned to the immediate treatment group and 58 to the delayed treatment group. The assignment was mainly done randomly based on spelling and reading skills assessed at T1. A full randomization of the participants was not possible due to ethical reasons and real-life circumstances of a field trial. Twelve parents of dyslexic children were not willing to participate in the study if their child would be assigned to the delayed treatment group and thus were assigned to the immediate treatment group. Three children, whose parents contacted us just before the start of the first training period, and four children who were sick at T1 were allocated to the delayed treatment group. Nine children of the immediate treatment group did not participate at T3 due to sickness or because they continued with a

Table 8.1: Descriptive data of the treatment groups (ITG = immediate treatment; DTG = delayed treatment).

Variables	ITG ($N = 58$)		DTG ($N = 58$)			
	M (SD)	*Range*	M (SD)	*Range*	t	p
Age in years	8.9 (0.9)	7.5–10.6	8.8 (1.0)	7.3–11.0	0.75	.45
Spelling[a]	37.7 (8.3)	23.5–57.9	41.6 (8.1)	25.8–61.8	−2.24	.03*
Reading fluency[b]	81.9 (14.0)	62–113	86.5 (13.8)	64–127	−1.22	.23
Word reading[c]	18.7 (22.7)	1–82	22.0 (23.7)	2–90	−0.69	.49
Syll. stress awareness[d]	7.4 (3.1)	1–13	7.9 (3.2)	1–14	−0.52	.60
	Frequencies				χ^2	p
Boys/girls	35/23		30/28		0.56	.45
Diagnosed dyslexics	25/33		8/50		10.84	< .001*
Grade 2/grade 3/grade 4	23/24/11		27/20/11		0.68	.71

[a] Spelling (DRT): T-scores, $M = 50$, $SD = 10$.
[b] Reading fluency (SLS 2–9): LQ-scores, $M = 100$, $SD = 15$.
[c] Word reading (SLRT-II): percentile ranks.
[d] Syllable stress awareness (self-designed test): raw scores (max = 15).
* Significant difference between the two treatment groups.

Table 8.2: Psychometric conversion table between standardized scores.

Percentile ranks	T-scores	LQ-scores	Performance level
91 − 100	≥ 64	≥ 120	good to excellent
76 − 90	57 − 63	110 − 119	above average
25 − 75	44 − 56	90 − 109	average
11 − 25	38 − 43	80 − 89	below average
0 − 10	≤ 37	≤ 79	very poor to poor

spelling remediation after T2. The flow diagram of the present study is depicted in Figure 8.1. As listed in Table 8.1, mean reading and spelling skills of the participants were significantly below average and ranged between very poor and below average, with few exceptions of (above) average performance, cf. Table 8.2.

For efficacy analyses, we only included children from the active training group who completed at least two-thirds of the training program (immediate treatment group during the first and delayed treatment group during the second training period) or who served as the control (vice versa). The first two-thirds of the training cover the acquisition phase. Children acquire new skills and learn to use their new knowledge. The last third covers a training and automation phase. Analyses including only participants who completed the whole intervention yield the same test decisions, with the drawback of smaller sample sizes and less ecological validity. As for training experience, we evaluate the data of 99 children who completed at least two-thirds of the training and who answered the training experience questionnaire (four children of the delayed treatment group did not answer the questionnaire).

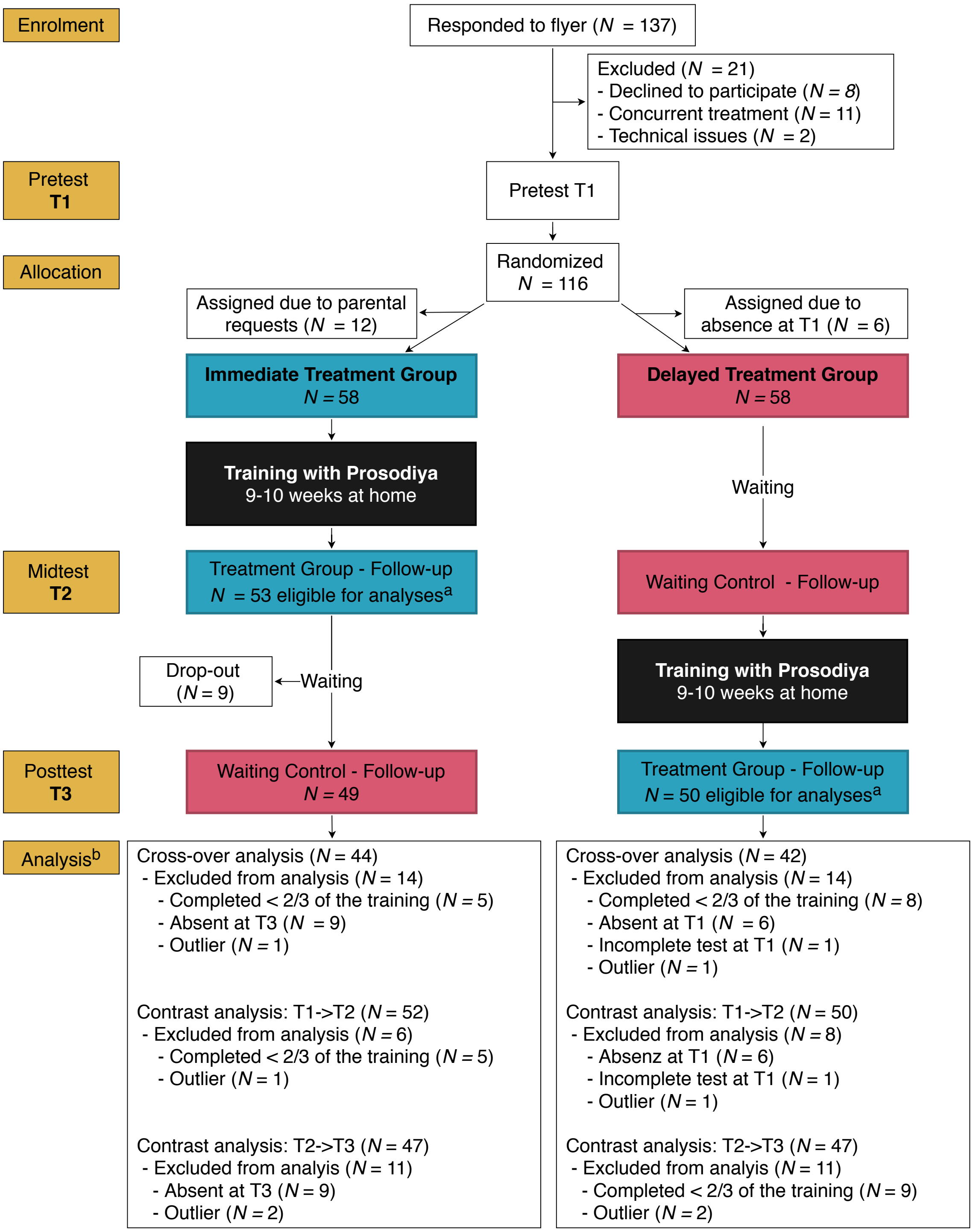
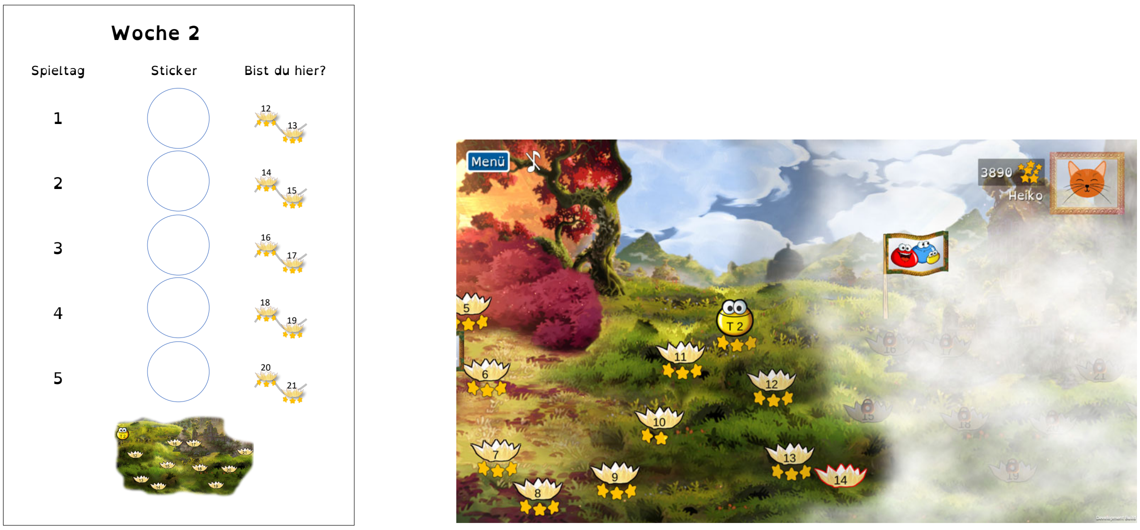


Figure 8.1: Flow diagram of the two-period, wait-list controlled crossover treatment design for the randomized controlled field trial of the Prosodiya training program. *Note.* [a] Only children from the active training group who completed at least two-thirds of the training program were included in the analyses of respective training periods.
[b] Analysis of the standardized spelling scores assessed with the DRT.



(a) One week of the training plan in the sticker book.

(b) Corresponding in-game map of the training. Glass blossoms are used as level symbols.

Figure 8.2: Training plan depicting what should be trained when to keep children on track and to engage them to complete their training.

8.1.3 Materials

Game. The mobile game described in Section 5.2 was used. For the present study, we excluded the subchapter on the "silent h" since words that feature a silent h are exceptions in terms of orthographic marking that do not follow explicit rules and must be memorized and learned by heart with memos such as "*das stumme h, das ist nicht schwer, steht meist vor l, m, n, und r*" [the silent h precedes mainly but not necessarily the letters l, m, n, and r after a long vowel phoneme]. Due to the brevity of the present study (training period: 8–10 weeks), we focused on the more consistent orthographic marking of long and short vowels. Further, the study version did not include capitalization rules. In spelling games, the available letters were displayed in lower- and uppercase, depending whether a noun was practiced or not, and the case could not be changed. For example, the available letters to spell the word *rennen* [to run] were all lowercase, whereas the available letters for the word *Biene* [bee] contained both lowercase and uppercase letters, e.g., a possible set of letters, including distractors, would be $\{B,n,i,n,e,P,h,ä\}$. Further, the study version did not include the games "Syllable counting" (cf. Figure 5.4 on page 47) and the simplified version of "Vowel length distinction" (cf. Figure 5.5b on page 49). The story implemented in the study version was limited to the prologue.

Training Plan. During respective training periods, families were given Android tablets and the children were asked to train at home five days per week twenty minutes each, following a training plan of eight weeks, see Figure 8.2. The training plan was given in the form of a sticker book with a set of 40 stickers to keep the children on track and to engage them to complete their training. The sticker book depicts for each training day and week the levels to be practiced, see Figure 8.2a. Each page contains one training week and corresponds to the map used in the game, see Figure 8.2b.

Due to school holidays during training, more levels than included in the sticker book were deployed in the game. In total, 80 levels were deployed. The training was officially completed at level 66, labeling the remaining levels as bonus. In each level, ten words were practiced. Depending on the levels' configuration and children's performance, the same levels may have to be practiced more than once. To avoid binge-playing and loss of training effect, content of a new training week was unlocked on Monday mornings.

8.2 Feasibility and Training Experience

The game's feasibility and training experience for the immediate treatment group is reported in "Validity and Player Experience of a Mobile Game for German Dyslexic Children" (Holz, Ninaus, et al., 2018), and the evaluation of individual game elements in "Design Rationales of a Mobile Game-Based Intervention for German Dyslexic Children" (Holz, Beuttler, & Ninaus, 2018). As the training behavior and game experience of the two treatment groups did not differ significantly, the drawn conclusions still hold.

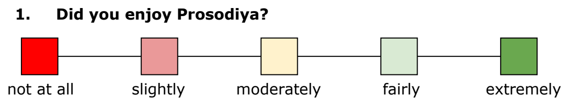
In this section, I report the training behavior of all 116 children eligible for analyses. Further, I report the game experience, perceived self-efficacy, and rating of individual game elements of those 99 children who answered the training experience questionnaire and who completed at least two-thirds of the training.

Investigating the feasibility and training experience of a digital game-based training program is necessary in order to interpret the results found in effectiveness analyses. If the training is not feasible, i.e., if it cannot be used by the target group unassisted and integrated into everyday life, then the conclusions drawn from efficacy analyses might be limited. That is, they might only represent training effects under optimal, controlled conditions and may not transfer to real-life contexts. Furthermore, a positive attitude towards the training and children's perceived impact on their abilities play an important role in learning. If children perceive the training negatively and think it's useless, it is unlikely that they willingly practice over a longer time and the training effect may be attenuated.

8.2.1 Methods

8.2.1.1 Measures

We assessed training behavior, game experience, user experience, usability, and perceived self-efficacy for trained literacy skills.



(a) Color and word coded rating scale.

(b) Smileyometer.

Figure 8.3: 5-point Smileyometer and 5-point color and word coded rating scale used in the questionnaires.

Training Behavior. Training behavior was computed from learner data extracted from game logs.

Training Experience. Training experience was evaluated based on a subset of a questionnaire of 69 questions. These 69 questions covered the following categories: general *game experience*, *user experience*, *self-efficacy* and *perceived influence on the awareness of specific linguist characteristics*, and questions on *individual game elements*. The complete questionnaire is listed in Appendix A.3.

Game Experience. First, we evaluated 19 questions from the Game Experience Questionnaire (GEQ; IJsselsteijn, de Kort, & Poels, 2013) using a 5-point word and color coded rating scale, see Figure 8.3a. We refer to the 19 questions from the GEQ as the $iGEQ_+$ that was composed of the in-game GEQ (iGEQ) and 6 additional questions from the GEQ's core module. The GEQ is intended to measure the seven subscales (i) *positive affect*, (ii) *competence*, (iii) *sensory & imaginative immersion*, (iv) *challenge*, (v) *flow*, (vi) *negative affect*, and (vii) *tension/annoyance*.

The $iGEQ_+$ adds one additional item to all subscales except flow. The items and Cronbach's alpha for the subscales of the iGEQ and $iGEQ_+$ are shown in Table 8.3. Due to an increase in Cronbach's alpha, we kept the additional item for all subscales except *negative affect*, which resulted in a severe decrease of Cronbach's alpha.

User Experience. In addition to the $iGEQ_+$, we evaluated 12 self-designed questions that covered the following six categories: (i) the children's *overall impression* of the game (1 question), (ii) *usability* (4 questions), (iii) *self-efficacy* (5 questions), (iv) *intention to use* (2 questions), (v) *likelihood to recommend* (1 question), and (vi) whether Prosodiya feels more like homework or like a game (1 question). Likelihood to recommend is inspired by the net promoter score by Reichheld (2004), which is one of the simplest loyalty measures. We used either a 5-point Smileyometer (Figure 8.3b, cf. Read, 2008), a bipolar rating scale, or the same color and word coded scale that was used for the $iGEQ_+$.

Table 8.3: Cronbach's alpha of the items used from the in-game GEQ (α_{iGEQ}) and the iGEQ+($\alpha_{\text{iGEQ+}}$) for the present study. For comparison, α^* refers to the reference Cronbach's alpha of the subscales of the GEQ's core module (cf. Poels et al., 2007).

	items iGEQ[a]	item+[a]	α_{iGEQ}	$\alpha_{\text{iGEQ+}}$	α^*
Positive Affect	[1,14]	4	.76	.85	.80
Competence	[17,2]	15	.62	.66	.83
Immersion	[3,27]	12	.75	.82	.81
Challenge	[26,33]	11	.70	.74	.74
Flow	[13,5]		.65		.87
Negative Affect	[16,9]	7	.50	.25	.71
Tension/Annoyance	[29,24]	22	.65	.79	.82

[a] The item number refers to an item's number in the core module of the GEQ (cf. IJsselsteijn et al., 2013).

Self-Efficacy and Perceived Impact on Specific Linguistic Awareness. As mentioned above, the category of general reading and spelling *self-efficacy* consisted of five questions. Besides this, the questionnaire also included five questions on the perceived influence of the training on children's awareness of specific linguistic characteristics that were explicitly trained in the program. That is, we asked children if they know, thanks to the training, what (i) stressed, (ii) open, and (iii) closed syllables are and how to spell (iv) open and (v) closed syllables.

Individual Game Elements. In addition to general training experience, we also assessed the perception of individual game elements, grouped into the following six categories: (i) *Kugellichter/pedagogical agents* (4 questions), (ii) *narrative, map, and environment* (6 questions), (iii) *background environment* within levels (4 questions), (iv) *tutorials* (4 questions) and *tooltips* (3 questions), (v) *enjoyment of individual games* (1 question per game), and (vi) the children's *favorite game* (1 question).

8.2.1.2 Procedure

The training experience questionnaires were conducted as follows: the children were told that they now have the chance to express anonymously what they think about the game with no right or wrong answers. We explained the rating scales of the questionnaire and provided explicit examples for positive and negative items with mock-up questions (e.g, "I like chocolate" vs. "I hate gummy bears"). Moreover, preliminary tests indicated that children had problems reading and understanding the questions due to their lack of proficient reading skills and unfamiliarity with such questionnaires. Thus, we read aloud each question individually and clarified posed questions to ensure that everyone understood the items. We continued with subsequent questions after every child had answered the preceding one. It took approximately 20 minutes to answer the questionnaire. Answering the questionnaire took approximately 20 minutes.

8.2.2 Results and Discussion

In the following section, I first report the results of children's training behaviors, followed by the reports on training experiences.

8.2.2.1 Training Behavior

Children practiced for about 18.5 minutes ($SD = 7.3$) over 27.9 days ($SD = 10.9$), reached, on average, level 69.0 ($SD = 18.7$), and practiced, on average, a total of 161.7 levels ($SD = 48.8$). Children spent an average of 10.1 hours ($SD = 3.5$) in total practicing with the game. It took them an average of 3.0 minutes ($SD = 0.8$) to complete a level that featured 10 words and they scored an average of 138.6 ($SD = 5.7$) out of 150 possible points per level, solving, on average, 8.2 ($SD = 0.9$) tasks on the first attempt. The training behavior did not differ significantly between the immediate treatment group and the delayed treatment group.

Out of the 116 children eligible for the evaluation, 103 children (89%) completed at least two-thirds of the training and 88 children (76%) fulfilled the complete training plan, reaching level 66 or higher. The number of children who successfully completed the training is comparable to that obtained in controlled intervention studies in which the training is carried out supervised in controlled settings, such as schools or learning facilities.

8.2.2.2 Training Experience

Answers from the training experience questionnaire were transformed into values 1 to 5. Answers with no clearly selected options were excluded. If children put marks between two options, we kept and transformed the answer into a floating point value. We excluded for each subscale the ratings of children who failed to answer more than one question.

Mean values of subscales were considered to reflect training experience. For the GEQ and general training experience, we used a conservative approach of analyzing each subscale by conducting one sample Wilcoxon signed rank tests against the middle value of the subscale's 5-point Likert scale (3 = moderately). Other categories are reported descriptively. Descriptive results are summarized in boxplots, see Figure 8.4, Figure 8.5, Figure 8.6, Figure 8.7, Figure 8.8, and Figure 8.9.

In the following, I first report children's perceived game experiences. Then, I report the game's user experience, followed by the subjectively perceived impact of the training on children's literacy skills. Finally, I report children's ratings of individual game elements.

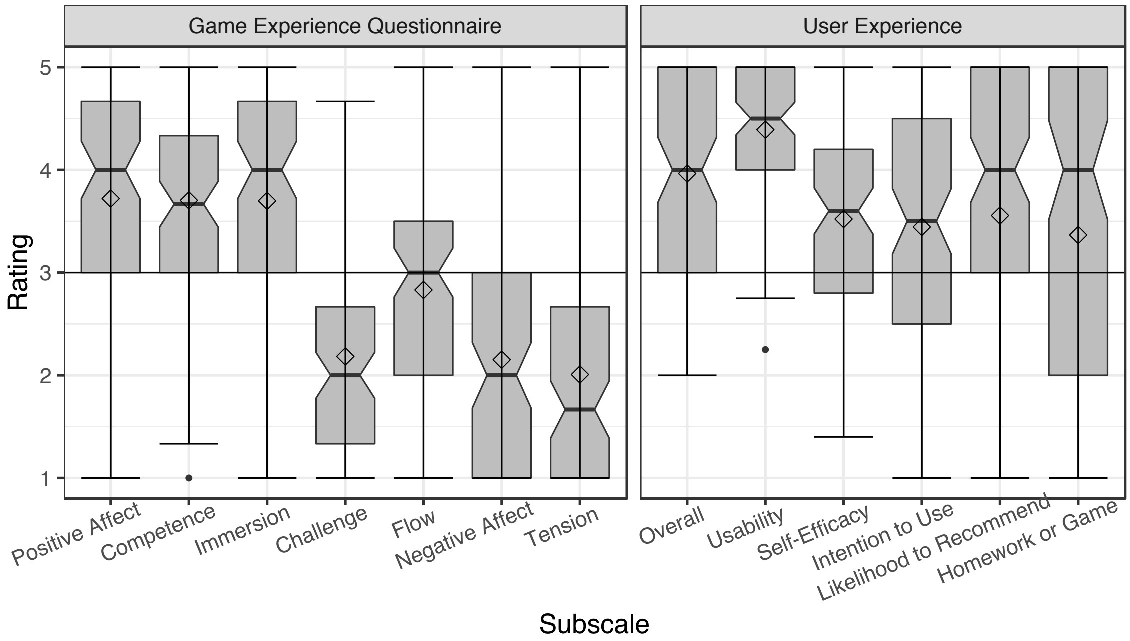


Figure 8.4: Children's ratings of game experience (based on the iGEQ$_+$) and user experience. Notches indicate a 95% confidence interval around the median ($1.58 * IQR/\sqrt{n}$, McGill et al., 1978). Diamond shapes represent mean values.

Game Experience. Children's ratings of the subscales *positive affect* ($M = 3.72,\ SD = 1.10,\ p < .001$), *competence* ($M = 3.70,\ SD = 0.97,\ p < .001$), and *immersion* ($M = 3.70,\ SD = 1.07,\ p < .001$) were significantly above moderately, indicating that children enjoyed Prosodiya and perceived high positive affects and that they felt competent and immersed while playing. In contrast, the game's *challenge* ($M = 2.18,\ SD = 0.91,\ p < .001$) was rated significantly lower than moderately, indicating that the children were not overstrained, but also that the game may have been too easy. Finally, the game's *flow* ($M = 2.83,\ SD = 1.22,\ p = .198$) was rated as moderately. The low challenge value may have been caused by the high percentage of tasks solved at the first go or by the study version of Prosodiya whose difficulty increased gradually to keep the training similar across children and to not overextend poor performers or young children. This may have led to a course of play that is too easy, which may have also influenced children's perceptions of flow.

The results indicate that Prosodiya provides good game experience and does not evoke negative feelings. This is promising as to the validity and effectiveness of Prosodiya, as positive engagement and positive emotions have been shown to positively affect learning (Hamari et al., 2016; Plass et al., 2014) and may increase motivation, satisfaction, and perception towards the learning material (Um et al., 2012). However, the results also indicate that the game's flow and challenge can be improved, which in turn can positively affect learning (Kiili, de Freitas, Arnab, & Lainema, 2012).

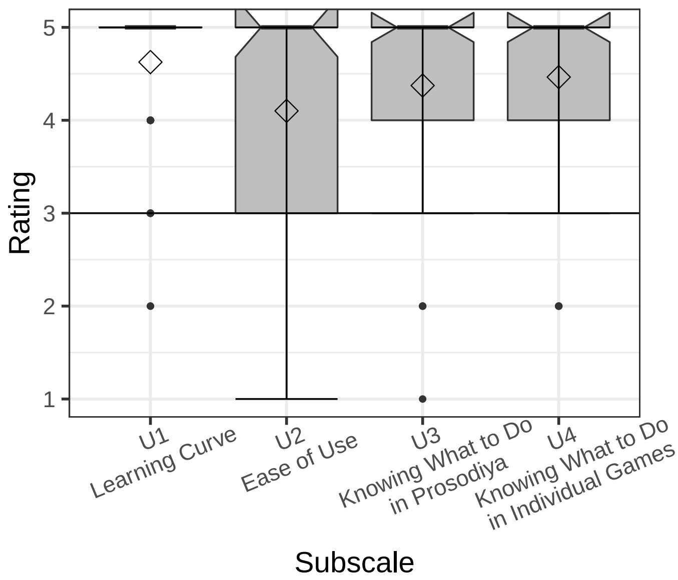


Figure 8.5: Children's usability rating. Notches indicate a 95% confidence interval around the median. Diamond shapes represent mean values.

In sum, the game experience of Prosodiya is promising towards engaging and motivating children to improve their literacy skills in a playful way.

User Experience. Children reported a significantly positive *overall impression* ($M = 3.96,\ SD = 0.93,\ p < .001$), rated the game's *usability* to be "very good" ($M = 4.39,\ SD = 0.66,\ p < .001$), and reported to perceive high *self-efficacy* after training ($M = 3.52,\ SD = 0.92,\ p < .001$), indicated by ratings significantly higher than moderately, see Figure 8.4. This result indicates that children were able to use Prosodiya unassisted and that they perceived a positive influence of the training on their literacy skills. The subscales *intention to use* ($M = 3.44,\ SD = 1.28,\ p < .001$) and *likelihood to recommend* (*UX4*, $M = 3.56,\ SD = 1.29,\ p < .001$) were rated significantly above moderately, indicating that children would likely recommend the game to friends and continue playing it themselves. Finally, children rated Prosodiya to be more like a game, as reflected by ratings significantly above "neither homework nor game"($M = 3.37,\ SD = 1.46,\ p = .042$).

Usability. Besides the averaged value of the usability scale, the ratings of individual usability items are listed in Figure 8.5. Children reported that they quickly learned how to play Prosodiya (*U1*, $M = 4.63,\ SD = 0.74$), that the game was easy to use (*U2*, $M = 4.10,\ SD = 1.14$), and that they knew what they had to do in the game (*U3*, $M = 4.37,\ SD = 0.89$) as well as in the various exercises (*U4*, $M = 4.46,\ SD = 0.84$).

In sum, Prosodiya is reportedly easy to learn and easy to use, which in turn can

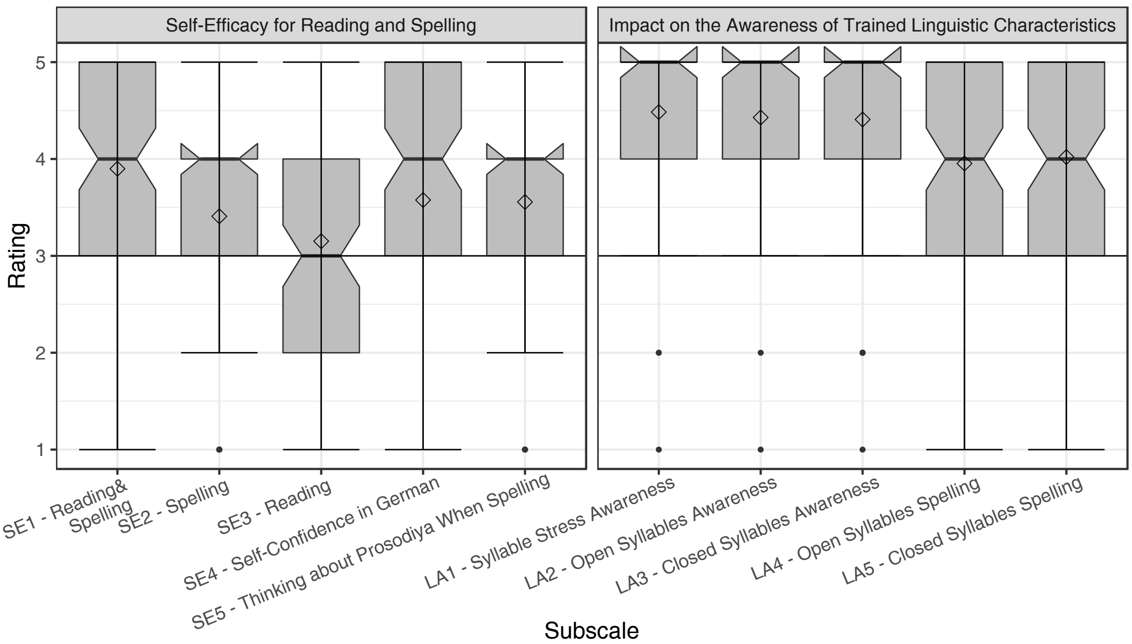


Figure 8.6: Children's perceived reading and spelling self-efficacy and perceived influence on their awareness of specifically trained linguistic characteristics. Notches indicate a 95% confidence interval around the median. Diamond shapes represent mean values.

increase children's learning gain (e.g., Ninaus, Moeller, McMullen, & Kiili, 2017) and, according to the Technology Acceptance Model (TAM; Davis, 1989), is related to the perceived usefulness of a system and its future usage.

Self-Efficacy and Perceived Impact on Specific Linguistic Awareness. Evaluation of individual items of the self-efficacy scale is provided in Figure 8.6. Children reported that they learned quite a lot that helps them in reading and spelling (*SE1*, $M = 3.90$, $SD = 0.99$) and that they perceived high spelling self-efficacy (*SE2*, $M = 3.41$, $SD = 1.16$) and a moderate reading self-efficacy (*SE3*, $M = 3.15$, $SD = 1.40$) after training. Further, they reported to feel more confident in German due to the training (*SE4*, $M = 3.58$, $SD = 1.28$) and that they often think about what they have learned in Prosodiya when they don't know how to spell a word (*SE5*, $M = 3.56$, $SD = 1.12$).

As for the specifically trained linguistic characteristics, children reported that they know after the training what stressed (*LA1*, $M = 4.48$, $SD = 0.86$), open (*LA2*, $M = 4.43$, $SD = 0.82$), and closed syllables are (*LA3*, $M = 4.41$, $SD = 0.92$) and how to spell open (*LA4*, $M = 3.95$, $SD = 1.08$) and closed syllables (*LA5*, $M = 4.02$, $SD = 1.02$), see Figure 8.6.

In sum, children reported that they perceived high self-efficacy after training for spelling-related abilities and a positive influence on linguistic awareness related to syllable stress and vowel length. Boosting self-efficacy and self-esteem are central aim of therapeutic interventions (Bender et al., 2017) and are related to self-awareness of skill increase and actual skill increase (Cleary, Velardi, & Schnaidman, 2017; Schunk, 1989). Furthermore, the perceived usefulness is a central factor influencing the future use of the training (cf. Davis, 1989). Thus, these findings provide further support as to the promising feasibility and validity of Prosodiya.

8.2.2.3 Evaluation of Individual Game Elements

In this section, I briefly report the children's ratings of the pedagogical agents, the narrative, map, and environment of Prosodiya, as well as the tutorials and tooltips explaining game mechanics and teaching linguistic knowledge.

Kugellichter. The ratings of our Kugellichter, the pedagogical agents and narrators of Prosodiya, are displayed in Figure 8.7. We combined the four questions about the agents *K1*–*K4* into a single value K, inverting the response to item *K4* (dislike). Children reported a positive overall impression of the agents (K, $M = 4.32$, $SD = 0.87$), high appeal (*K1*, $M = 4.32$, $SD = 0.90$), and that they enjoyed to play together with them (*K2*, $M = 4.45$, $SD = 0.96$). They also reported to consider them as their friends (*K3*, $M = 3.97$, $SD = 1.35$) and that they experienced very low dislike towards the Kugellichter (*K4*, $M = 1.47$, $SD = 0.98$).

Children qualitatively reported about the Kugellichter that they were one of the best aspects of Prosodiya, that they would like to have them as cuddly toys, and that they especially liked the yellow Kugellicht responsible for spoken feedback. The uniqueness of each Kugellicht was also emphasized positively.

In sum, the results indicate that we successfully designed the Kugellichter as emphatic pedagogical agents, narrators, companions, and maybe even friends. They were very well received and popular with children and the affection was not only limited to the game world, which is indicated by the children's desire for the Kugellichter as cuddly toys from their parents for their birthdays. Thus, the Kugellichter may help dyslexic children to face everyday challenges in literacy acquisition.

Narrative, Map, and Environment. Children reported to be interested in the game's story (*S1*, $M = 3.70$, $SD = 1.20$), that they liked the map of Prosodiya (*M1*, $M = 4.43$, $SD = 0.81$), and that they felt immersed into the fantasy world (*M2*, $M = 3.50$, $SD = 1.54$). They also reported that they liked the fog that covered the map (*M3*, $M = 4.27$, $SD = 1.02$), that they enjoyed dispelling the fog (*M4*, $M = 3.45$, $SD = 1.40$), and that they weren't bothered by the fog (*M5*, $M = 2.46$, $SD = 1.42$). Results of the map and its fog are summarized in Figure 8.7.

Regarding the environments of subchapters and levels, children reported that they liked the background images used (*E1*, $M = 4.36$, $SD = 0.98$) and the fog that they needed to fight back during a level (*E2*, $M = 3.63$, $SD = 1.42$), see Figure 8.7.

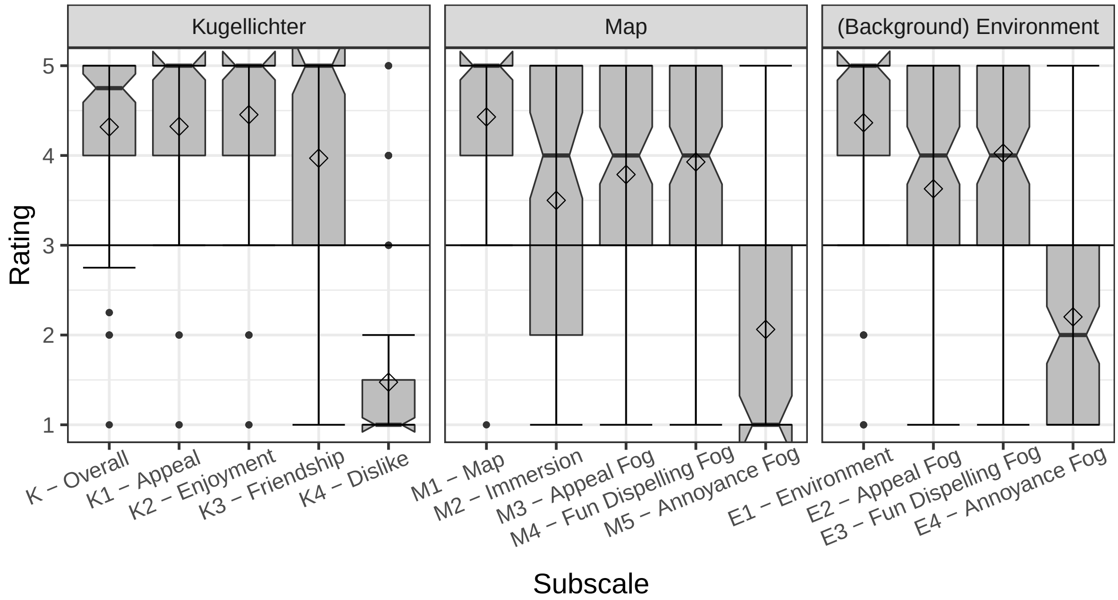


Figure 8.7: Children's ratings of individual game elements: the pedagogical agents (Kugellichter); the map and the fog covering it; and the background environment of levels and the fog that was dispelled in a level. Notches indicate a 95% confidence interval around the median. Diamond shapes represent mean values.

They enjoyed clearing the environment of its fog (*E3*, $M = 4.03,\ SD = 1.16$) while not being annoyed by the fog (*E4*, $M = 2.20,\ SD = 1.44$).

The high rating of Prosodiya's graphical appearance, which includes the design of the pedagogical agents, the world map, and the background environments of levels, indicates that the game appeals the children's tastes and that they felt comfortable in the world of Prosodiya. The story that was introduced with a prologue and continued implicitly in tutorials, the world map, and by changing level environments, has raised the children's interest. Also, the fog was received well, which children needed to dispel from the game's world map and its levels to increase their self-perception of progression and to reward themselves.

In qualitative feedback, children emphasized that they really liked the graphics of Prosodiya, that the map was "like a real world" and one of Prosodiya's highlights, and that "the mission to dispel the fog was great". As the study version of the game did not have a final sequence marking the end of the game, children proposed possible endings, such as a "castle in which the children can interact with the inhabitants", or a cut-scene with the sunrise of Prosodiya that was used as a background image in one of the levels.

Tutorials and Tooltips. Results of the tutorials are summarized in Figure 8.8. Children reported that the pedagogical agents explained the game mechanics and linguistic knowledge well (*T1*, $M = 4.17,\ SD = 0.96$), that the tutorials were kind

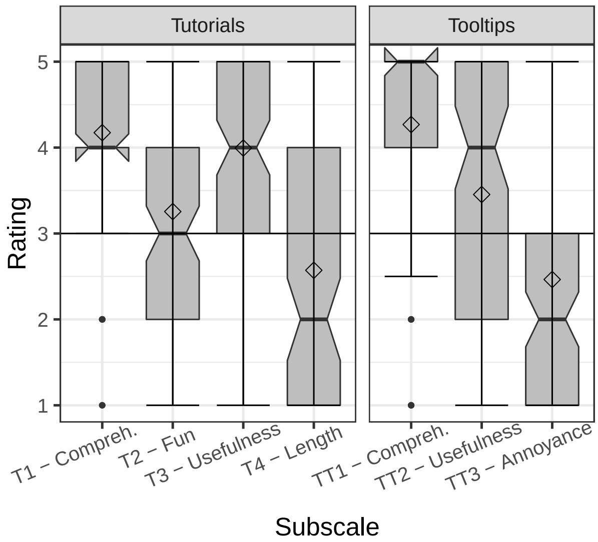


Figure 8.8: Children's ratings of tutorials and tooltips. Notches indicate a 95% confidence interval around the median. Diamond shapes represent mean values.

of fun (*T2*, $M = 3.26,\ SD = 1.36$), and that they were very useful in order to learn the "secret of words" (*T3*, $M = 3.99,\ SD = 1.09$). The duration of the tutorials were perceived to be not too long (*T4*, $M = 2.57,\ SD = 1.49$).

The tooltips were rated to be very comprehensible (*TT1*, $M = 4.27,\ SD = 1.02$) and to be very useful (*TT2*, $M = 3.45,\ SD = 1.40$) while not annoying the children (*TT3,* $M = 2.46,\ SD = 1.42$), see Figure 8.8.

The refined tutorials and newly introduced tooltips served their purposes to explain game mechanics and to convey linguistic knowledge. Children reported enjoying the tutorials and did not perceive them as too long. We may also infer that the tooltips popping up prior to each level were good solutions as regular reminders of task objectives and linguistic properties to pay attention.

Individual Games. Children reported enjoying all games of the study version, as indicated by high values to the question how much they enjoyed the games "Stress pattern" (*G1*, $M = 4.18,\ SD = 1.05$), "Open and closed syllables' (*G2*, $M = 3.84,\ SD = 1.13$), "Orthographic markers" (*G3*, $M = 3.93,\ SD = 1.12$), and "Spelling" (*G4*, $M = 3.79,\ SD = 1.34$), see Figure 8.9. Although the game "Stress pattern" received the highest value in enjoyment, the majority of 35 of the children chose "Spelling" (39%) as their favorite game, closely followed by "Stress pattern" favored by 25 children (28%) and "Orthographic markers" favored by 24 children (27%). The game "Open and closed syllables" was chosen only by five children (6%) as their favorite game. Children argued that while the other games are also fun, the game "Spelling" has potentially the highest impact on their literacy skills.

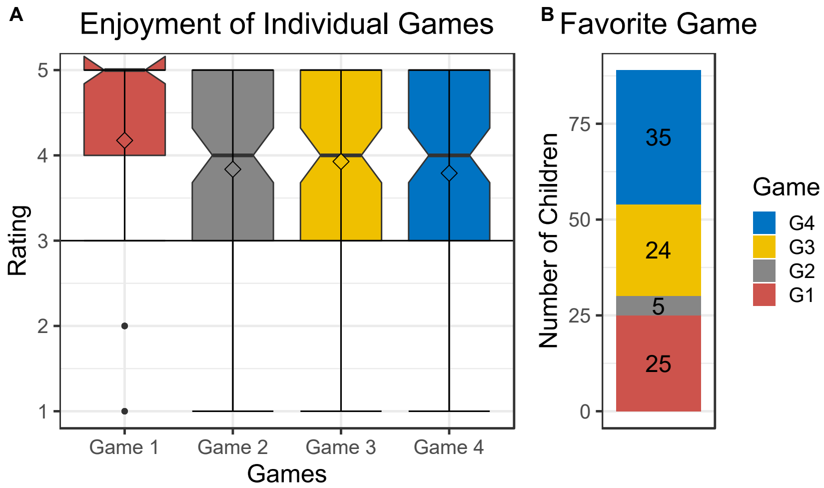


Figure 8.9: Children's game ratings. **A**: Rating of the individual games (G1="Stress pattern"; G2="Open and closed syllables"; G3="Orthographic markers"; G4="Spelling"). Notches indicate a 95% confidence interval around the median. Diamond shapes represent mean values. **B**: Overall favorite game.

Feedback and Rewards. Although we did not ask questions about feedback or rewards, children used the comment fields of the questionnaire to talk about it. Regarding feedback, they reported the praising words spoken by the yellow Kugellicht as one of Prosodiya's highlights. Other children complained that they were missing the joyful feedback in the later course of the game, in which the yellow Kugellicht was no longer a game element. This was especially the case when exercises became harder and one of the children's major challenges – spelling – was dealt with.

Regarding rewards, we investigated the times children additionally practiced levels after their completion and identified that children eagerly tried to gather three stars for each level. In the beginning of the present study, we defined that a flawless level grants the third star. We realized that poorer performing children had major problems solving all exercises in a level at the first go and hence were stuck in the game, trying to get the third star. In response, we lowered the threshold to 90% and implemented a fall-back system granting the third star after a level has been practiced a defined number of times.

8.2.3 Conclusion and Outlook

In the previous sections, I examined Prosodiya's feasibility based on training behavior and feedback collected in training experience questionnaires. That is, we investigated if Prosodiya can be used unassisted by primary school children in the home environment to support their literacy acquisition. Further, we evaluated children's training experience and perceived self-efficacy. Investigating the applicability of the training in the home environment – under real-world conditions – is important to determine whether the effects found in the present study may transfer to real-life context outside of scientific studies conducted in controlled environments.

Prosodiya was found to be easy to use and children spent on average 10 hours within the game and 76% of the children completed the training. This completion rate is comparable to studies conducted in controlled environments. Moreover, Prosodiya was able to engage and motivate the children throughout the training period and the children reported positive training experience. Further, the children considered Prosodiya more as a game than homework and would likely recommend it to friends and keep on playing it themselves. However, children also reported that the training might have not been challenging enough and that it offered only moderate flow experience. Regarding perceived influence of the training on children's literacy skills, children reported that they perceived high self-efficacy after training and that they perceived a high positive impact on spelling-related abilities and a moderate impact on reading. The training was received very positively also by parents and teachers and many families reported that they would continue the training or recommend it to others.

Additionally, the children indicated very positive perception of the game's graphics and elements. Especially the Kugellichter, which the kids befriended during the game, were very popular. Prosodiya's graphics appealed to the children's tastes and the story and map supported a sense of immersion and self-perception of progression. The tutorials and tooltips provided easy introductions into individual games, which is necessary when such trainings should be used unassisted. Further, the number of times children practiced old levels, even when new content had been unlocked, revealed that the stars awarded to level completions promoted self-competition and supported replayability of levels. Lastly, the importance of positive feedback and praise was emphasized especially in later course of play that dealt with spelling, one of the children's major challenges.

To conclude, the training behavior and positive feedback demonstrates the feasibility of Prosodiya. The results indicate that we succeeded in designing a game-based training program that can easily be used by children without additional instructions from parents, teachers, or learning therapists, and that is able to engage and motivate children over several months in order to positively affect their literacy acquisition. Hence, I may infer that such digital game-based trainings yield the potential to meliorate reading and spelling skills in dyslexic children in addition to school and learning therapy, and that the effectiveness of such trainings can successfully be investigated in the home environment under real-world conditions. For the present study, this means that the results reported in the following sections on the effectiveness of Prosodiya are not limited to controlled environments, but rather reflect real-life learning gains.

In the future, further development should address increasing the game's flow and challenge, e.g., by continuing the development of an adaptive user model and integration of daily narratives, in order to keep game and training experience high over a longer time. We already added more cutscenes to increase immersion and flow due to the feedback received in the study. Furthermore, we aim to explore new game mechanics, such as time constraints or moving objects, to increase the flow experience within the game. As children also reported that they would have liked to spend their collected points on rewards, we are planning the development of an extrinsic reward systems to foster long-term motivation. The reward systems is planned to feature a dragon's stronghold and cave in which children may redeem their collected points to buy a dragon egg, incubate it, and raise the hatched dragon to be their companion and play bonus games with.

8.2.4 Summary

Summary of Prosodiya's feasibility and training experience

Training behavior

- Children spent an average 10 hours with the game and 88 of the 116 children (76%) completed the entire training.

Training experience

- Children reported positive game experience with the training.
- Challenge and flow of the training can be improved.
- Prosodiya was found to be easy to learn and easy to use.
- Children reported that they would likely recommend Prosodiya to friends and keep on playing it themselves.
- Prosodiya was seen more as a game than as homework.

Self-efficacy and perceived impact on literacy skills

- Children reported that they perceived high spelling self-efficacy after training and a high positive impact on spelling-related abilities and linguistic awareness related to syllable stress and vowel length.
- Children reported that they perceived a moderate impact on reading.

Individual game elements

- The individual game elements (pedagogical agents; narrative, map, and environment; tutorials and tooltips; feedback and rewards) were positively perceived.
- Children's favorite game elements are the Kugellichter and their favorite game is "Spelling".

→ Prosodiya seems to be an engaging mobile game-based spelling intervention. The training was perceived very positively and can be used unassisted by primary school children in the home environment and children perceived a positive influence on their syllable stress awareness and spelling skills.

8.3 Effectiveness

So far, we were able to demonstrate that Prosodiya is feasible for the use at home and that children enjoyed the training. Moreover, children reported to perceive a positive impact on their literacy skills. However, the million-glass-blossom question whether the training actually improves syllable stress awareness and reading and spelling skills has not been answered. This question is taken care of in this section.

The evaluation of Prosodiya with regard to educational effectiveness is reported in our article "A Digital Game-Based Training Improves Spelling in German Primary School Children – A Randomized Controlled Field Trial" (Holz et al., unpublished). This section is mainly based on that article.

8.3.1 Methods

8.3.1.1 Measures

To evaluate the efficacy of Prosodiya, we examine syllable stress awareness, general spelling ability, spelling performance of specific orthographic learning categories, reading fluency, and word reading.

Syllable Stress Awareness was assessed using an individually administered paper version of the game "Stress pattern", in which children have to identify the stress pattern of 15 training words. Scoring is based on the number correctly identified stress pattern. The test items are listed in Appendix A.1.

Spelling was assessed with a standardized classroom cloze spelling test (*DRT 2/3/4, Diagnostischer Rechtschreibtest für 2./3./4. Klassen* [Diagnostic Spelling Test for 2nd/3rd/4th Grade]; Grund, Leonhart, & Nauman, 2017; Müller, 2003a, 2003b) and a self-designed classroom cloze spelling test. Scoring of the spelling tests is based on the number of correctly spelled words.

For the standardized spelling test, norm-referenced scores are standard scores (T-scores) with a mean of 50 and standard deviation of 10. For analyses, we used the standard score for the entire test. For the category *vowel length marking*, we used the raw score of the number of words with spelling mistakes in vowel length marking (error category *"D"* in Müller, 2003a)[1] since standardized scores of this error category are not available in the DRT 4.

The self-designed spelling test was used to further investigate transfer of learning. The test was the same for all grades and consisted of 30 words. The test was administered only at T2 and T3. For analyses, we used in total five raw scores: (i) the number of correctly spelled words in the entire test, (ii) the number of words with spelling mistakes in vowel length marking, and (iii) the number of correctly

[1] Transforming the raw scores into grade-specific z-scores lead to the same statistical test decisions.

spelled training words (no transfer learning, 9 words). Additionally, we assessed transfer of learning to untrained domains, i.e., we analyzed the raw scores of (iv) correctly spelled uninflected untrained words (near transfer learning, 10 words) and (v) correctly spelled inflected training words (far transfer learning, 11 words). The self-designed spelling test is listed in Appendix A.2.

Reading. In addition to syllable stress awareness and spelling skills, we also analyzed transfer effects on untrained reading skills, i.e., reading fluency and word reading, that are related to phonological awareness and spelling but that are not explicitly trained in the program.

Reading fluency was assessed with a standardized classroom reading test (*SLS 2–9, Salzburger Lese-Screening für die Schulstufen 2–9* [Salzburg Reading Screening for Grades 2–9]; Mayringer & Wimmer, 2014) in which children read silently as many sentences as possible in three minutes and mark them as either true or false (e.g., "you can drink water" is true while "strawberries can speak" is false). Scoring is based on the number of correctly marked sentences. Norm-referenced scores are standard reading scores (reading quotient, LQ-score) with a mean of 100 and standard deviation of 15. The norm table of the handbook is limited to LQ-scores in the range between 62 and 138. For the analyses, we use the standard reading score.

Word reading was assessed in a standardized one-minute reading speed test (*SLRT-II: Lese- und Rechtschreibtest* [SLRT-II: Reading and Spelling Test]; Moll & Landerl, 2010) in which children read aloud words as fast as possible without making errors from a reading list. The test contains a word and a pseudoword reading list with increasing word length and complexity. Scoring is based on the number of correctly read words. Norm-reference scores are percentile ranks. For the analyses, we calculated and used z-scores based on the norm sample.

8.3.1.2 Analysis

The evaluation of Prosodiya's educational effectiveness was carried out in a two-step process. First, cross-over analyses were performed to investigate if children's learning gains induced by the training is significantly higher than that obtained during waiting periods without extra training. For this, we compared the within-subject differences between the two training periods from the immediate treatment and the delayed treatment group with regard to the outcome variables, following the analysis for two-group two-period cross-over trials proposed by Hills and Armitage (1979). That is, we calculated changes in the outcome variables for both training periods (T2−T1 and T3−T2, respectively) by group and analyzed the within-subject period differences ([T2 − T1] − [T3 − T2]) in our outcome measures between the immediate and the delayed treatment group with two-sample t-tests. This analysis is recommended as the standard approach to investigate treatment effects for two-group two-period cross-over trials when controlling for possible time effects (Senn, 2002; Wellek & Blettner, 2012). In case of significant treatment effects, Cohen's d

effect sizes based on the pooled standard deviations were calculated. According to Hattie (2008), effect sizes can be considered small if $d = 0.2$, medium if $d = 0.4$, and large if $d = 0.6$ when evaluating educational outcomes.

In the second step, we examined whether potential training effects are found during the first and second training period. For this, we applied planned contrasts to analyze separately changes in the outcome measures from pre- (T1) to mid- (T2) and from mid- (T2) to posttest (T3). We analyzed the group differences in learning gains between T1 and T2 and between T2 and T3 by means of one-way ANCOVAs, comparing group effects on gain scores of the outcome variables at T2 and at T3 with the pretest scores of respective tests of the respective training period (T1 for the first training period, T2 for the second), diagnosis of dyslexia, sex, and grade treated as covariates. In case of significant group effects, we estimated between-group effect sizes $\hat{d}$ separately for the learning gain using the adjusted mean difference between the active training group and the control group divided by the estimated pooled standard deviation obtained from the square root of the mean squared error of the ANCOVA models, i.e., $\hat{d} = \frac{\overline{X'}_{training} - \overline{X'}_{control}}{\sqrt{MSE'}}$ (Grissom & Kim, 2012, p. 349).

Exclusion of Participants. We excluded individual participants from respective analyses due to absence, non-completion of a test, flawed test administrations, and based on outlier analyses. In outlier analyses, we excluded participants from respective analyses whose period differences (cross-over analyses) or standardized residuals (planned contrast ANCOVAs) deviated more than 2.5 standard deviation from the mean of the respective treatment group. Detailed description of the exclusion criteria and exact sample sizes for the individual analyses can be found in our article "A Digital Game-Based Training Improves Spelling in German Primary School Children – A Randomized Controlled Field Trial" (Holz et al., unpublished).

8.3.2 Results and Discussion

8.3.2.1 Effects on Syllable Stress Awareness

The cross-over analysis revealed a significant training effect of large effect size, $t(84.83) = 7.32$, $p < .001$, $d = 1.57$. The training-induced learning gain in syllable stress awareness was significantly higher than the change induced by waiting periods without extra training ($M_{\text{diff}} = 4.0$, 95% CI_{diff} [2.92, 5.10]). Figure 8.10 indicates that, during both training periods, the active training group improved at a significantly higher rate in syllable stress awareness than the control group that did not receive the training. improved at a significantly higher rate in syllable stress awareness than children in the control condition without extra training.

The planned contrast analyses confirm the training effect. We found a significant group effect of large effect size on syllable stress awareness during the first training period, $F(1,\ 96) = 46.86$, $p < .001$, $\hat{d} = 1.49$, as well as during the second training period, $F(1,\ 88) = 26.22$, $p < .001$, $\hat{d} = 1.25$.

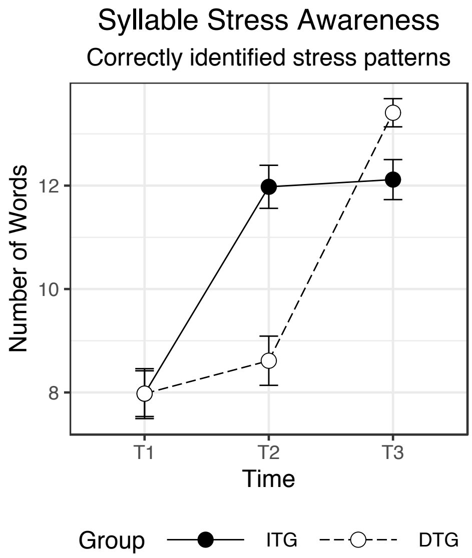


Figure 8.10: Mean scores of syllable stress awareness (max=15) by group (ITG = immediate treatment; DTG = delayed treatment) at the pre- (T1), mid- (T2), and posttest (T3). Bars represent the standard errors of the mean.

In sum, the analyses revealed that children's abilities to correctly identify stress patterns improved at a significantly higher rate when they received the training, which is confirmed by significant effects in favor of our training program found during both training periods. Thus, we may infer that Prosodiya has a strong positive impact on children's syllable stress awareness, providing first evidence of the training's pedagogical approach to support literacy acquisition.

8.3.2.2 Effects on Spelling

To examine if the positive impact of Prosodiya goes beyond improving syllable stress awareness and if the perceived-self efficacy for spelling-related abilities can actually be confirmed in spelling tests, I analyze in the following section the results of the standardized spelling test followed by the analyses of our self-designed spelling test.

Standardized Spelling Test. The cross-over analysis revealed a large significant training effect on standardized spelling scores, $t(80.82) = 2.79$, $p = .007$, $d = 0.60$, see Figure 8.11. The training-induced learning gain was significantly higher than the learning gain obtained during waiting periods ($M_{\text{diff}} = 2.45$ T-scores, 95% CI_{diff} [0.71, 4.20]). We also found a large significant training effect on vowel length marking, $t(84.30) = 3.28$, $p = .001$, $d = 0.70$. The training-induced improvement in the orthographic vowel length marking was significantly higher than the learning gain during waiting periods without extra training ($M_{\text{diff}} = 1.59$, 95% CI_{diff} [0.63, 2.56]).

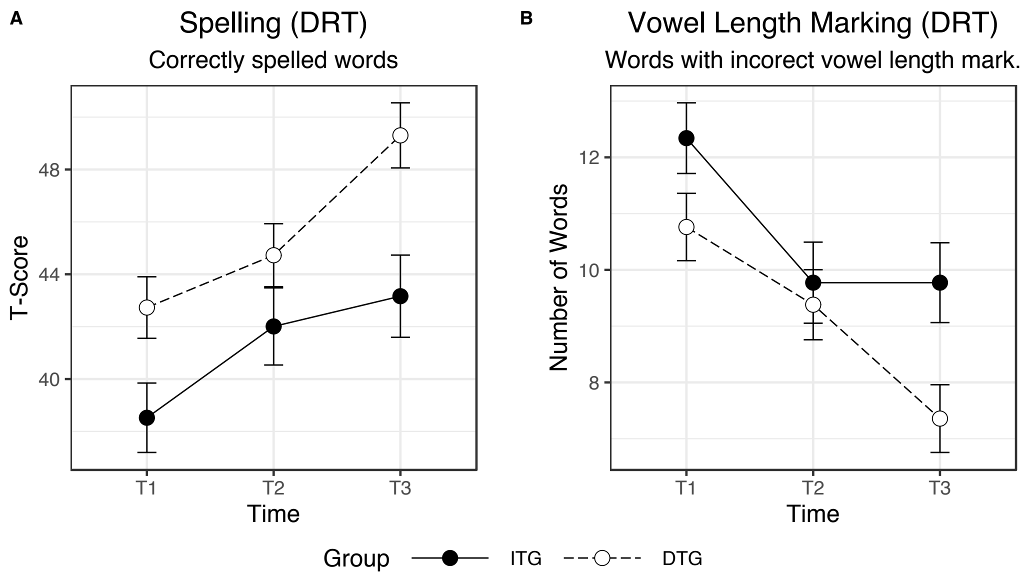


Figure 8.11: Mean scores on the standardized spelling test by group (ITG = immediate treatment; DTG = delayed treatment) at the pre- (T1), mid- (T2), and posttest (T3). Bars represent the standard errors of the mean. **A**: Total score (standardized); **B**: Number of words with incorrect vowel length marking.

The training effect is confirmed in the planned contrast analyses, see Figure 8.12. The analyses of covariance revealed a large significant group effect on the standardized spelling scores during the first training period, $F(1,\ 95) = 7.13,\ p\ =\ .009,\ \hat{d} = 0.60$, and a significant group effect of medium to large effect size during the second training period, $F(1,\ 90) = 5.85,\ p\ =\ .018,\ \hat{d} = 0.51$. The contrast analyses of the orthographic marking of short and long vowels yield similar results. A large significant group effect was found during the first training period, $F(1,\ 95) = 12.02,\ p < .001,\ \hat{d} = 0.76$, as well as during the second training period, $F(1,\ 92) = 8.77,\ p\ =\ .004,\ \hat{d} = 0.63$. Figure 8.11 and Figure 8.12 indicate that children in the active training condition improved at a significantly higher rate in spelling than children in the control condition.

In summary, the analyses of the standardized spelling test indicate that children's spelling abilities improved at a significantly higher rate when they received the training, proving the effectiveness of Prosodiya. This applies to the general spelling ability as well as to the explicitly practiced orthographic marking of long and short vowels. Moreover, the immediate treatment group did not decline in spelling during the second training period, i.e., they could maintain their performance level at T3 without further training, indicating a long-term effect of the training.

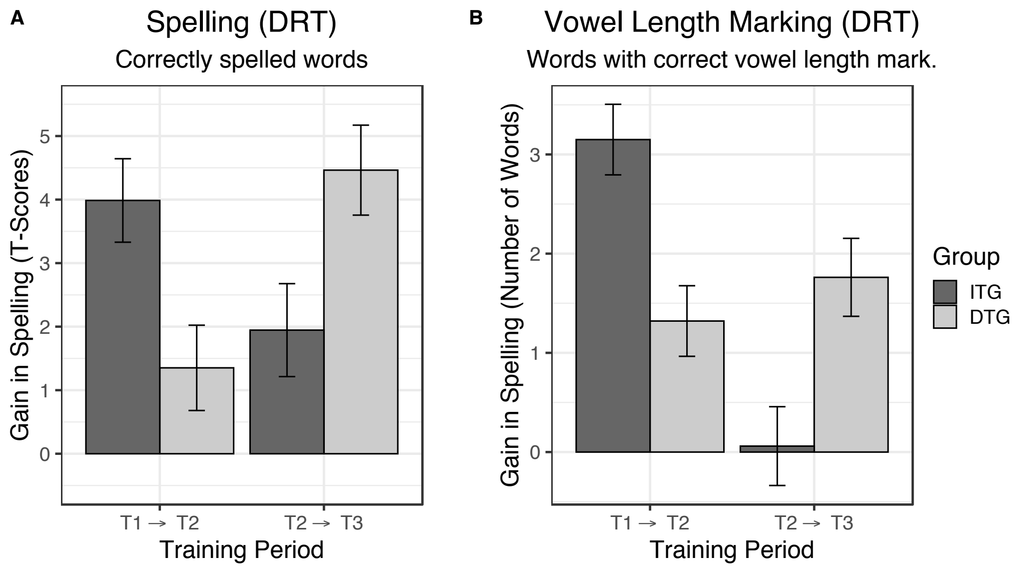


Figure 8.12: Estimated marginal means of spelling gains on the standardized spelling test by group (ITG = immediate treatment; DTG = delayed treatment) and training period. Bars represent standard errors of the mean. **A**: Gain in the standardized total score; **B**: Gain in vowel length marking (i.e., words with *correct* vowel length marking).

Self-Designed Spelling Test. To investigate if the previously found training effect is confirmed in specific orthographic learning categories, I examine the results of our self-designed spelling test in the following. The self-designed spelling test was administered at T2 and T3 and covered learning categories that were not available across grades in the standardized spelling test. During the second training period, the delayed treatment group received the training while the immediate treatment group served as the control.

The evaluation of the self-designed spelling test yields similar results as those found in the standardized spelling test. Figure 8.13 indicates that the spelling improvement was significantly higher among children from the delayed treatment group, who received the training, than in the immediate treatment group in all learning categories but one (training words). In fact, we found a large significant group effect on the total score, $F(1,\ 85) = 14.80$, $p < .001$, $\hat{d} = 0.86$, as well as on the orthographic marking of long and short vowels, $F(1,\ 85) = 16.40$, $p < .001$, $\hat{d} = 0.92$.

Furthermore, we found transfer effects on untrained learning categories. That is, we found a large significant group effect on spelling of uninflected untrained words (near transfer learning), $F(1,\ 84) = 20.40$, $p < .001$, $\hat{d} = 1.02$, and a medium significant group effect on spelling of inflected training words (far transfer learning), $F(1,\ 84) = 5.17$, $p = .026$, $\hat{d} = 0.51$.

For uninflected training words (no transfer learning), we found a marginal though not significant group effect, $F(1,\ 84) = 3.61$, $p = .061$, $\hat{d} = 0.43$. As Figure 8.13 indicates, the group difference in uninflected training words is not significant due

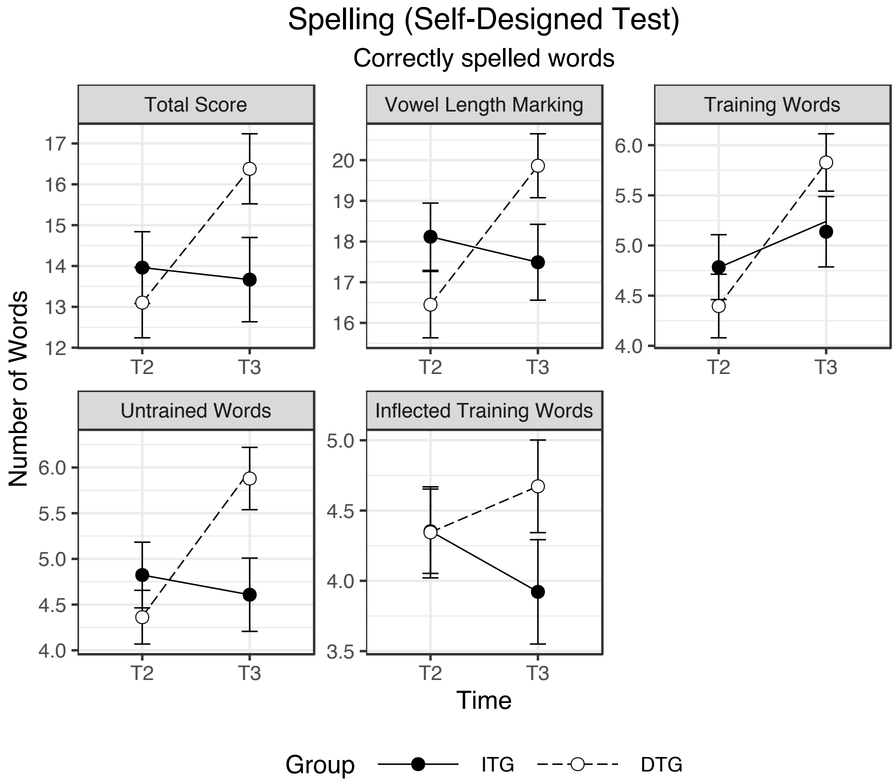


Figure 8.13: Mean scores on the self-designed spelling test by group (ITG = immediate treatment; DTG = delayed treatment) at the mid- (T2) and posttest (T3). Categories from top-left to right-bottom: (i) total score (max=30); (ii) number of words with correct vowel length marking (max=30); (iii) training words (no transfer learning, max=9); (iv) untrained words in basic form (near transfer learning, max=10); and (v) inflected training words (far transfer learning, max=11). Bars represent the standard errors of the mean.

to a noteworthy learning gains in the immediate treatment group, which may result from consolidation effects.

The results of the self-designed spelling tests confirm the findings of the standardized spelling test. We found a significantly higher spelling improvement in the active training group compared to the control group in the general spelling ability, in the orthographic marking of long and short vowels, as well as in the categories of near and far transfer of learning. Children did not only improve in spelling of training items, but were also able to apply the acquired knowledge on trained spelling rules to untrained uninflected words as well as to inflected training words.

In sum, the results of the spelling tests indicate a significant training effect on spelling abilities, which confirms children's perceived positive influence on spelling as reported in the training experience questionnaire and further supports Prosodiya's pedagogical approach and effectiveness.

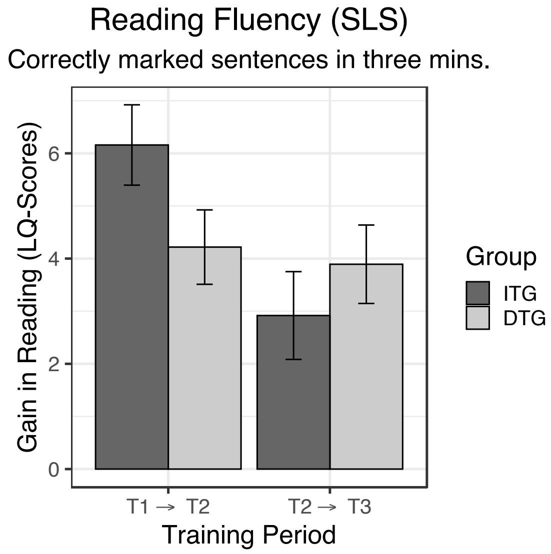


Figure 8.14: Estimated marginal means of the learning gains in reading fluency by group (ITG = immediate treatment; DTG = delayed treatment) and training period. Bars represent the standard errors of the mean.

8.3.2.3 Effects on Reading

To this point, we have shown that Prosodiya has a beneficial effect on trained literacy skills, i.e., syllable stress awareness and spelling skills. In the following, I further investigate if Prosodiya has a transfer effect on untrained reading skills by analyzing children's reading fluencies and word reading.

Reading fluency. The cross-over analysis revealed no significant training effect on reading fluency, $t(66.97) = 0.74,\ p = .465$.

As for the planned contrasts, the ANCOVA of the first training period revealed a marginal though not significant group effect, $F(1,\ 77) = 3.24,\ p = .076$, while the group effect during the second training period was not significant, $F(1,\ 75) = 1.12,\ p = .293$. Figure 8.14 indicates that the improvement in reading fluency was more pronounced though not significantly higher in the active training condition than in the control condition. When we exploratively examined the data for sources that might have inhibited a potential training effect on reading fluency, we found a strikingly high development of fourth graders from the immediate treatment group during the second training period, suggesting that they undperformed at T2 and thus might have skewed the results. Excluding the fourth graders from the analyses leads to a marginal though not statistically significant group effect during the second period, $F(1,\ 59) = 2.78,\ p = .100$.

Word Reading. The cross-over analysis revealed no significant training effect on word reading, $t(63.31) = 0.11,\ p = .909$, nor on pseudoword reading, $t(66.43) = 0.37,\ p = .710$. As for the planned contrasts, the group effects on word and on pseudoword reading were not significant during the first training period, $t(88.95) = -1.53,\ p = .130$, and $W = 1097.5,\ p = .695$, nor during the second training period, $F(1,\ 75) = 0.15,\ p = .700$, and $F(1,\ 75) = 0.36,\ p = .551$.

In sum, we did not find significant training effects on untrained reading skills. This result is not too surprising as the training program did not include explicit exercises on reading-related (precursor) skills. However, the different stages of the acquisition of each literacy skill require specific treatment approaches (cf. Galuschka et al., 2014; Galuschka & Schulte-Körne, 2016) . Yet, we found primary indications that the training meliorates the reading fluency of some children. In the future, we plan to thoroughly examine children's characteristics that relate to the responsiveness to the training with regard to reading fluency. Further, the training could be complemented with exercises covering reading related (precursor) skills.

8.3.2.4 Additional Analyses

After the effectiveness analyses revealed a significant training effect on syllable stress awareness and spelling abilities, I investigated potential factors that may have influenced the success of the intervention, detailed in the following.

We calculated the total change in the T-scores of the standardized spelling test that can be attributed to the training, computed by subtracting the waiting-induced improvement from the training-induced improvement for each child included in the cross-over analysis. This absolute improvement was subjected to a bidirectional stepwise linear regression analysis with pre-treatment score (T1 for the immediate and T2 for the delayed treatment group), diagnosis of dyslexia, grade, sex, and group assignment as possible predictors. We used Akaike's information criterion (Akaike, 1998) to find the most appropriate model in the stepwise regression analysis. The stepwise regression analysis revealed that only two of the possible predictors reliably predicted the improvement in spelling attributed to the training, namely the pre-treatment spelling score, $\beta = -0.2,\ SE = 0.1,\ t(83) = -2.09,\ p = .039$, and sex, $\beta = 3.7,\ SE = 1.7,\ t(83) = 2.17,\ p = .034$. Group assignment, diagnosis of dyslexia, and grade did not significantly predict the improvement. The results indicate that the training success increased with a decreasing pre-treatment spelling ability and was more pronounced in girls than in boys. Interestingly, upon further investigation, we found a marginal significant interaction between pre-treatment score and sex on the spelling improvement attributed to the training. While the spelling improvement in girls only increased slightly with decreasing spelling ability, boys tend to improve in spelling more strongly with decreasing initial spelling ability. Possibly, the attitude towards the training, i.e., the children's awareness that they need the training and may benefit from it and the willingness to practice conscientiously, may be differently pronounced in boys and girls with different spelling abilities.

8.3.3 Conclusion and Outlook

In the previous sections, I have examined the training effect of Prosodiya on syllable stress awareness and reading and spelling skills. For this, we analyzed the outcome measures of 116 German primary school children in a randomized controlled field trial with a two-period, wait-list controlled crossover treatment design. During respective training periods of 9–10 weeks, children in the active treatment condition received the training program and trained at home on Android tablets.

We have shown that children improved their syllable stress awareness and spelling skills at a significantly higher rate when they actively trained with Prosodiya at home, compared to waiting periods in which they did not receive extra training. We found significant training effects of medium to large effect sizes in cross-over analyses evaluating within-subject period differences as well as in planned contrasts analyzing the individual training periods separately by means of analyses of covariance. As for spelling, the training effect was found in the general spelling ability as well as in the orthographic marking of long and short vowels. Moreover, the immediate treatment group maintained their improved spelling level even 9–10 weeks after they completed the training, indicating the long-term effect of the intervention. Additionally, we found evidence of near and far transfer of learning in the delayed treatment group. The results of the self-designed spelling test show that children improved in spelling of untrained uninflected words as well as of inflected training words at a significantly higher rate than their peers without training. Thus, the program does not only improve the spelling of training words, but also affects untrained spelling skills, such as the spelling of untrained words and inflected word forms. Our results are in line with the consistent finding that improving orthographic knowledge improves the spelling abilities in German (dyslexic) primary school children (cf. Galuschka & Schulte-Körne, 2016; Ise et al., 2012).

In the current state, the training had no significant impact on untrained reading skills, i.e., reading fluency and (pseudo-) word reading. Interestingly, the results are in line with the perceived reading impact, which was reported as only moderate and significantly lower than the perceived impact on spelling-related abilities. However, we found first indications that some children's reading fluency may benefit from the training, which could be investigated more thoroughly in future data analyses and studies. Further, as the different stages of the acquisition of each literacy skill require specific treatment approaches (cf. Galuschka et al., 2014; Galuschka & Schulte-Körne, 2016), the training could extend its current spelling-specific focus by adding modules that specifically target reading (precursor) skills.

In the present study, the average training-induced improvement in spelling, obtained from the estimated marginal means of the ANCOVAs of a standardized spelling test, was +4.0 T-scores in the immediate treatment group and +4.5 T-scores in the delayed treatment group. These learning gains are comparable to empirically evaluated computer-based interventions to improve spelling in German primary school children in which the children either trained during school lessons (e.g., Klatte et al., 2018) or in supervised training sessions and at home (e.g., Kargl et al., 2008), as well as to paper-based interventions in which children had weekly training sessions with trained personnel (e.g., Ise & Schulte-Körne, 2010; Reuter-Liehr, 1993; Schulte-Koerne, Deimel, Huelsmann, Seidler, & Remschmidt, 2001). Of the referenced interventions, our approach is most similar to the *Marburger Rechtschreibtraining* (Marburg Spelling Training; Schulte-Körne & Mathwig, 2013), which has been shown to improve the spelling in dyslexic children in grades 2–4 by around +3.2 T-scores (twelve weekly training sessions with trained personnel of 45 minutes each; Schulte-Koerne et al., 2001) and the spelling in dyslexic children in grades 5–6 by between +3.5 and +5.3 T-scores (twelve to fifteen weekly training sessions with trained personnel of 60 minutes each; Ise & Schulte-Körne, 2010). Considering the treatment duration and absolute training time in the present study, our results show that digital game-based interventions can significantly improve spelling in primary school children with comparable learning gains that may even outperform individually administered training sessions. Moreover, digital game-based trainings like Prosodiya can be used by children independently without permanent supervision of parents or trained personnel. Consequently, the training can take place anytime and anywhere – as long as the children have access to a tablet or smartphone. Further, the results demonstrate that our innovative approach yields results comparable to traditional training methods. The approach to systematically teach orthographic knowledge in combination with the awareness of syllable stress seems to be equally beneficial. It may therefore expand the traditional pool of training methods.

However, the present study also has some limitations. First, due to the scope and complexity of the training, the learning gains in spelling cannot explicitly be attributed. It is not clear whether they result from specific training components (e.g., syllable stress awareness), the combination of specific components (e.g., syllable stress awareness and orthographic marking), or the integration of all components in the holistic intervention, and to what extent the playful implementation meliorated the learning gains. Yet, it seems reasonable that a holistic approach in the orthographic stage of spelling acquisition is effective when it includes – besides morphological skills, lexical knowledge and knowledge of spelling rules (cf. Galuschka & Schulte-Körne, 2016; Ise & Schulte-Körne, 2010) – syllable stress awareness, particularly in the spelling of long and short vowels (Sauter et al., 2012). Second, we observed significant differences in learning gains among children. While the majority profited from the training, each treatment group also included some non-responders, i.e., children whose spelling scores did not change or even declined over time. In the future, predictors of children's responsiveness could be addressed, e.g., by enhancing the adaptive learner model, to ensure effective training for each child.

To summarize, the results show that Prosodiya evidentially improves syllable stress awareness and spelling abilities outside the classroom and learning therapy, confirming children's perceived positive impact of the training. In the future, we plan to include modules focusing on morphological skills, such as word stem identification, to explicitly teach children the spelling of inflected word forms.

8.3.4 Summary

Summary of Prosodiya's effectiveness

Effects on syllable stress awareness

- Prosodiya significantly improves children's syllable stress awareness.

Effects on spelling

- Prosodiya significantly improves children's spelling skills.
- We found significant training effects on general spelling ability, on the orthographic marking of long and short vowels, as well as on untrained spelling categories of near (uninflected untrained words) and far transfer learning (inflected training words).
- The average training-induced improvement in spelling was +4.0 T-scores in the immediate treatment group and +4.5 T-scores in the delayed treatment group. These learning gains are comparable to other empirically evaluated spelling trainings for German primary school children.
- Children from the immediate treatment group maintained their improved spelling levels even months after the intervention, indicating long-term effectiveness.

Effects on reading

- No significant training effects were found on untrained reading abilities, i.e., reading fluency and (pseudo-) word reading.
- However, we found primary indications that Prosodiya potentially meliorates children's reading fluencies.

Factors influencing training success

- Gender and pre-treatment spelling ability significantly predicted spelling improvement attributed to the training.
- The results indicate that the training success increased with a decreasing pre-treatment spelling ability and was more pronounced in girls than in boys.
- Group assignment, diagnosis of dyslexia, and grade did not significantly predict spelling improvement.

→ Prosodiya significantly improves children's syllable stress awareness and spelling skills. The approach to systematically teach orthographic knowledge in combination with the awareness of syllable stress seems to be equally beneficial as traditional training methods, expanding the traditional pool of training methods. Thus, digital game-based trainings are essential tools to support literacy acquisition in the home environment.

8.4 Validity

After demonstrating the feasibility of Prosodiya in the home environment and its positive training effect on syllable stress awareness and spelling skills, I address the validity of Prosodiya's pedagogical motivation in the following sections. Proving the validity of an educational training is as important as proving its effectiveness. This is because training effects could result from mere exposure of training material or practice of common skills, such as spelling, without a theoretically sound foundation of the teaching approach and its implementation. Accordingly, I investigate the relationship between literacy skills as well as the extent to which the exercises implemented in the training relate to real-life challenges of children with poor spelling and reading skills. The validity of Prosodiya is reported in our articles "Validity and Player Experience of a Mobile Game for German Dyslexic Children" (Holz, Ninaus, et al., 2018), which is based on the data of the immediate treatment group, and "A Digital Game-Based Training Improves Spelling in German Primary School Children – A Randomized Controlled Field Trial" (Holz et al., unpublished), which is based on the data of both treatment groups. This section is mainly based on these articles.

8.4.1 Results and Discussion

We investigated the validity of Prosodiya by means of correlation analyses. We computed partial correlations in order to determine the relationship between the assessed literacy skills and between literacy skills and training performances, while controlling for sex and grade. We included the data of children who participated at T3 and completed at least two-thirds of the training. We opted for Spearman's rank correlation due to non-normal distribution of the learner data obtained from game logs. The partial correlations are listed in Figure 8.15.

8.4.1.1 Relationship Between Syllable Stress Awareness and Reading and Spelling

We found significant positive correlations between syllable stress awareness and reading. Particularly, we found that syllable stress awareness significantly correlated with reading fluency, $r_s = .31,\ p < .001$, word reading, $r_s = .37,\ p < .001$, and pseudoword reading, $r_s = .35,\ p < .001$. Moreover, syllable stress awareness was significantly correlated with the spelling score of the standardized spelling test, $r_s = .48,\ p < .001$, with the more specific spelling score of our self-designed spelling test, $r_s = .51,\ p < .001$, as well as with mistakes in vowel length marking, $r_s = -.50,\ p < .001$.

These correlations of moderate effect sizes concur with the current state of research on the role of stress awareness in the development of reading and spelling, which is specifically impaired in children with poor reading and spelling skills (Goswami, Gerson, & Astruc, 2010; Goswami et al., 2013; Jiménez-Fernández et al., 2015; Leong et

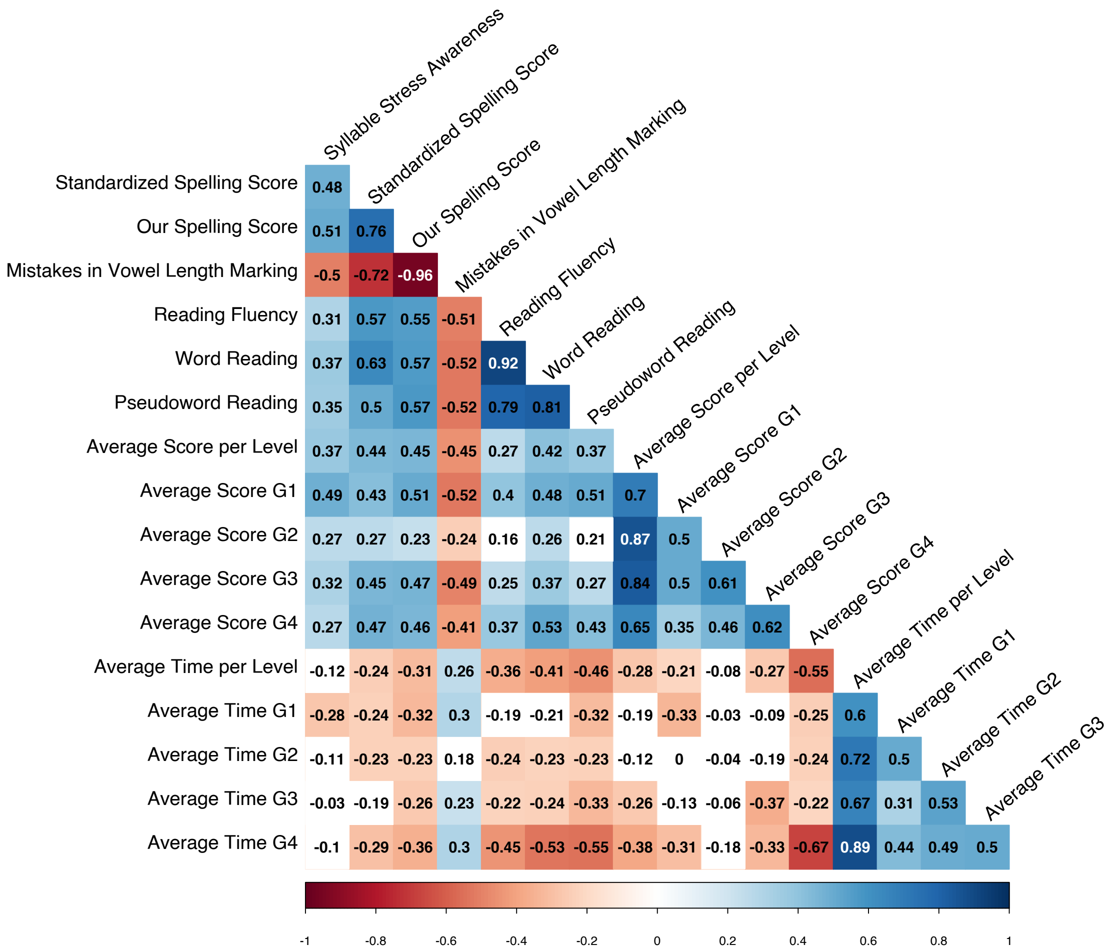


Figure 8.15: Correlations between literacy skills assessed at T3 (syllable stress awareness, standardized spelling score, our spelling score, words with incorrect vowel length marking, reading fluency, and (pseudo-) word reading) and average in-game scores and times per game type (G1 = "Stress pattern"; G2 = "Open and closed syllables"; G3 = "Orthographic markers"; G4 = "Spelling"). Correlations significant on $\alpha = .05$ are colored.

al., 2011; Sauter et al., 2012; Weber, Hahne, Friedrich, & Friederici, 2004). Accordingly, the results further validate the approach to improve literacy skills by focusing on syllable stress awareness and linking the linguistic features of the stressed syllable to orthographic regularities, in particular to vowel length marking.

8.4.1.2 Relationship Between Assessed Literacy Skills and In-Game Performances

Moreover, we found that spelling, reading, and syllable stress awareness were significantly correlated with the overall average score achieved in-game as well as the average score achieved in levels of individual game types, see Figure 8.15. The standardized spelling score significantly correlated positively with all in-game scores, particularly a moderate positive correlation with the average score achieved in its in-game counterpart “Spelling” (G4) was found, $r_s = .47,\ p < .001$. The spelling score of our self-designed spelling test, which addresses the educational content of Prosodiya, correlated even more strongly with the average score achieved per level, $r_s = .45,\ p < .001$, as well as with average score of the game “Stress pattern” (G1), $r_s = .51,\ p < .001$, and with the average score of the game “Orthographic markers” (G3), $r_s = .47,\ p < .001$. The correlations of the mistakes in vowel length marking are inverted but strikingly similar to the total score of our spelling test.

Syllable stress awareness and all in-game scores were also significantly correlated positively, particularly a moderate positive correlation between syllable stress awareness and the average score of its in-game counterpart “Stress pattern” (G1) was found, $r_s = .49,\ p < .001$.

Reading fluency correlated significantly with the overall average in-game score as well as with the average score of all game types except for the game “Open and closed syllables” (G2), whereas reading fluency most strongly correlated with the average score of the game “Stress pattern” (G1), $r_s = .40,\ p < .001$, and with the average score of the game “Spelling” (G4), $r_s = .37,\ p < .001$.

Word reading also correlated significantly with all in-game scores, particularly with the average score of the game “Stress pattern” (G1), $r_s = .48,\ p < .001$, with the average score of the game “Orthographic markers”, $r_s = .37,\ p < .001$, and with the average score of the game “Spelling” (G4), $r_s = .53,\ p < .001$

Furthermore, the relationship found among these literacy skills is also present in the correlations between the performance metrics of the individual game-based exercises of the training, speaking for the implementation of the game-based exercises.

The indications of the correlations between literacy skills and in-game performances are twofold. First, they provide support for the validity of the implementation of the game’s pedagogical approach. In particular, the results indicate that the game addresses the difficulties of children with poor literacy skills and that we successfully embedded and implemented these literacy skills in the game-based exercises of Prosodiya. This applies to syllable stress awareness, to the general reading and spelling abilities assessed by standardized tests, as well as to the more specific spelling categories included in our self-designed spelling test, particularly the orthographic marking of long and short vowels. Second, the results are in line with previous research providing evidence that in-game measures such as times (e.g., Sense, Behrens, Meijer, & van Rijn, 2016) and scoring (e.g., Ninaus, Kiili, et al., 2017) may allow for valid assessment of skills and knowledge.

In sum, the correlation analyses confirm the theoretic background of our teaching approach to focus on linguistic characteristics of syllable stress to improve reading and spelling, particularly vowel length marking.

8.4.2 Summary

Summary of Prosodiya's validity

Relationship between syllable stress awareness and reading and spelling

- Correlation analyses revealed moderate associations between syllable stress awareness and reading and spelling skills.
- The results confirm recent empirical findings that stress perception is impaired in children with poor reading and/or spelling skills and support our pedagogical approach to improve literacy skills by training syllable stress awareness and shifting the attention to the stressed syllable to teach related spelling rules.

Relationship between literacy skills and in-game performances

- Correlation analyses further revealed significant associations between these literacy skills (syllable stress awareness, reading, spelling) and training performances computed from game logs.
- The results show that Prosodiya addresses the difficulties of children with poor literacy skills and that we successfully embedded and implemented these literacy skills in the game-based exercises of Prosodiya.

8.5 Summary

In this chapter, I presented the evaluation of our digital game-based spelling training for German primary school children with regard to feasibility, validity, and effectiveness. To recap, Prosodiya is the first digital intervention that focuses on training syllable stress awareness and linking the linguistic features of the stressed syllable to orthographic regularities of German orthography, primarily the marking of long and short vowels. The development and evaluation of Prosodiya are the major contributions of this thesis.

The evaluation was carried out with 116 German primary school children from second to fourth grades (aged 7–11 years) in a randomized controlled field trial with a two-period, wait-list controlled crossover treatment design. During respective training periods of 9–10 weeks, children from the active training group practiced at home with Prosodiya on Android tablets. The evaluation was guided by three major research questions on the feasibility of the spelling training, i.e., the appropriateness of the digital training in the home environment, the validity of its pedagogical approach and, most importantly, its training effect on literacy skills.

Feasibility. Feasibility was addressed by evaluating training behavior and training experience questionnaires. Prosodiya was reported to be easy to use and children spent an average 10 hours with the game. 76% of the children completed the training. Moreover, the training seemed to engage and motivate children throughout the training period, which is confirmed in that children would likely continue the training or recommend it to friends. Further, the children reported positive training experiences and liked the individual game elements, in particular the Kugellichter. Children also reported to perceive high self-efficacy for spelling-related domains and a moderate impact on reading. However, children also reported that challenge and flow of the training can be improved. In sum, the training behavior and positive feedback indicates the feasibility of our digital game-based training for the use in the home environment.

Effectiveness. We found significant training effects of medium to large effect sizes on children's syllable stress awareness and spelling skills. Although the training had no significant impact on untrained reading skills, i.e., reading fluency and (pseudo-) word reading, we found first indications that children's reading fluency may benefit from the training. As for spelling, we found that the training had a significant positive effect on the general spelling ability, on the orthographic marking of long and short vowels, as well as on untrained spelling categories of near (uninflected untrained words) and far transfer learning (inflected training words). Moreover, the improvements were maintained even months after the training, indicating long-term effectiveness. The average training-induced improvement in spelling ranged between 4.0 and 4.5 T-scores. The learning gains are comparable to empirically evaluated computer-based and therapeutic interventions, proving digital game-based trainings as essential tools to support literacy acquisition in the home environment. Further, our innovative approach yields results comparable to traditional training methods. The approach to systematically teach orthographic knowledge in combination with the awareness of syllable stress seems to be equally beneficial and might therefore expand the traditional pool of training methods.

Validity. Regarding the validity of our training, we found moderate positive correlations between syllable stress awareness and reading and spelling skills. This confirms recent empirical findings and supports our pedagogical approach to improve literacy skills by training syllable stress awareness and shifting the attention to the stressed syllable to teach related spelling rules. Moreover, we found significant associations between literacy skills (syllable stress awareness, reading fluency, and spelling skills) and training performances obtained from game logs. Most interestingly, we found moderate correlations between syllable stress awareness and its in-game counterpart "Stress pattern", $r_s = .49$, between reading skills and the average score of the game "Stress pattern", $r_s = [.40,\ .51]$, and between spelling skills and the average score of the games "Stress pattern", "Orthographic markers", and "Spelling", whose correlation coefficients ranged between $r_s = [.41,\ .52]$. Hence, we may conclude that the pedagogical content implemented in our digital game-based training deals with real challenges of children with poor literacy skills and is tailored to the improvement of spelling abilities of poor spellers.

To summarize, the digital game-based training Prosodiya was found to be engaging and feasible for the use in the home environment. Furthermore, the training evidently improves syllable stress awareness and spelling abilities, proving its educational effectiveness. Lastly, the validity of our pedagogical approach to teach orthographic knowledge in combination with the awareness of syllable stress could be further confirmed in correlation analyses. Thus, our findings demonstrate the benefits of digital game-based training on children's literacy acquisition.

The summary box for this chapter can be found on the next page.

Summary of the empirical evaluation of Prosodiya

In this chapter, I assessed the feasibility, validity, and effectiveness of the digital game-based spelling training "Prosodiya". For this, we evaluated the training program in a randomized controlled field trial with 116 German primary school children (age 7–11). We used a two-period, wait-list controlled crossover treatment design in which children from the immediate treatment group ($N = 58$) received Prosodiya during the first training period and the delayed treatment group ($N = 58$) during the second, while the treatment groups served as the control in the opposite training periods. While in the active training condition, children practiced with Prosodiya at home for 9–10 weeks.

Feasibility and training experience

- Prosodiya can easily be used unassisted by primary school children in the home environment.
- Children enjoyed the training and reported to perceive high positive influence on syllable stress awareness and spelling abilities.
- Children's favorite game elements are the Kugellichter.

Effectiveness

- Prosodiya significantly improves children's syllable stress awareness.
- Prosodiya also significantly improves children's spelling skills. We found significant training effects on general spelling ability, on the orthographic marking of long and short vowels, as well as on untrained spelling categories of near (uninflected untrained words) and far transfer learning (inflected training words).
- Children maintained their improved spelling levels even months after the training, indicating long-term effectiveness.
- No significant training effects were found on untrained reading abilities. However, we found first indications that Prosodiya meliorates children's reading fluency.

Validity

- Correlation analyses revealed moderate associations between syllable stress awareness and reading and spelling skills, demonstrating the validity of our pedagogical approach.
- Correlation analyses further revealed significant associations between these literacy skills and training performances computed from game logs, speaking for the implementation of individual game-based exercises.

→ Prosodiya evidently improves syllable stress awareness and spelling skills in primary school children and has proven its feasibility as a digital game-based training for the home environment.

Part IV

Conclusion and Implications

Chapter 9

Conclusion and Implications

Digital tools offer many possibilities to enhance different processes in language learning and teaching. In particular, children with special learning needs, such as reading and spelling disorders, can benefit greatly from the use of state-of-the-art technology. The aim of the present thesis was to design, develop, and evaluate such educational tools to support reading and spelling development of primary school children. For this, I worked on three aspects of different granularity to enhance children's experience in language learning: (1) the development and evaluation of a digital game-based spelling training to improve children's literacy skills in the home environment, (2) the systematic comparison of touch interaction styles for use with children, and (3) exploring the use of state-of-the art technology to automatically generate and enhance language learning material.

The most extensive contribution of this thesis is the development and evaluation of a digital game-based spelling training. Digital training programs have been shown to effectively improve children's reading and spelling abilities and are recognized treatment approaches for children with reading or spelling disorders. However, I have argued in this thesis for the need for empirical evaluations of spelling trainings carried out in the home environment: The efficacies of evidence-based digital reading and spelling trainings for German primary school children were investigated almost exclusively in studies where the training was carried out completely or partially in supervised training sessions. Yet, this leaves unanswered the question of whether the training also effectively improves literacy skills when children practice independently in the home environment. Moreover, the range of evidence-based treatment approaches implemented as digital game-based training programs that can be used unassisted is limited. In response to this paucity, I assessed whether Prosodiya – our novel digital game-based and independently performed spelling training – can effectively improve spelling in German primary school children with below average spelling skills. Besides its playful approach, Prosodiya differs from similar concepts in that it focuses on improving syllable stress and on teaching children the associations between syllable stress and orthographic marking of long and short vowels, finally leading to a rule-based orthographic spelling training.

The results of a randomized controlled field trial with 116 German primary school children show that our digital game-based training carried out independently at home improves syllable stress awareness and spelling skills. Moreover, the training seems to have a long-term effect on improved literacy skills. Importantly, the improvement in spelling was found in trained and untrained word material, and the training seems to induce transfer effects on non-explicitly trained spelling skills, such as the spelling of inflected training words. Moreover, by integrating innovative learning and game features, such as interactive instructions, immediate feedback, adaptivity, and rewarding, the training program does not need professional or parental support and keeps children engaged and motivated over several months. Additionally, significant associations found between syllable stress awareness, reading, spelling, and training performances validate or novel approach to systematically teach spelling rules in combination with syllable stress awareness. Our findings concur with recent research that digital game-based trainings carried out unassisted at home can meliorate children's literacy acquisition (Görgen et al., 2020). Thus, the training program can be particularly useful for children who don't have access to or are waiting for special spelling support. Further, the training can be used in addition to class and learning therapy to increase frequency of support. To improve other literacy skills, such as reading fluency, word reading, or the spelling of plosive sounds, we must complement the training by specifically developed modules.

The evaluation of Prosodiya has two implications for the practical application in reading and spelling promotion. On one hand, the results indicate that the approach to systematically teach orthographic knowledge in combination with the awareness of syllable stress yields results comparable to traditional training methods. On the other hand, digital game-based trainings seem to facilitate enjoyable learning experiences in children with reading or spelling disorders and can be used in addition to class or therapy. In summary, Prosodiya might therefore expand the pool of evidence-based training methods.

The main contribution of this thesis addressed an entire learning system. However, the development process of such systems involves many design decisions that can influence the experience with and the effectiveness of a system. For the second contribution of this thesis, I addressed one particular design decision, namely whether to implement a drag-and-drop, point-and-click, or click interaction style in educational applications to move and interact with objects. As I have argued in this thesis, the decision is not trivial and has been shown to impact performance, usability, and educational effectiveness. Thus, this also affects the implementation of game mechanics in digital trainings such as Prosodiya. For example, children reported that the spelling game is their favorite game of Prosodiya. However, they also stated that the drag-and-drop interaction style to move the letters, which was implemented in the study version of Prosodiya, is sometimes exhausting, particularly when they practiced multiple levels consecutively. Choosing the most appropriate interaction style could therefore further enhance children's learning experience. However, the current state of research provides conflicting results and recommendations as to this choice. For example, point-and-click is said to be more appropriate than drag-and-drop – and vice versa. This contradictory state presumably arises from the fact that the studies differed in a range of aspects, such as input modality, type of task, as well as the ages of children and their typical interaction experience with technology. Importantly, as technology emerges, conclusions from studies with different contexts must be drawn with caution. Thus, without further research, it was not possible to make an evidence-based decision about the appropriateness of different interaction styles to maximize children's playing experience.

In response to this, we systematically compared the drag-and-drop, point-and-touch, and touch interaction styles in a touch-based spelling game with regard to subjectively perceived workload, user experience, and writing time. For this, we conducted a within-subject lab experiment with 25 German primary school children aged 8–11 years. Our results showed that the touch interaction style outperformed drag-and-drop and point-and-touch in writing time and in most workload and user experience subscales. Further, most children selected touch as their interaction style of choice. Interestingly, our findings showed small advantages of drag-and-drop over point-and-touch in reported mental demand and direct user rankings, contradicting earlier results and design recommendations concluding that point-and-click or point-and-touch is more appropriate. To conclude, while the touch interaction style seems to be most appropriate in a touch-based spelling game, drag-and-drop seems to be unproblematic or even more appropriate than point-and-touch when used with touch-based devices. This study has two implications for practical application. First, the results suggest that we should change to a touch or hybrid interaction style in the spelling game of Prosodiya to further enhance the training experience. Further, the literature review and result of our study can help other designers and developers to choose appropriate game mechanics in the development process. We advice researchers and designers to select the interaction style carefully with regard to age group, input modality, and context – and to not rely on results of studies that did not address the requirements of their applications.

The third contribution of this thesis addressed another aspect of language learning systems, namely content creation and input enhancement. The provision of age- and skill-appropriate content is costly and thus limits the availability of learning material. I explored in two studies the use of state-of-the-art technology to automatically generate speech and visually enhance reading material to facilitate content provision. In particular, I presented our novel method to improve the pronunciation of German pseudowords generated artificially with text-to-speech (TTS) tools to reduce the cost and effort of providing large audio-dictionaries for digital minimal pair therapy. Training the distinction of vowel lengths or learning to differentiate between voiced and voiceless plosive sounds in form of minimal pair therapy (e.g., *beaver* vs. *peaver*) is one recognized treatment fostering phonological awareness of children with language disorders. For example, Prosodiya uses minimal pairs consisting of a lexical word and its pseudoword counterpart to teach the distinction of long and short vowels to improve spelling. As I have argued, TTS tools pronounce lexical words sufficiently well, but most of the tools perform poorly in the pronunciation of German pseudowords when plaintext is used as input. In response to this, I presented our novel method that consistently modifies the input of TTS tools using X-SAMPA and SSML in order to improve the pronunciation of synthetically generated German pseudowords. Our novel approach was evaluated in a crowdsourcing experiment. The results showed that distinguishing the lexical word from its pseudoword counterpart was equally successful when the minimal pair was generated by our method or produced by a human. This study has one practical implication for language learning systems. While TTS systems may perform poorly in the pronunciation of German pseudowords when the word is entered in plaintext, a consistent modification of the input into TTS tools seems promising as to the provision of synthetically generated pseudowords for the language learning context. Thus, the use of TTS tools seems promising to extend audio material in language learning systems such as Prosodiya.

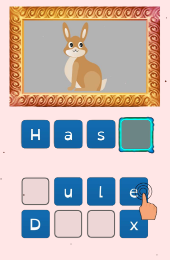
Lastly, I addressed the automatic visual input enhancement of reading material. Visually enhanced texts with custom spacing and syllables alternately displayed in different colors are commonly used in teaching and learning therapy to support literacy acquisition. This so-called *Silbenmethode* [Syllable Method] teaches children to focus and understand syllables and their structures rather than single characters, and is commonly used in Germany. However, as I have argued in this thesis, the range of applications that provide automatic visual input enhancement in order to provide age- and skill-appropriate reading material are limited and lack full customization. In response, I presented COAST, our web-based tool to automatically enhance reading material. COAST offers a high degree of customization for text enhancement (i.e., syllable structure, word stress, and spacing) and supports management of annotation schemes. COAST was successfully tested with teaching practitioners in user tests validating the concept of our system. This study has one practical implication. Tools such as COAST, using state-of-the-art web technology and resources of natural language processing, can facilitate the provision of internally differentiated reading material and thus support teaching staff and parents.

To conclude, this thesis builds a unique bridge between scientific evidence and practical application in supporting children's reading and spelling development. As the results of this thesis suggest, the promotion of reading and spelling development of children can be enhanced on different levels with digital learning tools. First and foremost, I have shown that digital game-based trainings carried out unassisted at home improve literacy skills in German primary school children and provide enjoyable training experience. Further, I have shown that the design of individual game mechanics in such training programs can impact learning experiences. Lastly, I have shown that web technologies, natural language processing, and text-to-speech systems are promising as to automatically generate and enhance language learning material in order to facilitate the provision of age- and skill-appropriate content.

Take-Home Messages

1. **Digital game-based trainings carried out unassisted at home improve literacy skills in primary school children**
   - The novel digital game-based spelling training "Prosodiya" significantly improves syllable stress awareness and spelling skills in German primary school children.
   - The training program can be used unassisted at home and offers enjoyable training experience.

2. **The touch interaction style outperforms drag-and-drop and point-and-touch in a touch-based spelling game for children**
   - Constructing interaction design in educational applications for children is challenging and should consider a range of factors, such as type of task and children's ages and previous interaction experiences with technology.
   - For a touch-based spelling game for children, the touch interaction style – compared to drag-and-drop and point-and-touch – seems to be most appropriate with regard to mental workload, user experience, and writing time.



3. **Automatic generation and enhancement of language learning material can facilitate content provision**
   - Text-to-speech (TTS) tools seem promising to extend audio dictionaries of language learning systems: the consistent modification of the input of TTS tools improves the pronunciation of artificially generated German pseudowords for use in minimal pair distinction tasks.
   - Visual input enhancement: COAST, a web application using resources of natural language processing, supports automatic and customizable visual input enhancement of texts to facilitate the provision of age- and skill-appropriate reading material.

Ich be rat schlag te
mein Mei ster werk
mit ei nem
E le fan ten und ei ner
Rie sen schlan ge.

Bibliography

Abrami, P., Borohkovski, E., & Lysenko, L. (2015). The effects of ABRACADABRA on reading outcomes: A meta-analysis of applied field research. *Journal of Interactive Learning Research*, *26*(4), 337–367.

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, *37*(4), 445–456. https://doi.org/10.3233/WOR-2010-1109

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 199–213). https://doi.org/10.1007/978-1-4612-1694-0_15

Amazon Webservices, Inc. (2018). Amazon Polly [Computer software]. https://aws.amazon.com/de/polly/

Antle, A. N. (2013). Research opportunities: Embodied child–computer interaction. *International Journal of Child-Computer Interaction*, *1*(1), 30–36. https://doi.org/10.1016/j.ijcci.2012.08.001

Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., de Freitas, S., Louchart, S., ... De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, *46*(2), 391–411. https://doi.org/10.1111/bjet.12113

Aro, M. (2004). *Learning to read: the effect of orthography* (Unpublished doctoral dissertation).

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1-2), 46–63. https://doi.org/10.1159/000208930

Azah, N., Syuhada, N., Batmaz, F., Stone, R., & Wai, P. (2014). Selection of touch gestures for children's applications: Repeated experiment to increase reliability. *International Journal of Advanced Computer Science and Applications*, *5*(4), 97–102. https://doi.org/10.14569/IJACSA.2014.050415

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). CELEX2 LDC96L14. *Web Download. Philadelphia: Linguistic Data Consortium*.

Barendregt, W. (2015). Determining the most appropriate interaction style for children: Taking spontaneous use into account. https://doi.org/10.13140/RG.2.1.2463.2163

Barendregt, W., & Bekker, M. M. (2011). Children may expect drag-and-drop instead of point-and-click. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (p. 1297–1302). https://doi.org/10.1145/1979742.1979764

Barlow, J. A., & Gierut, J. A. (2002). Minimal pair approaches to phonological remediation. *Seminars in Speech and Language*, *23*(1), 57–68. https://doi.org/10.1055/s-2002-24969

Baschera, G.-M. (2011). *Modeling and Evaluation of Computer-Assisted Spelling Learning in Dyslexic Children* (Doctoral dissertation, ETH Zurich). https://doi.org/10.3929/ethz-a-007086954

Beddington, J., Cooper, C. L., Field, J., Goswami, U., Huppert, F. A., Jenkins, R., … Thomas, S. M. (2008). The mental wealth of nations. *Nature*, *455*(7216), 1057–1060. https://doi.org/10.1038/4551057a

Bender, F., Brandelik, K., Jeske, K., Lipka, M., Löffler, C., Mannhaupt, G., … von Orloff, M. (2017). Die integrative Lerntherapie. *Lernen und Lernstörungen*, *6*(2), 65–73. https://doi.org/10.1024/2235-0977/a000167

Berkling, K. (2017). Phontasia: A phonics game for German and its effect on orthographic skills–first corpus explorations. In *Proc. WOCCI 2017: 6th International Workshop on Child Computer Interaction* (pp. 7–16). https://doi.org/10.21437/WOCCI.2017-2

Berkling, K., Faller, H., & Piertzik, M. (2017). Avoiding failure in modern game design with academic content - a recipe, an anti-pattern and applications thereof. In *Proceedings of the 9th International Conference on Computer Supported Education* (pp. 25–36). https://doi.org/10.5220/0006281800250036

Berkling, K., Pflaumer, N., & Lavalley, R. (2015). German phonics game using speech synthesis - A longitudinal study about the effect on orthography skills. In *SLaTE* (pp. 167–172).

Berninger, V. W., Nagy, W., Tanimoto, S., Thompson, R., & Abbott, R. D. (2015). Computer instruction in handwriting, spelling, and composing for students with specific learning disabilities in grades 4–9. *Computers & Education*, *81*, 154–168. https://doi.org/10.1016/j.compedu.2014.10.005

Bhide, A., Power, A., & Goswami, U. (2013). A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind, Brain, and Education*, *7*(2), 113–123. https://doi.org/10.1111/mbe.12016

Boyer, N., & Ehri, L. C. (2011). Contribution of phonemic segmentation instruction with letters and articulation pictures to word reading and spelling in beginners. *Scientific Studies of Reading*, *15*(5), 440–470. https://doi.org/10.1080/10888438.2010.520778

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., … Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, *94*(Supplement C), 178–192. https://doi.org/10.1016/j.compedu.2015.11.003

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read – a causal connection. *Nature*, *301*(3), 419–421. https://doi.org/10.1038/301419a0

Brainy Ape Studio, LLP. (2017). Kinder lernen deutsche Wörter - lesen und schreiben [Children Learn German Words – Reading and Spelling] [Mobile application software]. https://play.google.com/store/apps/details?id=com.puzzlingapps.spelling.de.free

Bußmann, H. (2008). *Lexikon der Sprachwissenschaft* [Encyclopedia of linguistics]. Stuttgart: Kröner.

Chiasson, S., & Gutwin, C. (2005). Design principles for children's technology. *interfaces*, *7*(28), 1–9. http://www.hci.usask.ca/publications/2005/HCI_TR_2005_02_Design.pdf

Cidrim, L., & Madeiro, F. (2017). Information and Communication Technology (ICT) applied to dyslexia: literature review. *Revista CEFAC*, *19*(1), 99–108. https://doi.org/10.1590/1982-021620171917916

Cleary, T. J., Velardi, B., & Schnaidman, B. (2017). Effects of the Self-Regulation Empowerment Program (SREP) on middle school students' strategic skills, self-efficacy, and mathematics achievement. *Journal of School Psychology*, *64*(Supplement C), 28–42. https://doi.org/10.1016/j.jsp.2017.04.004

Corbalan, G., Kester, L., & J.G. van Merriënboer, J. (2009). Dynamic task selection: Effects of feedback and learner control on efficiency and motivation. *Learning and Instruction*, *19*(6), 455–465. https://doi.org/10.1016/j.learninstruc.2008.07.002

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, *88*(4), 715–730. https://doi.org/10.1037/0022-0663.88.4.715

Creative-Solutions. (2018). GSpeech [Computer software]. https://wordpress.org/plugins/gspeech/

Crystal, D. (2011). *A Dictionary of Linguistics and Phonetics.* John Wiley & Sons.

Daniel, S. S., Walsh, A. K., Goldston, D. B., Arnold, E. M., Reboussin, B. A., & Wood, F. B. (2006). Suicidality, school dropout, and reading problems among adolescents. *Journal of Learning Disabilities*, *39*(6), 507–514. https://doi.org/10.1177/00222194060390060301

Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality*, *57*, 119–130. https://doi.org/10.1016/j.jrp.2015.05.004

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15). https://doi.org/10.1145/2181037.2181040

Deunk, M. I., Smale-Jacobse, A. E., de Boer, H., Doolaard, S., & Bosker, R. J. (2018). Effective differentiation practices: A systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education. *Educational Research Review*, *24*, 31–54. https://doi.org/https://doi.org/10.1016/j.edurev.2018.02.002

DirectINNOVATION, UG. (2018). MWS Reader 5 [Computer software]. https://www.mwsreader.com/en/

Donker, A., & Reitsma, P. (2007a). Drag-and-drop errors in young children's use of the mouse. *Interacting with Computers*, *19*(2), 257–266. https://doi.org/10.1016/j.intcom.2006.05.008

Donker, A., & Reitsma, P. (2007b). Young children's ability to use a computer mouse. *Computers & Education*, *48*(4), 602–617. https://doi.org/10.1016/j.compedu.2005.05.001

Drigas, A., & Batziaka, E. (2016). Dyslexia and ICTs, assessment and early intervention in Kindergarten. *International Journal of Emerging Technologies in Learning (iJET)*, *11*(02), 53. https://doi.org/10.3991/ijet.v11i02.5193

Druin, A. (1999). Cooperative inquiry: Developing new technologies for children with children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 592–599). https://doi.org/10.1145/302979.303166

Druin, A. (2002). The role of children in the design of new technology. *Behaviour & Information Technology*, *21*(1), 1–25. https://doi.org/10.1080/01449290110108659

Dummer-Smoch, L., & Hackethal, R. (2011). *Kieler Leseaufbau* [Kiel Reading

Course] (Vol. 8). Veris.

Durkin, K., Boyle, J., Hunter, S., & Conti-Ramsden, G. (2013). Video games for children and adolescents with special educational needs. *Zeitschrift für Psychologie*, *221*(2), 79–89. https://doi.org/10.1027/2151-2604/a000138

Eisenberg, P. (1998). Der Laut und die Lautstruktur des Wortes [The sound and the sound structure of the word]. In G. Drosdowski, P. Eisenberg, & G. Bibliographisches Institut (Mannheim (Eds.), *Duden, Grammatik der deutschen Gegenwartssprache* (4th ed.). Mannheim: Dudenverlag.

Eisenberg, P. (2013). *Grundriss der deutschen Grammatik. Band 1: Das Wort* [Outline of German grammar. Volume 1: The word] (4th ed.; N. Fuhrhop, Ed.). Stuttgart: J.B. Metzler. https://doi.org/10.1007/978-3-476-00757-5

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, *67*(September), 156–167. https://doi.org/10.1016/j.compedu.2013.02.019

Ernst Klett Verlag. (2015). Lesen lernen 1 mit Zebra [Learning to read 1 with zebra] [Mobile application software]. https://www.klett.de/produkt/isbn/ECN00001APA99

Evans, G. W., Brauchle, G., Haq, A., Stecker, R., Wong, K., & Shapiro, E. (2007). Young children's environmental attitudes and behaviors. *Environment and Behavior*, *39*(5), 635–658. https://doi.org/10.1177/0013916506294252

Falloon, G. (2013). Young students using ipads: App design and content influences on their learning pathways. *Computers & Education*, *68*, 505–521. https://doi.org/10.1016/j.compedu.2013.06.006

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in german- and english-speaking children. *Scientific Studies of Reading*, *2*(1), 31–54. https://doi.org/10.1207/s1532799xssr0201_2

Fullerton, T. (2008). *Game Design Workshop: A Playcentric Approach to Creating Innovative Games* (2nd ed.). Heidelberg: Morgan Kaufmann. https://books.google.de/books?id=OjIYWtqWxtAC

Gabe Zichermann and Christopher Cunningham. (2011). *Gamification by Design-Implementing Game Mechanics in Web and Mobile Apps* (Vol. 27; Intergovernmental Panel on Climate Change, Ed.) (No. 3). Cambridge: Cambridge University Press. http://ebooks.cambridge.org/ref/id/CBO9781107415324A009

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS ONE*, *9*(2), e89900. https://doi.org/10.1371/journal.pone.0089900

Galuschka, K., & Schulte-Körne, G. (2016). Clinical practice guideline: The diagnosis and treatment of reading and/or spelling disorders in children and adolescents. *Dtsch Arztebl Int.*, *113*(16), 279–86. https://doi.org/10.3238/arztebl.2016.0279

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment*, *1*(1), 20. https://doi.org/10.1145/950566.950595

Gelderblom, H., & Kotzé, P. (2009). Ten design lessons from the literature on child development and children's use of technology. In *Proceedings of the 8th International Conference on Interaction Design and Children - IDC '09* (p. 52). https://doi.org/10.1145/1551788.1551798

Girard, S., & Johnson, H. (2009). Educational software for children: analysis of interaction techniques for direct manipulation. In (pp. 259–262). https://doi.org/10.1145/1629826.1629868

Goldberg, B., & Cannon-Bowers, J. (2015). Feedback source modality effects on training outcomes in a serious game: Pedagogical agents make a difference. *Computers in Human Behavior*, *52*(1-2), 1–11. https://doi.org/10.1016/j.chb.2015.05.008

Gonzalez, A. (2014). *OpenDyslexic.* https://www.opendyslexic.org/

Görgen, R., Huemer, S., Schulte-Körne, G., & Moll, K. (2020). Evaluation of a digital game-based reading training for german children with reading disorder. *Computers & Education*, *150*, 103834. https://doi.org/10.1016/j.compedu.2020.103834

Goswami, U. (1999). Integrating orthographic and phonological knowledge as reading develops: Onsets, rimes and analogies in children's reading. *Converging methods for understanding reading and dyslexia*, 57–75.

Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing*, *23*(8), 995–1019. https://doi.org/10.1007/s11145-009-9186-6

Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., & Leong, V. (2013). Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language*, *69*(1), 1–17. https://doi.org/10.1016/j.jml.2013.03.001

Griffiths, M. (2002). The educational benefits of videogames. *Education and Health*, *20*(3), 47–51.

Grissom, R. J., & Kim, J. J. (2012). *Effect Sizes for Research: Univariate and Multivariate Applications* (2nd ed.). https://doi.org/10.4324/9780203803233

Gros, B. (2017). Game dimensions and pedagogical dimension in serious games. In M. K. Zheng, Robert and Gardner (Ed.), *Handbook of Research on Serious Games for Educational Applications* (pp. 402–417). Hershey, PA, USA: IGI Global. https://doi.org/10.4018/978-1-5225-0513-6.ch019

Groth, K., Hasko, S., Bruder, J., Kunze, S., & Schulte-Körne, G. (2013). Interventionseffekte bei Lese-Rechtschreibstörung: Evaluation von zwei Förderkonzepten unter besonderer Betrachtung methodischer Aspekte [Measurement of treatment in developmental dyslexia under methodological considerations]. *Lernen und Lernstörungen*, *2*(3), 161–175. https://doi.org/10.1024/2235-0977/a000038

Grund, M. (). *GUT1. Grundwortschatz 500* [GUT1. Basic vocabulary 500]. [2020-02-13]https://www.gut1.de/grundwortschatz/grundwortschatz-500/

Grund, M., Leonhart, R., & Nauman, C. L. (2017). *Diagnostischer Rechtschreibtest für 4. Klassen (DRT 4)* [Diagnostic Spelling Test for 4th Grade (DRT 4)] (3rd ed.; M. Hasselhorn, W. Schneider, & U. Trautwein, Eds.). Hogrefe Schultests. https://www.testzentrale.de/shop/diagnostischer-rechtschreibtest-fuer-4-klassen.html

Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., & Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. *Computers & Education*, *102*, 202–223. https://doi.org/10.1016/j.compedu.2016.09.001

Hall, L., Hume, C., & Tazzyman, S. (2016). Five degrees of happiness: Effective smiley face likert scales for evaluating with children. In *Proceedings of the The 15th International Conference on Interaction Design and Children - IDC '16* (pp. 311–321). https://doi.org/10.1145/2930674.2930719

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? – a literature review of empirical studies on gamification. In *2014 47th Hawaii International Conference on System Sciences* (pp. 3025–3034). IEEE. https://doi.org/10.1109/HICSS.2014.377

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, *54*, 170–179. https://doi.org/10.1016/j.chb.2015.07.045

Hamza, Z., & Salivia, G. (2015). Study of touch gestures performance in touch devices by young children. *International Journal on Recent and Innovation Trends in Computing and Communication*, *3*(3), 1395–1400. https://doi.org/10.17762/ijritcc2321-8169.1503107

Handley, Z. (2009). Is text-to-speech synthesis ready for use in computer-assisted language learning? *Speech Communication*, *51*(10), 906–919.

https://doi.org/10.1016/j.specom.2008.12.004

Handt, R., Kuhn, K., & Mrowka-Nienstedt, K. (2019). *ABC der Tiere 1 – Lesen in Silben. Die Silbenfibel* [Animal alphabet 1 – Reading in syllables. The fibula of syllables] (K. Kuhn, Ed.). Mildenberger Verlag. https://www.thalia.de/shop/home/artikeldetails/ID19676097.html?retn=thdedvbu:devi

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (Vol. 52, pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Hattie, J. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement.* Routledge. https://doi.org/10.4324/9780203887332

Hills, M., & Armitage, P. (1979). The two-period cross-over clinical trial. *British Journal of Clinical Pharmacology*, *8*(1), 7–20. https://doi.org/10.1111/j.1365-2125.1979.tb05903.x

Holmes, W. (2011). Using game-based learning to support struggling readers at home. *Learning, Media and Technology*, *36*(1), 5–19. https://doi.org/10.1080/17439884.2010.531023

Holz, H., Beuttler, B., & Kirsch, A. (2017). Bewegungserkennung mit Wearables für Embodied Trainings in Serious Games [Movement recognition with wearables for embodied trainings in serious games]. In M. Burghardt, R. Wimmer, C. Wolff, & C. Womser-Hacker (Eds.), *Mensch und Computer 2017 - Tagungsband* (pp. 259–262). https://doi.org/10.18420/muc2017-mci-0368

Holz, H., Beuttler, B., & Ninaus, M. (2018). Design rationales of a mobile game-based intervention for german dyslexic children. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 205–219). https://doi.org/10.1145/3270316.3272053

Holz, H., Brandelik, K., Beuttler, B., Brandelik, J., & Ninaus, M. (2018). How to train your syllable stress awareness – A digital game-based intervention for German dyslexic children. *International Journal of Serious Games*, *5*(3), 37–59. https://doi.org/10.17083/ijsg.v5i3.242

Holz, H., Brandelik, K., Brandelik, J., Beuttler, B., Kirsch, A., Heller, J., & Meurers, D. (2017). Prosodiya – A Mobile Game for German Dyslexic Children. In J. Dias, P. A. Santos, & R. C. Veltkamp (Eds.), *Games and Learning Alliance. GALA 2017* (pp. 73–82). https://doi.org/10.1007/978-3-319-71940-5_7

Holz, H., Chinkina, M., & Vetter, L. (2018). Optimizing the quality of synthetically generated pseudowords for the task of minimal-pair distinction. In *2018 IEEE Spoken Language Technology Workshop (SLT)* (pp. 470–476). IEEE. https://doi.org/10.1109/SLT.2018.8639037

Holz, H., & Meurers, D. (submitted). Interaction styles in context: Comparing drag-and-drop, point-and-touch, and touch in a mobile spelling game.

Holz, H., Ninaus, M., Beuttler, B., Brandelik, K., & Meurers, D. (unpublished). A digital game-based training improves spelling in german primary school children – a randomized controlled field trial.

Holz, H., Ninaus, M., Meurers, D., & Kirsch, A. (2018). Validity and player experience of a mobile game for German dyslexic children. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 469–478). https://doi.org/10.1145/3270316.3271523

Holz, H., Weiss, Z., Brehm, O., & Meurers, D. (2018). COAST - Customizable online syllable enhancement in texts. a flexible framework for automatically enhancing reading materials. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 89–100). https://doi.org/10.18653/v1/W18-0509

Homer, B. D., Kinzer, C. K., Plass, J. L., Letourneau, S. M., Hoffman, D., Bromley, M., … Kornak, Y. (2014). Moved to learn: The effects of interactivity in a kinect-based literacy game for beginning readers. *Computers & Education*, *74*, 37–49. https://doi.org/10.1016/j.compedu.2014.01.007

Honnibal, M., & Johnson, M. (2015). An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1373–1378). Lisbon, Portugal: Association for Computational Linguistics. https://aclweb.org/anthology/D/D15/D15-1162

Hourcade, J. P. (2008). Interaction design and children. *Human–Computer Interaction*, *1*(4), 277–392. https://doi.org/10.1561/1100000006

Huotari, K., & Hamari, J. (2012). Defining gamification: A service marketing perspective. In *Proceeding of the 16th International Academic MindTrek Conference* (pp. 17–22). https://doi.org/10.1145/2393132.2393137

IBM. (2018). Watson Text to Speech [Computer software]. https://www.ibm.com/cloud/watson-text-to-speech

IJsselsteijn, W., de Kort, Y., & Poels, K. (2013). *The Game Experience Questionnaire*. Eindhoven: Technische Universität Eindhofen. https://pure.tue.nl/ws/files/21666907/Game_Experience_Questionnaire_English.pdf

Inkpen, K. (2001). Drag-and-drop versus point-and-click mouse interaction styles for children. *ACM Transactions on Computer-Human Interaction*, *8*(1), 1–33. https://doi.org/10.1145/371127.371146

Inkpen, K., Booth, K. S., & Klawe, M. (1996). *Drag-and-Drop vs. Point-and-Click Mouse Interaction for Children* (Technical Report No. 96-17). Vancouver, BC, Canada: Department of Computer Science, University of British Columbia.

Ise, E., Engel, R. R., & Schulte-Körne, G. (2012). Effective treatment of dyslexia: A meta-analysis of intervention studies. *Kindheit und Entwicklung*, *21*(2), 122–136. https://doi.org/10.1026/0942-5403/a000077

Ise, E., & Schulte-Körne, G. (2010). Spelling deficits in dyslexia: Evaluation of an orthographic spelling training. *Annals of Dyslexia*, *60*(1), 18–39. https://doi.org/10.1007/s11881-010-0035-8

ISpeech, Inc. (2018). iSpeech [Computer software]. https://www.ispeech.org/

Jessen, M., Marasek, K., Schneider, K., & Claßen, K. (1995). Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German. In *Proceedings of the International Congress of Phonetic Sciences* (Vol. 13, pp. 428–431).

Jiménez, J. E., Hernández-Valle, I., Ramírez, G., Del Rosario Ortiz, M., Rodrigo, M., Estévez, A., ... De La Luz Trabaue, M. (2007). Computer speech-based remediation for reading disabilities: The size of spelling-to-sound unit in a transparent orthography. *Spanish Journal of Psychology*, *10*(1), 52–67. https://doi.org/10.1017/S1138741600006314

Jiménez-Fernández, G., Gutiérrez-Palma, N., & Defior, S. (2015). Impaired stress awareness in Spanish children with developmental dyslexia. *Research in Developmental Disabilities*, *37*, 152–161. https://doi.org/10.1016/j.ridd.2014.11.002

Joiner, R., Messer, D., Light, P., & Littleton, K. (1998). It is best to point for young children: a comparison of children's pointing and dragging. *Computers in Human Behavior*, *14*(3), 513–529. https://doi.org/10.1016/S0747-5632(98)00021-1

Kargl, R., & Purgstaller, C. (2010). *Morphemunterstütztes Grundwortschatz-Segmentierungstraining* [Morpheme-oriented basic vocabulary and segmenting training]. Göttingen: Hogrefe. http://afs-weiterstadt.de/iWeb/AFS_-_Podcasts/BFZ_undMorpheme/Eintrage/2012/6/30_Literatur-Diagnostik-Medien_files/Morpheus-Testzentrale.pdf

Kargl, R., Purgstaller, C., Weiss, S., & Fink, A. (2008). Effektivitätsüberprüfung eines morphemorientierten Grundwortschatz-Segementierungstrainings (MORPHEUS) bei Kindern und Jugendlichen [Effectiveness of a morpheme-oriented basic vocabulary segmenting training (MORPHEUS) for children and adolescents]. *Heilpädagogische Forschung*, *34*(3), 147–156.

Kast, M., Baschera, G.-M., Gross, M., Jäncke, L., & Meyer, M. (2011). Computer-based learning of spelling skills in children with and without dyslexia. *Annals of Dyslexia*, *61*(2), 177–200. https://doi.org/10.1007/s11881-011-0052-2

Katusic, S. K., Colligan, R. C., Barbaresi, W. J., Schaid, D. J., & Jacobsen, S. J. (2001). Incidence of reading disability in a population-based birth cohort, 1976-1982, Rochester, Minn. *Mayo Clinic Proceedings*, *76*(11), 1081–1092. https://doi.org/10.4065/76.11.1081

Kazakou, M. N., & Soulis, S. (2015). Feedback and the speed of answer of pupils with dyslexia in digital activities. In *Procedia Computer Science* (Vol. 67, pp. 204–212). https://doi.org/10.1016/j.procs.2015.09.264

Ke, F., & Abras, T. (2013). Games for engaged learning of middle school children with special learning needs. *British Journal of Educational Technology*, *44*(2), 225–242. https://doi.org/10.1111/j.1467-8535.2012.01326.x

Kerawalla, L., & Crook, C. (2005). From promises to practices: The fate of educational software in the home. *Technology, Pedagogy and Education*, *14*(1), 107–125. https://doi.org/10.1080/14759390500200195

Kiili, K., de Freitas, S., Arnab, S., & Lainema, T. (2012). The Design Principles for Flow Experience in Educational Games. *Procedia Computer Science*, *15*(Supplement C), 78–91. https://doi.org/10.1016/j.procs.2012.10.060

Klatte, M., Bergström, K., Steinbrink, C., Konerding, M., & Lachmann, T. (2018). Effects of the computer-based training program Lautarium on phonological awareness and reading and spelling abilities in German second-graders. In T. Lachmann & T. Weis (Eds.), *Reading and Dyslexia: From Basic Functions to Higher Order Cognition* (pp. 317–333). https://doi.org/10.1007/978-3-319-90805-2_15

Klicpera, C., & Gasteiger Klicpera, B. (2000). Sind Rechtschreibschwierigkeiten Ausdruck einer phonologischen Störung? [Are spelling disorders an indicator of phonological deficits?] *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *32*(3), 134–142. https://doi.org/10.1026//0049-8637.32.3.134

Klicpera, C., Gasteiger-Klicpera, B., & Schabmann, A. (1993). Lesen und Schreiben–Entwicklung und Schwierigkeiten [Reading and spelling – Development and difficulties]. *Bern: Huber*.

Klicpera, C., Schabmann, A., & Gasteiger-Klicpera, B. (2013). *Legasthenie - LRS* [Dyslexia] (4th ed.). München: Reinhardt. http://www.reinhardt-verlag.de/de/titel/52968/Legasthenie_LRS/

Klicpera, C., Weiss, J., & Gasteiger-Klicpera, B. (2013). Erfolg einer schulischen Legastheniker-Förderung für Kinder der 3. und 4. Schulstufe nach dem Programm von Reuter-Liehr [Success of a school support program for dyslexic children of the 3rd and 4th grade based on the training program of Reuter Liehr]. *Heilpädagogische Forschung*, *39*(2), 87–95.

Klische, A. (2007). *Leseschwächen gezielt beheben: Individuelle Diagnose und Ther-*

apie mit dem Programm celeco [Remediation of reading difficulties: Individual diagnosis and therapy with the celeco program]. Tectum-Verl.

Kohler, K. J. (1986). Invariance and variability in speech timing: From utterance to segment in German. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and Variability in Speech Processes* (pp. 268–289). Hillsdale: Erlbaum.

Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior*, *35*, 179–188. https://doi.org/10.1016/j.chb.2014.03.007

Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG Rime and GG Phoneme. *Reading Research Quarterly*, *48*(1), 61–76. https://doi.org/10.1002/rrq.038

Landerl, K. (2003). Categorization of vowel length in german poor spellers: An orthographically relevant phonological distinction. *Applied Psycholinguistics*, *24*(4), 523–538. https://doi.org/10.1017/S0142716403000262

Landerl, K., & Reitsma, P. (2005). Phonological and morphological consistency in the acquisition of vowel duration spelling in Dutch and German. *Journal of Experimental Child Psychology*, *92*(4), 322–344. https://doi.org/10.1016/j.jecp.2005.04.005

Landerl, K., & Thaler, V. (2013). Reading and spelling acquisition and dyslexia in german. In R. M. Joshi & P. Aaron (Eds.), *Handbook of Orthography and Literacy* (pp. 135–148). Routledge.

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A german-english comparison. *Cognition*, *63*(3), 315–334. https://doi.org/https://doi.org/10.1016/S0010-0277(97)00005-X

Laurie-Rose, C., Frey, M., Ennis, A., & Zamary, A. (2014). Measuring perceived mental workload in children. *The American Journal of Psychology*, *127*(1), 107. https://doi.org/10.5406/amerjpsyc.127.1.0107

Lenhard, W., & Lenhard, A. (2016). Evidenzbasierte Förderung schulischer Fertigkeiten am Computer: Lernspiele mit Elfe und Mathis. *Förderprogramme für Vor- und Grundschule*, *2006*, 87–113.

Leong, V., Hämäläinen, J., Soltész, F., & Goswami, U. (2011). Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language*, *64*(1), 59–73. https://doi.org/10.1016/j.jml.2010.09.003

Lim, S., & Reeves, B. (2010). Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player. *International Journal of Human-Computer Studies*, *68*(1-2), 57–68.

https://doi.org/10.1016/j.ijhcs.2009.09.008

Loewe Verlag. (2020). *Leselöwen & Bildermaus* [Reading lion & picture mouse]. https://www.loewe-verlag.de/content-1013-1013/leseloewen/

Mayer, J. (2014). *Linguistische Phonetik* [Linguistic phonetics]. http://praatpfanne.lingphon.net/downloads/LinguistischePhonetik.pdf

Mayringer, H., & Wimmer, H. (2014). *Salzburger Lese-Screening für die Schulstufen 2-9.* [Salzburg reading screening for grades 2–9]. Hogrefe.

McArthur, G., Eve, P. M., Jones, K., Banales, E., Kohnen, S., Anandakumar, T., ... Castles, A. (2012). Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews*(12). https://doi.org/10.1002/14651858.CD009115.pub2

McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, *32*(1), 12. https://doi.org/10.2307/2683468

McKnight, L., & Fitton, D. (2010). *Touch-screen technology for children: Giving the right instructions and getting the right responses* (No. January 2010). https://doi.org/10.1145/1810543.1810580

Mildenberger Verlag. (2018). *Eine Einführung in die Silbenmethode* [An introduction to the Syllable Method]. Mildenberger Verlag.

Mildenberger Verlag GmbH. (2016). ABC der Tiere 1 [Animal Alphabet 1] [Mobile application software]. https://play.google.com/store/apps/details?id=air.de.mildenbergerverlag.abc1android

Mitkov, R. (2004). *The Oxford handbook of computational linguistics.* Oxford University Press.

Moll, K., Kunze, S., Neuhoff, N., Bruder, J., & Schulte-Körne, G. (2014). Specific learning disorder: Prevalence and gender differences. *PLoS ONE*, *9*(7), e103537. https://doi.org/10.1371/journal.pone.0103537

Moll, K., & Landerl, K. (2009). Double dissociation between reading and spelling deficits. *Scientific Studies of Reading*, *13*(5), 359–382. https://doi.org/10.1080/10888430903162878

Moll, K., & Landerl, K. (2010). *SLRT II – Lese- und Rechtschreibtest. Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT)* [SLRT II – Salzburg Reading and Spelling Test]. Bern: Verlag Hans Huber. https://www.testzentrale.de/shop/lese-und-rechtschreibtest.html

Moll, K., Wallner, R., & Landerl, K. (2012). Kognitive Korrelate der Lese-, Leserechtschreib- und der Rechtschreibstörung [Cognitive correlates of deficits in reading and/or spelling]. *Lernen und Lernstörungen*, *1*(1), 7–19. https://doi.org/10.1024/2235-0977/a000002

Müller, R. (2003a). *Diagnostischer Rechtschreibtest für 2. Klassen (DRT 2)* [Diagnostic Spelling Test for 2nd Grade (DRT 2)]. Göttingen: Beltz. https://www.testzentrale.de/shop/diagnostischer-rechtschreibtest-fuer-2-klassen.html

Müller, R. (2003b). *Diagnostischer Rechtschreibtest für 3. Klassen (DRT 3)* [Diagnostic Spelling Test for 3rd Grade (DRT 3)] (4, Ed.). Göttingen: Beltz. https://www.testzentrale.de/shop/diagnostischer-rechtschreibtest-fuer-3-klassen.html

Müller, R. (2013). ABC der Tiere – Silben-Generator für die Grundschule. [Computer software]. Mildenberger Verlag. www.abc-der-tiere.de/index.php?id=388

Ninaus, M., Kiili, K., McMullen, J., & Moeller, K. (2016). A game-based approach to examining students' conceptual knowledge of fractions. In R. Bottino, J. Jeuring, & R. C. Veltkamp (Eds.), *Games and Learning Alliance. GALA 2016* (pp. 37–49). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-50182-6_4

Ninaus, M., Kiili, K., Mcmullen, J., & Moeller, K. (2017). Assessing fraction knowledge by a digital game. *Computers in Human Behavior*, *70*, 197–206. https://doi.org/10.1016/j.chb.2017.01.004

Ninaus, M., Moeller, K., McMullen, J., & Kiili, K. (2017). Acceptance of Game-Based Learning and Intrinsic Motivation as Predictors for Learning Success and Flow Experience. *International Journal of Serious Games*, *4*(3), 15–30. https://doi.org/10.17083/ijsg.v4i3.176

Norman, D. A., & Draper, S. W. (1986). User centered system design; new perspectives on human-computer interaction. *Journal educational computing research*, *3*(1), 129–134.

Nwana, H. S. (1990). Intelligent tutoring systems: An overview. *Artificial Intelligence Review*, *4*(4), 251–277. https://doi.org/10.1007/BF00168958

Oddcast, Inc. (2018). Oddcast Text to Speech [Computer software]. https://www.oddcast.com/technologies/tts/

Olson, R. K., & Wise, B. W. (1992). Reading on the computer with orthographic and speech feedback. *Reading and Writing*, *4*(2), 107–144. https://doi.org/10.1007/BF01027488

Overy, K. (2003). Dyslexia and music: From timing deficits to musical intervention. *Annals of the New York Academy of Sciences*, *999*(1), 497–505. https://doi.org/10.1196/annals.1284.060

Overy, K., Nicolson, R. I., Fawcett, A. J., & Clarke, E. F. (2003). Dyslexia and music: measuring musical timing skills. *Dyslexia*, *9*(1), 18–36. https://doi.org/10.1002/dys.233

Parker, L. E., Lepper, M. R., Bartholomew, K., Cordova, D., & Mayer, J. (1992). Effects of fantasy contexts on children's learning and motivation: Making learning more fun. *Journal of personality and social psychology*, *62*(4), 625–633.

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction*, *29*, 128–140. https://doi.org/10.1016/j.learninstruc.2013.02.006

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, *50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

Poels, K., de Kort, Y. A. W., & IJsselsteijn, W. A. (2007). *D3.3 : Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games*. Eindhoven: Technische Universiteit Eindhoven.

Pompino-Marschall, B. (2009). *Einführung in die Phonetik* [Introduction to phonetics]. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110224818

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment*, *1*(1), 21. https://doi.org/10.1145/950566.950596

Ptok, M., & Meisen, R. (2008). Die Fähigkeit, Minimalpaare nachzusprechen, korreliert mit der Rechtschreibleistung [The ability to repeat minimal pairs correlates with spelling performance]. *HNO*, *56*(1), 73–80.

Quinn, C. N. (2005). *Engaging learning: Designing e-learning simulation games*. John Wiley & Sons.

Ramus, F. (2003). Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, *13*(2), 212–218. https://doi.org/https://doi.org/10.1016/S0959-4388(03)00035-7

Rau, A. K., Moll, K., Snowling, M. J., & Landerl, K. (2015). Effects of orthographic consistency on eye movement behavior: German and english children and adults process the same words differently. *Journal of Experimental Child Psychology*, *130*, 92–105. https://doi.org/10.1016/j.jecp.2014.09.012

Rauschenberger, M., Fuechsel, S., Rello, L., Bayarri, C., & Gòrriz, A. (2015). A game to target the spelling of german children with dyslexia. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (pp. 445–446). https://doi.org/10.1145/2700648.2811345

Read, J. C. (2008). Validating the fun toolkit: An instrument for measuring children's opinions of technology. *Cognition, Technology & Work*, *10*(2), 119–128. https://doi.org/10.1007/s10111-007-0069-9

Read, J. C., Macfarlane, S., & Casey, C. (2002). Endurability, engagement and expectations: Measuring children's fun. In *Interaction design and children* (Vol. 2, pp. 1–23).

Reichheld, F. (2004). The one number you need to grow. *Harvard Business Review*, *81*, 46–54.

Reid, G., Strnadová, I., & Cumming, T. (2013). Expanding horizons for students with dyslexia in the 21st century: Universal design and mobile technology. *Journal of Research in Special Educational Needs*, *13*(3), 175–181. https://doi.org/10.1111/1471-3802.12013

Rello, L., Bayarri, C., & Gòrriz, A. (2013). Dyslexia exercises on my tablet are more fun. *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility - W4A '13*, 1. https://doi.org/10.1145/2461121.2461148

Rello, L., Bayarri, C., Otal, Y., & Pielot, M. (2014). A computer-based method to improve the spelling of children with dyslexia. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility - ASSETS '14* (pp. 153–160). https://doi.org/10.1145/2661334.2661373

Reuter-Liehr, C. (1993). Behandlung der Lese-Rechtschreibschwäche nach der Grundschulzeit: Anwendung und Überprüfung eines Konzeptes [Treatment of dyslexia after primary school: Application and verification of a concept]. *Zeitschrift für Kinder- und Jugendpsychiatrie*, *21*(3), 135–147.

Röber, C. (2009). Die Leistungen der Kinder beim Lesen-und Schreibenlernen [Children's performances in learning to read and write]. *Grundlagen der Silbenanalytischen Methode. Ein Arbeitsbuch mit Übungsaufgaben. Baltmannsweiler: Schneider Hohengehren*.

Röber, C. (2012). Die Systematik der Orthographie als Basis von Analysen von Kinderschreibungen. Eine empirische Untersuchung zur Schreibung der i-Laute [The systematics of orthography as a basis for analyses of children's spellings. An empirical study of the spelling of the "i" sounds]. In U. Bredel & H. Günther (Eds.), *Orthographietheorie und Rechtschreibunterricht* (pp. 71–102). Tübingen: Niemeyer.

Roman, K. (2015). *Children's ability to navigate through a streaming services such as TV4 Play* (Master's thesis, KTH, School of Computer Science and Communication (CSC)). http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-169931

Ronimus, M., Kujala, J., Tolvanen, A., & Lyytinen, H. (2014). Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge. *Computers & Education*, *71*, 237–246. https://doi.org/10.1016/j.compedu.2013.10.008

Ronimus, M., & Lyytinen, H. (2015). Is school a better environment than home for digital game-based learning? the case of GraphoGame. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*, *11*(2), 123–

147. https://doi.org/10.17011/ht/urn.201511113637

Rose, S. J. (2009). *Identifying and Teaching Children and Young People with Dyslexia and Literacy Difficulties.* Department for Children, Schools and Families.

Sauter, K., Heller, J., & Landerl, K. (2012). Sprachrhythmus und Schriftspracherwerb [Speech rhythm skills and literacy]. *Lernen und Lernstörungen*, *1*(4), 225–239. https://doi.org/10.1024/2235-0977/a000023

Scheerer-Neumann, G. (1981). The utilization of intraword structure in poor readers: Experimental evidence and a training program. *Psychological Research*, *43*(2), 155–178. https://doi.org/10.1007/BF00309827

Schmitt, K. L., Hurwitz, L. B., Duel, L. S., Linebarger, D. L. N., & Nichols Linebarger, D. L. (2018). Learning through play: The impact of web-based games on early literacy development. *Computers in Human Behavior*, *81*, 378–389. https://doi.org/10.1016/j.chb.2017.12.036

Schröder, M., & Trouvain, J. (2003). The German text-to-speech synthesis system MARY: A tool for research, development and teaching. *International Journal of Speech Technology*, *6*(4), 365–377. https://doi.org/10.1023/A:1025708916924

Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015). childLex: aa lexical database of German read by children. *Behavior Research Methods*, *47*(4), 1085–1094. https://doi.org/10.3758/s13428-014-0528-1

Schulte-Koerne, G., Deimel, W., Huelsmann, J., Seidler, T., & Remschmidt, H. (2001). Das Marburger Rechtschreib-Training - Ergebnisse einer Kurzzeit-Intervention [The Marburg Spelling Training Program - Results of a short-term intervention]. *Zeitschrift fuer Kinder- und Jugendpsychiatrie und Psychotherapie*, *29*(1), 7–15. https://doi.org/10.1024//1422-4917.29.1.7

Schulte-Körne, G., Deimel, W., & Remschmidt, H. (1998). Das Marburger Eltern-Kind-Rechtschreibtraining–Verlaufsuntersuchung nach zwei Jahren [The Marburg parent-child-spelling training – follow up after two years]. *Zeitschrift für Kinder-und Jugendpsychiatrie*, *26*(3), 167–173.

Schulte-Körne, G., & Mathwig, F. (2013). *Das Marburger Rechtschreibtraining: ein regelbegleitetes Förderprogramm für rechtschreibschwache Kinder* [The Marburg Spelling Training: A rule-based training program for children with spelling difficulties]. Hogrefe. https://www.testzentrale.de/shop/das-marburger-rechtschreibtraining.html

Schulte-Körne, G., & Remschmidt, H. (2003). Legasthenie - Symptomatik, Diagnostik, Ursachen, Verlauf und Behandlung [Dyslexia – Symptomatology, diagnosis, etiology, long-term outcome and therapy]. *Dtsch Arztebl International*, *100*(7), 396–406.

Schulte-Körne, G., Schäfer, J., Deimel, W., & Remschmidt, H. (1997). Das Marburger Eltern-Kind-Rechtschreibtraining [The Marburg Parent-Child Spelling Training]. *Zeitschrift für Kinder-und Jugendpsychiatrie*, *25*(3), 151–159.

Schulte-Körne, G. (2010). The prevention, diagnosis, and treatment of dyslexia. *Deutsches Arzteblatt international*, *107*(41), 718–727. https://doi.org/10.3238/arztebl.2010.0718

Schunk, D. H. (1989). Self-efficacy and achievement behaviors. *Educational Psychology Review*, *1*(3), 173–208. https://doi.org/10.1007/BF01320134

Schwartz, R. N., & Plass, J. L. (2014). Click versus drag: User-performed tasks and the enactment effect in an interactive multimedia environment. *Computers in Human Behavior*, *33*, 242–255. https://doi.org/10.1016/j.chb.2014.01.012

Senn, S. S. (2002). *Cross-over trials in clinical research* (Vol. 5). John Wiley & Sons.

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2015). Stability of individual parameters in a model of optimal fact learning. In *Proceedings of the 13th International Conference on Cognitive Modeling* (pp. 136–141). http://www.iccm2015.org/proceedings/papers/0034/paper0034.pdf

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Topics in Cognitive Science*, *8*(1), 305–321. https://doi.org/10.1111/tops.12183

Seymour, P. H., Aro, M., Erskine, J. M., Wimmer, H., Leybaert, J., Elbro, C., … Olofsson, Å. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174. https://doi.org/10.1348/000712603321661859

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Snowling, M. J. (1995). Phonological processing and developmental dyslexia. *Journal of Research in Reading*, *18*(2), 132–138. https://doi.org/10.1111/j.1467-9817.1995.tb00079.x

Snowling, M. J. (2001). From language to reading and dyslexia. *Dyslexia*, *7*(1), 37–46. https://doi.org/10.1002/dys.185

Soegaard, M. (2015). Interaction Styles. In *The Glossary of Human Computer Interaction.* The Interaction Design Foundation. https://www.interaction-design.org/literature/book/the-glossary-of-human-computer-interaction/interaction-styles

Soliman, S., & Nathan-Roberts, D. (2018). Evaluating children's interaction with touchscreens from 0 to 8 years old. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 260–264.

https://doi.org/10.1177/1541931218621060

Soni, N., Aloba, A., Morga, K. S., Wisniewski, P. J., & Anthony, L. (2019). A framework of touchscreen interaction design recommendations for children (TIDRC). In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (pp. 419–431). https://doi.org/10.1145/3311927.3323149

Sonnenwald Apps. (2013). ABC Deutsch lernen Grundschule [ABC – Learning German in Primary School] [Mobile application software]. Sonnenwald Apps. https://play.google.com/store/apps/details?id=kellerm.org.school.writer

Squire, K. D. (2008). Video games and education: Designing learning systems for an interactive age. *Educational Technology*, *48*(2), 17–26. http://www.jstor.org/stable/44429558

Staffeldt, S. (2010). *Einführung in die Phonetik, Phonologie und Graphematik des Deutschen. Ein Leitfaden für den akademischen Unterricht* [An introduction to phonetics, phonology, and graphematics of German. A guide to academic instruction]. Tübingen: Stauffenburg.

Tacke, G. (2005). Evaluation eines Lesetrainings zur Förderung lese-rechtschreibschwacher Grundschüler der zweiten Klasse [Evaluation of a reading training for second-grade primary school children with poor reading and spelling skills]. *Psychologie in Erziehung und Unterricht*, *52*(3), 198–209.

Tacke, G., Wörner, R., Schultheiss, G., & Brezing, H. (1993). Die Auswirkung rhythmisch-syllabierenden Mitsprechens auf die Rechtschreibleistung [The effect of rhythmic-syllabic speaking on spelling performance]. *Zeitschrift für Pädagogische Psychologie*, *7*(2), 3.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., ... Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science*, *271*(5245), 81–84. https://doi.org/10.1126/science.271.5245.81

Thomson, J. M., & Jarmulowicz, L. (2016). *Linguistic Rhythm and Literacy* (No. volume 17). John Benjamins Publishing Company. https://doi.org/10.1075/tilar.17

Thomson, J. M., Leong, V., & Goswami, U. (2013). Auditory processing interventions and developmental dyslexia: A comparison of phonemic and rhythmic approaches. *Reading and Writing*, *26*(2), 139–161. https://doi.org/10.1007/s11145-012-9359-6

Um, E. R., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology*, *104*(2), 485–498. https://doi.org/10.1037/a0026609

Vatavu, R.-D., Cramariuc, G., & Schipor, d. M. (2015). Touch interaction for children aged 3 to 6 years: Experimental findings and relationship to motor skills. *International Journal of Human-Computer Studies*, *74*, 54–76. https://doi.org/10.1016/j.ijhcs.2014.10.007

Vennemann, T. (2011). *Preference Laws for Syllable Structure: And the Explanation of Sound Change with Special Reference to German, Germanic, Italian, and Latin.* Walter de Gruyter.

Ward, M. (2014). HCI requirements for young primary school CALL learners. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration* (Vol. 8524, pp. 339–348). https://doi.org/10.1007/978-3-319-07485-6_34

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2004). Discrimination of word stress in early infant perception: Electrophysiological evidence. *Cognitive Brain Research*, *18*(2), 149–161. https://doi.org/10.1016/j.cogbrainres.2003.10.001

Wellek, S., & Blettner, M. (2012). On the proper use of the crossover design in clinical trials. *Deutsches Aerzteblatt Online*. https://doi.org/10.3238/arztebl.2012.0276

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education*, *60*(1), 412–425. https://doi.org/10.1016/j.compedu.2012.07.018

Part V

Publications and Contributorship

Chapter 10

Contributorship

Parts of this thesis have been published elsewhere:

1. **Holz, H.**, Beuttler, B., & Kirsch, A. (2017). Bewegungserkennung mit Wearables für Embodied Trainings in Serious Games [Movement Recognition with Wearables for Embodied Trainings in Serious Games]. In M. Burghardt, R. Wimmer, C. Wolff, & C. Womser-Hacker (Eds.), *Mensch und Computer 2017 – Tagungsband* (pp. 259–262). doi: 10.18420/muc2017-mci-0368

    Own contribution
    I designed and carried out the study with input from A.K. The methodology was mainly designed and developed by myself with input from A.K. All programming was done by myself. I performed the analyses and interpreted the data with input from B.B. and A.K. The data were collected by myself. The manuscript was mainly written by myself with input from B.B. and A.K. I consider my contribution to this work to be approximately 90% of total work.

2. **Holz, H.**, Brandelik, K., Brandelik, J., Beuttler, B., Kirsch, A., Heller, J., & Meurers, D. (2017). Prosodiya – A Mobile Game for German Dyslexic Children. In J. Dias, P. A. Santos, & R. C. Veltkamp (Eds.), *Games and Learning Alliance. GALA 2017* (pp. 73–82). doi: 10.1007/978-3-319-71940-5_7

    Own contribution
    Together with K.B. and J.B., I designed and carried out the study with input from J.H. The methodology was designed and developed by K.B., J.B., and myself. The training was designed by K.B., J.B., and myself. I developed the mobile application and other software used in the study. The data were analyzed and interpreted by myself and K.B. with input from J.B. The data were collected by K.B., J.B., and myself. The manuscript was mainly written by myself with contributions from K.B. and input from A.K. and D.M. I consider my contribution to this work to be approximately 50% of total work.

3. **Holz, H.**, Brandelik, K., Beuttler, B., Brandelik, J., & Ninaus, M. (2018). How to train your syllable stress awareness – A digital game-based approach for German dyslexic children. *International Journal of Serious Games*, 5(3), 37–59. doi: 10.17083/ijsg.v5i3.242

   Own contribution
   I conducted the literature review on different forms of treatment approaches for learning disorders with input from K.B. and M.N. The proposed training was designed by K.B., myself, J.B., and B.B.† The mobile application and other software were developed by myself and B.B. The manuscript was mainly written by myself with contributions from M.N. and input from K.B. I consider my contribution to this work to be approximately 70% of total work.

4. **Holz, H.**, Beuttler, B., & Ninaus, M. (2018). Design Rationales of a Mobile Game-Based Intervention for German Dyslexic Children. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 205–219). doi: 10.1145/3270316.3272053

   Own contribution
   I designed, supervised, and carried out the study with input from M.N. and B.B. The methodology was mainly designed and developed by myself with input from N.M. and B.B. The training was designed by K.B., myself, J.B., and B.B.† The mobile application and other software was developed by myself and B.B. I performed the analyses and interpreted the data with input from M.N. and D.M. The data were collected by myself, B.B., and student assistants. The manuscript was mainly written by myself with contributions from B.B. and M.N. I consider my contribution to this work to be approximately 60% of total work.

5. **Holz, H.**, Ninaus, M., Meurers, D., & Kirsch, A. (2018). Validity and Player Experience of a Mobile Game for German Dyslexic Children. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 469–478). doi: 10.1145/3270316.3271523

   Own contribution
   I designed, supervised, and carried out the study with input from M.N., D.M, and A.K. The methodology was mainly designed and developed by myself with input from N.M. and D.M. The training was designed by K.B., myself, J.B., and B.B.† The mobile application and other software was developed by myself and B.B. I performed the analyses and interpreted the data with input from M.N. and D.M. The data were collected by myself and student assistants. The manuscript was

† Although not all core team members of the Prosodiya project are always listed as authors of a manuscript, the training was designed by Katharina Brandelik, Heiko Holz, Jochen Brandelik, and Benedikt Beuttler.

mainly written by myself with contributions from M.N. and input from A.K. I consider my contribution to this work to be approximately 70% of total work.

6. **Holz, H.**, Chinkina, M., & Vetter, L. (2018). Optimizing the Quality of Synthetically Generated Pseudowords for the Task of Minimal-Pair-Distinction. In *2018 IEE Spoken Language Technology Workshop (SLT)* (pp. 470–476). doi: 10.1109/SLT.2018.8639037

   Own contribution
   The idea of this work was originally mine and was further developed together with L.V. The study was designed and carried out by L.V. under my supervision. The methodology was designed and developed by L.V. and myself. The items were generated by L.V. with my input. The experiment was programmed by myself with input from L.V. The data were collected by L.V. and myself. The analyses were performed by L.V. with input from M.C. The data were interpreted by L.V., M.C., and myself. The manuscript was mainly written by myself and M.C. with contributions from L.V. I consider my contribution to this work to be approximately 35% of total work.

7. **Holz, H.**, Weiss, Z., Brehm, O., & Meurers, D. (2018). COAST – Customizeable Online Syllable Enhancement in Texts. A flexible framework for automatically enhancing reading materials. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 89–100). doi: 10.18653/v1/W18-0509

   Own contribution
   The idea of this work was originally mine and was further developed together with O.B. The study was designed and carried out by O.B. under my supervision. The methodology was designed and developed by O.B. and myself. The COAST software was developed by O.B. with my input. The analyses were performed by O.B. with my input. The data were interpreted by O.B., myself, and Z.W. The manuscript was mainly written by myself and Z.W. with contributions from O.B. and input from D.M. I consider my contribution to this work to be approximately 35% of total work.

8. **Holz, H.**, & Meurers, D. (submitted). Interaction Styles in Context: Comparing Drag-and-Drop, Point-and-Touch, and Touch in a Mobile Spelling Game.

   Own contribution
   I designed, supervised, and carried out the study with input from D.M. The methodology was mainly designed and developed by myself with input from D.M. The experiment was programmed by myself. I performed the analyses and interpreted the data with input from D.M. The data were collected by myself and student assistants. The manuscript

was mainly written by myself with contributions from D.M. I consider my contribution to this work to be approximately 90% of total work.

9. **Holz, H.**, Ninaus, M., Beuttler, B., Brandelik, K., & Meurers, D. (unpublished). A Digital Game-Based Training Improves Spelling in German Primary School Children – A Randomized Controlled Field Trial.

    Own contribution
    I designed, supervised, and carried out the study with input from B.B., K.B., and D.M. The methodology was mainly designed and developed by myself with input from B.B., K.B. and D.M. The training was designed by K.B., myself, J.B., and B.B.† The mobile application and other software was developed by myself and B.B. I performed the analyses and interpreted the data with input from M.N. and D.M. The data were collected by myself, B.B., K.B., and student assistants. The manuscript was mainly written by myself with contributions from M.N., K.B., and D.M. I consider my contribution to this work to be approximately 60% of total work.

Chapter 11

Publications

The listed publications are attached on the following pages.

1. **Holz, H.**, Beuttler, B., & Kirsch, A. (2017). Bewegungserkennung mit Wearables für Embodied Trainings in Serious Games [Movement Recognition with Wearables for Embodied Trainings in Serious Games]. In M. Burghardt, R. Wimmer, C. Wolff, & C. Womser-Hacker (Eds.), *Mensch und Computer 2017 – Tagungsband* (pp. 259–262). doi: 10.18420/muc2017-mci-0368
2. **Holz, H.**, Brandelik, K., Brandelik, J., Beuttler, B., Kirsch, A., Heller, J., & Meurers, D. (2017). Prosodiya – A Mobile Game for German Dyslexic Children. In J. Dias, P. A. Santos, & R. C. Veltkamp (Eds.), *Games and Learning Alliance. GALA 2017* (pp. 73–82). doi: 10.1007/978-3-319-71940-5_7
3. **Holz, H.**, Brandelik, K., Beuttler, B., Brandelik, J., & Ninaus, M. (2018). How to train your syllable stress awareness – A digital game-based approach for German dyslexic children. *International Journal of Serious Games*, 5(3), 37–59. doi: 10.17083/ijsg.v5i3.242
4. **Holz, H.**, Beuttler, B., & Ninaus, M. (2018). Design Rationales of a Mobile Game-Based Intervention for German Dyslexic Children. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 205–219). doi: 10.1145/3270316.3272053
5. **Holz, H.**, Ninaus, M., Meurers, D., & Kirsch, A. (2018). Validity and Player Experience of a Mobile Game for German Dyslexic Children. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 469–478). doi: 10.1145/3270316.3271523
6. **Holz, H.**, Chinkina, M., & Vetter, L. (2018). Optimizing the Quality of Synthetically Generated Pseudowords for the Task of Minimal-Pair-Distinction. In *2018 IEE Spoken Language Technology Workshop (SLT)* (pp. 470–476). doi: 10.1109/SLT.2018.8639037

7. **Holz, H.**, Weiss, Z., Brehm, O., & Meurers, D. (2018). COAST – Customizeable Online Syllable Enhancement in Texts. A flexible framework for automatically enhancing reading materials. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 89–100). doi: 10.18653/v1/W18-0509

8. **Holz, H.**, & Meurers, D. (submitted). Interaction Styles in Context: Comparing Drag-and-Drop, Point-and-Touch, and Touch in a Mobile Spelling Game.

9. **Holz, H.**, Ninaus, M., Beuttler, B., Brandelik, K., & Meurers, D. (unpublished). A Digital Game-Based Training Improves Spelling in German Primary School Children – A Randomized Controlled Field Trial.

Bewegungserkennung mit Wearables für Embodied Trainings in Serious Games

Heiko Holz[1], Benedikt Beuttler[2], Alexandra Kirsch[3]

Graduiertenschule und Forschungsnetzwerk LEAD, Universität Tübingen[1]
Lehrstuhl für Forschungsmethoden und Mathematische Psychologie, Universität Tübingen[2]
Lehrstuhl für Mensch-Computer-Interaktion, Universität Tübingen[3]

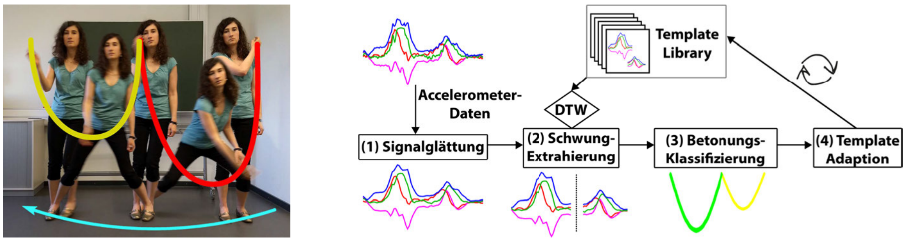
Zusammenfassung

Embodied Trainings – Übungen mit Ganzkörperbewegungen und Gesten – können das Lernen mit computergestützten Lernspielen für Kinder mit Lernschwächen effizienter und motivierender gestalten. In dem Lernspiel „Silbenschwingen mit Betonung" nutzen wir die in Smartphones und Fitnessarmbänder eingebauten Accelerometer zur Bewegungs- und Gestenerkennung. Damit untersuchen wir die Möglichkeit, inwiefern solche Embodied Trainings auch für mobile Serious Games entwickelt werden können. Wir erreichen bei der Klassifizierung von Silbenschwüngen unter Einsatz von Dynamic Time Warping und nur einem Training pro Geste bei Erwachsenen eine Genauigkeit von 99.3 % und bei Kindern zwischen 82.2 % und 93.9 %. Die Ergebnisse zeigen, dass einfache Gesten wie Silbenschwünge zuverlässig erkannt werden können, verdeutlichen jedoch auch die Herausforderung bei der Entwicklung von Embodied Trainings für Kinder.

1 Einleitung

In der Lerntherapie können körperbetonte Übungen – sogenannte Embodied Trainings – lernschwache Kinder durch den Einsatz von (Ganzkörper-) Bewegungen zusätzlich fördern. Embodied Trainings zielen auf die Entwicklung körpereigener Gedächtnisstützen ab, bei denen der Lernprozess positiv durch Körperbewegungen und -erfahrungen beeinflusst wir. So kann der gelernte Inhalt später besser abgerufen werden (Dackermann et al., 2017).
Für Kinder mit einer Lese-Rechtschreib-Schwäche (LRS) werden Embodied Trainings in vielen analogen Förderkonzepten erfolgreich eingesetzt Das Silbenschwingen ist eine dieser Fördermaßnahmen (Michel, 2008; Reuter-Liehr, 1993). Indem die Kinder den Sprachrhythmus von Worten schwingen (Abb. 1), wird durch ein gezieltes Training der Silbenanalyse eine Verbesserung der Lese- und Rechtschreibleistung erwartet.
Um Embodied Trainings auch in mobilen Serious Games zu unterstützen, können Accelerometer verwendet werden, die in mobilen Geräten verbaut sind. Mit den Accelerometern können körperbasierte Interaktionen wie Bewegung oder Gestik erkannt und klassifiziert werden.

Veröffentlicht durch die Gesellschaft für Informatik e. V. 2017 in
M. Burghardt, R. Wimmer, C. Wolff, C. Womser-Hacker (Hrsg.):
Mensch und Computer 2017 – Tagungsband, 10.–13. September 2017, Regensburg.
 https://doi.org/10.18420/muc2017-mci-0368



(a) *(b)*

Abbildung 1: Das „Silbenschwingen mit Betonung“ für das Wort „malen“. Kinder lernen durch „Tanzen“ des Sprachrhythmus' (Abb. 1a), Wörter zu silbieren und das Betonungsmuster zu bestimmen. Dabei sprechen sie jede Silbe deutlich aus, schreiten gleichzeitig seitwärts durch den Raum und schwingen mit der Schreibhand – für betonte Silben einen größeren, stärkeren und für unbetonte einen kleineren – Silbenbogen von der linken zur rechten Schulter. Abb. 1b zeigt das System zur Klassifizierung der Schwungbewegungen.

Obwohl nahezu perfekten Ergebnisse von bis zu 99.79 % der Gestenerkennung mit solchen Geräten (Xie und Cao, 2016; Akl und Valaee, 2010) dafür sprechen, sind uns keine mobilen Serious Games bekannt, welche durch Gestenerkennung Embodied Trainings unterstützen.
Als Teil des mobilen Serious Games „Prosodiya“ (Holz et al., im Druck) haben wir das Embodied Training „Silbenschwingen mit Betonung“ (SmB) entwickelt (Abb. 1a). Prosodiya fokussiert im Besonderen die Sensibilisierung auf die sprachrhythmische Struktur der Silbenbetonung. Aufgabe der Kinder beim SmB ist es, für vorgegebene Wörter das Betonungsmuster zu schwingen. Die Schwungbewegungen werden mit Accelerometern aufgenommen und durch Gestenerkennung als Betonungsmuster klassifiziert, sodass die Lösung der Kinder automatisch ausgewertet werden kann. Im Gegensatz zum klassischen Silbenschwingen unterscheiden wir zwischen betonten und unbetonten Silben, da Schwierigkeiten im Erkennen der Betonung mit dem Phänomen der LRS assoziiert zu sein scheinen (Sauter et al., 2012).
In diesem Artikel untersuchen wir anhand des SmB, inwiefern Smartphones und Fitnessarmbänder mit eingebauten Accelerometern für die Entwicklung von Embodied Trainings mit einfachen, körperbasierten Interaktionen für mobile Serious Games verwendet werden können.

2 Methodik

Für die Datenaufnahme haben wir Android-Smartphones und herkömmliche Fitnessarmbänder verwendet, deren Accelerometer eine Sampling-Rate von 50 Hz, bzw. 10 Hz, besitzen. Wie wir in Nutzertests festgestellt haben, sind Smartphones aufgrund ihrer Größe für Grundschulkinder zum Schwingen ungeeignet. Deshalb wird bei ihnen nur das Fitnessarmband verwendet. Die Extrahierung und Klassifizierung der Schwungbewegungen erfolgt in vier Schritten (Abb. 1b):
In Schritt 1 wird das Beschleunigungssignal mittels eines Tiefpassfilters geglättet.
In Schritt 2 werden Schwünge einzelner Silben extrahiert. Ein Schwellenwertverfahren durchsucht das Beschleunigungssignal nach Maxima (Peaks), welche sich zwischen zwei Minima befinden. Die Minima müssen zwischen 500 ms und 1800 ms auseinander liegen. Von diesen Signalausschnitten wird die Ähnlichkeit zu bereits klassifizierten Schwüngen der Template

Library berechnet. Die Berechnung erfolgt mittels Dynamic Time Warping (DTW). DTW bildet zwei zu vergleichende Sequenzen mittels nichtlinearer zeitlicher Verzerrung aufeinander ab (vgl. u. a. Keogh und Ratanamahatana, 2004). Der Vorteil gegenüber statistischen Methoden liegt darin, dass für personalisierte Gesten bei gleicher oder besserer Erkennungsleistung jeweils nur ein Trainingsdatensatz benötigt wird.Bei ausreichender Ähnlichkeit werden die Signalausschnitte als Silbenschwung klassifiziert, ansonsten verworfen.
Da die Kinder sehr unbeständig schwingen – unbetonte Silben werden oft in einem Wort stärker geschwungen als betonte in anderen und umgekehrt – kann die Betonung nicht mit der berechneten Ähnlichkeit zuverlässig bestimmt werden. Deshalb wird in Schritt 3 die Betonung anhand des Peaks bestimmt: Der Schwung mit dem größten Peak innerhalb eines Wortes wird als betont, die anderen als unbetont klassifiziert. Dies wurde gemeinsam mit Lerntherapeuten bestimmt, da alle im Lernspiel verwendeten Wörter nur eine betonte Silbe besitzen.
In Schritt 4 werden korrekt klassifizierter Schwünge zur Template Library hinzugefügt. Da die Schwünge bei längerer Spieldauer schwächer und ungenauer ausgeführt wurden, muss die Template Library mit der Zeit angepasst werden. Sie enthält nur die Schwünge der letzten fünf Minuten, mindestens jedoch fünf Schwünge.

3 Evaluation

Das vorgestellte System wurde anhand zweier Nutzertests evaluiert. Für jedes Wort wurden die Accelerometerdaten ab ca. 500 ms vor dem Schwingen bis ca. 500 ms nach dem Schwingen mit einer Android-App aufgenommen. Mit dem ersten Wort wird die Template Library initiiert.
Im ersten Nutzertest haben fünf Erwachsene im Alter zwischen 20 und 32 Jahren ($M = 26$, $SD = 2.7$) mit einem Smartphone das SmB durchgeführt. Insgesamt wurden 11 Wörter mit den vier häufigsten Betonungsmustern der deutschen Sprache aufgenommen. Das System erzielte bei der Klassifikation eine Genauigkeit von 99.3%.
Im zweiten Nutzertest wurde das SmB von 9 Kindern im Alter zwischen 10 und 13 Jahren ($M = 11.1$, $SD = 1.05$) mit einem Fitnessarmband durchgeführt. Je nach Motivation wurden zwischen 11 und 14 Wörter aufgenommen. Zu Beginn wurde den Kindern das SmB samt Prinzip der betonten Silbe ausführlich erklärt. Um die korrekte Bestimmung der Betonung sicherzustellen, wurde das SmB zusätzlich vor jedem Wort vorgeführt.
Dennoch konnten wir beobachten, dass einigen Kinder Schwierigkeiten haben, Silben gemäß des Förderkonzepts zu schwingen: Betonte Silben wurden oft schwächer geschwungen als unbetonte oder waren kaum von diesen zu unterscheiden. Es ist nicht klar, ob dies an mangelhaftem Verständnis der Wortbetonung oder an fehlendem Feedback zur Schwungintensität liegt. Deshalb wurden sowohl alle 115 geschwungenen Wörter ($D1$, 307 Silbenschwünge) als auch ein bereinigter Datensatz ($D2$) evaluiert. $D2$ enthält nur die 95 Wörter (258 Silbenschwünge), bei welchen die Kinder das Betonungsmuster korrekt benannt haben und die betonte Silbe stärker oder nur geringfügig schwächer geschwungen wurde ($Peak_{betont} \geq \max_{Acc} - \Theta_{Peak}$).
Ohne Berücksichtigung der Betonung wurden die Silbenschwünge mit einer Genauigkeit von $96.8\,\%$ für $D1$ und $97.7\,\%$ für $D2$ erkannt. Die Extrahierung der Silbenschwünge mit anschließender Klassifizierung der Betonung erreicht eine Genaugikeit von 83% ($M = 82.8\,\%$, $SD = 13.8\,\%$-Pkt.) für $D1$ und 94.6% ($M = 93.9\%$, $SD = 6.5\,\%$-Pkt.) für $D2$.

4 Diskussion und Ausblick

Das vorgestellte System ist in der Lage, einfache Gesten wie Silbenschwünge zuverlässig zu erkennen. Bei Erwachsenen funktioniert dies mit Smartphones und Fitnessarmbändern. Kinder konnten sich aufgrund der Größe von Smartphones nur mit Fitnessarmbänder uneingeschränkt bewegen. Die Klassifikation der Silbenbetonung ist bei Erwachsenen ausreichend, während sie bei Kindern stark variiert. Die abweichenden Ergebnisse entstehen durch die im Nutzertest festgestellte Schwierigkeit, die Silben gemäß der Anleitung zu schwingen. Dies verdeutlicht drei wichtige Punkte, die bei der Entwicklung sensorgestützter Embodied Trainings für Kinder zu beachten sind: Erstens sollten Fitnessarmbänder aufgrund ihrer Größe und Tragekomfort verwendet werden. Zweitens müssen Systeme an die Bedürfnisse und das Verhalten der Kinder angepasst werden. Drittens darf von den Kindern das korrekte Ausführen der Gesten nicht vorausgesetzt werden, sondern muss verstanden und ausführlich trainiert werden.
Insgesamt konnten wir feststellen, dass Embodied Trainings mit einfachen, körperbasierten Interaktionen für mobile Serious Games entwickelt werden können. Abhängig der Zielgruppe eignen sich unterschiedliche mobile Geräte, in denen ein Accelerometer verbaut ist.
In Zukunft wollen wir die aufgedeckten Limitierungen durch eine bessere und intensivere Trainingsphase überwinden, welche direkt in das mobile Lernspiel integriert wird. Eine Figur aus der Spielwelt soll den Kindern ausführlich das Silbenschwingen erklären und direktes Feedback geben, ob die betonte Silbe ausreichend unterschiedlich geschwungen wurde.

Literaturverzeichnis

Akl, A. & Valaee, S. (2010). Accelerometer-based gesture recognition via dynamic-time warping, affinity propagation, & compressive sensing. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing* (S. 2270–2273).

Dackermann, T., Fischer, U., Nuerk, H.-C., Cress, U. & Moeller, K. (2017). Applying embodied cognition: from useful interventions and their theoretical underpinnings to practical applications. *ZDM*, 1–13.

Holz, H., Beuttler, B., Brandelik, K. & Brandelik, J. (im Druck). Prosodiya - ein Lernspiel zur Förderung des Sprachrhythmus bei Kindern mit LRS. In *Mensch und Computer 2017 - Tagungsband*. Gesellschaft für Informatik.

Keogh, E. & Ratanamahatana, C. A. (2004). Exact indexing of dynamic time warping. *Knowledge and Information Systems*, *7*(3), 358–386.

Michel, H.-J. (Hrsg.). (2008). *FRESCH. Freiburger Rechtschreibschule. Fit trotz LRS*. Buxtehude: AOL-Verlag.

Reuter-Liehr, C. (1993). Behandlung der Lese-Rechtschreibschwäche nach der Grundschulzeit: Anwendung und Überprüfung eines Konzeptes. *Zeitschrift für Kinder- und Jugendpsychiatrie*, *21*(3), 135–147.

Sauter, K., Heller, J. & Landerl, K. (2012). Sprachrhythmus und Schriftspracherwerb. *Lernen und Lernstörungen*, *1*(4), 225–239.

Xie, R. & Cao, J. (2016). Accelerometer-Based Hand Gesture Recognition by Neural Network and Similarity Matching. *IEEE Sensors Journal*, *16*(11), 4537–4545.

Prosodiya – A Mobile Game for German Dyslexic Children

Heiko Holz[1,2(✉)], Katharina Brandelik[2], Jochen Brandelik[2], Benedikt Beuttler[2,3], Alexandra Kirsch[4], Jürgen Heller[3], and Detmar Meurers[1,5]

[1] LEAD Graduate School & Research Network, Eberhard Karls Unviersität Tübingen, Tübingen, Germany
{heiko.holz,detmar.meurers}@uni-tuebingen.de
[2] Tübinger Institut für Lerntherapie GmbH, Tübingen, Germany
{k.brandelik,j.brandelik}@til-lerntherapie.de
[3] Department of Psychology, Eberhard Karls Unviersität Tübingen, Tübingen, Germany
{benedikt.beuttler,juergen.heller}@uni-tuebingen.de
[4] Department of Computer Science, Eberhard Karls Unviersität Tübingen, Tübingen, Germany
alexandra.kirsch@hci.uni-tuebingen.de
[5] Department of Linguistics, Eberhard Karls Unviersität Tübingen, Tübingen, Germany

Abstract. Approximately 4–10% of the German population suffers from developmental dyslexia. The learning disorder afffects educational, personal, and social development of children in a negative way. Mobile serious games have the potential to support dyslexic primary-school children in addition to school support and learning therapy. We propose such a mobile serious game called "Prosodiya", with the help of which dyslexic children can improve their reading and writing performance. Prosodiya includes innovative and evidence-based interventions that focus on improving the awareness of linguistic features related to syllable stress. We report the results of a pilot study of a preliminary version of the game. Results indicate that the children enjoyed playing the game, that their motivation was maintained, and that they wanted more levels.

1 Introduction

Developmental dyslexia is one of the most frequent learning disorders [24]. Affected children suffer massively from an impaired literacy acquisition – compared to their classmates, they acquire reading and writing skills in a much slower pace and not as proficient [29]. Usually, they lose motivation for the learning process as well as faith that they will ever be able to develop a comprehension of literacy language [2]. If these children do not receive appropriate treatments, negative consequences in the long run may arise, such as poor graduation and

J. Dias et al. (Eds.): GALA 2017, LNCS 10653, pp. 73–82, 2017.
https://doi.org/10.1007/978-3-319-71940-5_7

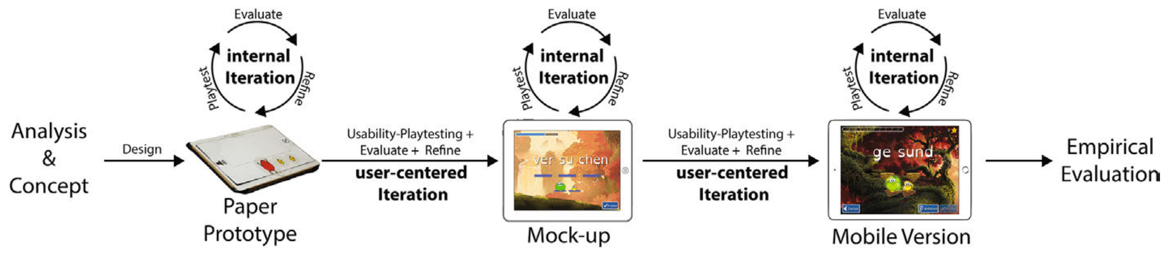


Fig. 1. Exemplary Iterative Children-Centered Game Design Process (ICCGD) for the game "stress pattern".

higher chances to drop out from school, resulting in poor employment prospects or unemployment [10]. Overall, the learning disorder has a negative impact not only on mental health but also on social and cultural participation [1]. Serious Games have the potential to address the aforementioned difficulties. In shaping an individual's learning curve, they boost motivation [11], and lead to successful learning processes (cf. [5]). With the great advantage of independence of time and location, mobile games can additionally help the children to overcome their learning disorder outside of learning therapy and classrooms – serious games for dyslexic children have been proven to have positive effects on the process of literacy acquisition (e.g., [4,18]).

In the first part of this article, we present such a mobile serious game for dyslexic children aged 6–12. Focus of this game called "Prosodiya" is on well-founded user-centered development, on the adaptivity of the digital interventions, and on embodied training. To the best of our knowledge, Prosodiya adds two novelties to the field of serious games for dyslexic children. First, it is the first digital therapy approach that focuses on improving the awareness of syllable stress and associates the stressed syllable's linguistic features to orthographic principles of the German orthography. Second, it is the first mobile game supporting embodied training using sensor-based gesture recognition.

In the second part of this article we present the results of a pilot testing a preliminary version of the game. Focus is on user experience and usability, but the impact on literacy development is also considered.

2 Prosodiya

To avoid failure caused by losing focus on the target audience, we particularly emphasized the involvement of primary-school children as the end user. We followed an approach called iterative children-centered game design (ICCGD, see Fig. 1) during the whole development process. The ICCGD combines the two familiar approaches of user-centered design [22] (UCD) and iterative game design [12] (IGD).

Prosodiya is based on recent empirical findings and on evidence-based interventions (e.g., [17,27]). For example, a main component is training phonological awareness as children with dyslexia often struggle with this basic skill [6,32].

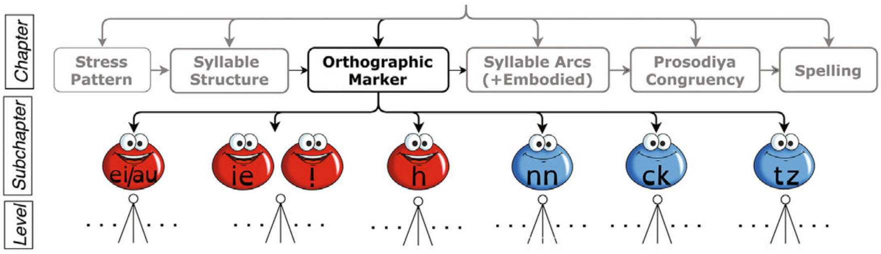


Fig. 2. Structure of Prosodiya. The third chapter (Orthographic Marker) is illustrated in detail. Red blobs refer to open and blue blobs to closed syllables. (Color figure online)

Phonological awareness refers to the ability to deal with the sound system of a language and to detect, distinguish and manipulate segments of a language like syllables, rimes, or even single sounds [20]. It also includes the adequate perception and processing of prosodic features such as syllable stress. Performance in detecting stress highly correlates with reading and writing skills [7,14,28] and recent research shows that a shortcoming in syllable stress detection in the context of words or sentences is a very strong predictor of dyslexia (e.g., [7,14,21]).

Prosodiya builds on this research and trains stress detection (e.g., Fig. 3a). First, this might boost a child's ability to segment words into relevant components. Second, children learn to focus on relevant areas in words, as orthographic challenges mainly occur in stressed syllables: In the German orthographic system, there is a strong association between stress and vowel length markers – vowel length markers generally occur in stressed syllables [33]. Prosodiya aims at clarifying this association. It helps children to focus on the stressed syllable and to learn how such syllables are spelled. In doing so, it finally leads to a rule-based orthographic spelling training inspired by the empirically evaluated Marbuger Rechtschreibtraining [17].

Little inhabitants called "Kugellichter" ("spherical lights"), kindred to will-o-wisps, guide the children through the world of syllables and orthography and accompany them through the story: The magical land of Prosodiya is haunted by a mysterious and maleficent fog, causing the inhabitant to live in sadness. Only the children can relieve the world from its sorrow. Prosodiya consists of six chapters, each corresponding to a different linguistic or orthographic challenge for which different mini-games were developed. Each chapter is composed of various subchapters, each targeting a specific level of linguistic or orthographic competence. Each subchapter is again composed of various levels. The levels increase in difficulty in that the target words' structures get more complex and the objective(s) of tasks get more challenging. The structure of Prosodiya can be seen in Fig. 2. Each part of the game is introduced and instructed by an interactive tutorial that explains game mechanics, imparts linguistic knowledge, and narrates the story. An integrated learner model aims at adapting the game to their individual proficiency, trying to keep up the learning curve, motivation, and fun. We give a brief introduction into the therapeutic process and its games. A further overview with videos is available at http://youtube.prosodiya.de.

(a) Stress Pattern (b) Open & Closed Syllables (c) Orthographic Marker

(d) Syllable Arcs (e) Embodied Syllable Arcs (f) Spelling

Fig. 3. Exercises of Prosodiya. See http://youtube.prosodiya.com for videos. (Color figure online)

In the first chapter, children develop and improve their awareness for syllable stress. They rebuild stress patterns of audio- and/or visually presented words by dragging and dropping cartoon blobs (Fig. 3a, big, green blobs for stressed and yellow blobs for unstressed syllables). This awareness is extended in the second chapter in which the structure of the stressed syllable is explored (Fig. 3b). They additionally have to decide whether the stressed syllable of a word is open (ends with a long vowel, represented by red blobs) or closed (ends with a consonant that closes the syllable and "squeezes" the vowel, represented by blue blobs).

Processing the structure of the stressed syllables provides a basis for acquiring the complex spelling rules that underlie spelling of long vs. short vowels in the German orthography. In the third chapter, children need to find out the spelling of these structures (Fig. 3c). They acquire knowledge about the rules that underlie the spelling of open and closed syllables by playing various minigames that cover the recognition of each of the special orthographic markers (c.f. [15] for a detailed description). In the fourth chapter, they learn to divide written words into relevant components (e.g., syllables) and thus foster their orthographic representation. The traditional intervention "draw syllable arcs" is commonly used to train syllable analysis. In Prosodiya, this training is enhanced by emphasizing syllable stress (children draw outstanding arcs for stressed syllables). This game is also being developed as a so-called embodied training that uses body movement and gestures. In this version, children speak each syllable clearly and loudly, simultaneously do a sidestep, and swing their writing hand from their left to their right shoulder (Fig. 3e). Fitness trackers with built-in accelerometers are used to record and classify the swung stress pattern. We refer to [16] for a detailed description about this embodied training and its implications.

In the last chapter, children foster their previously acquired knowledge by spelling out words using a predefined set of letters (Fig. 3f). This set of letters can - depending on the difficulty - contain distractors.

Table 1. The questionnaires given to children (C) and parents (P).

Question		Response options
	Children and Parents	
Q1	C: *How do you like Prosodiya?* P: —"—	awful ○○○○○ great
Q2	C: *Did you enjoy training with Prosodiya?* P: *Did your child enjoy —"—?*	not at all ○○○○○ very much
Q3	C: *Would you like to continue training with Prosodiya?* P: *Would you continue using Prosodiya for your child?*	☐ yes ☐ yes, if new games added ☐ don't know ☐ no
Q4	C: *Do you think Prosodiya helped you to improve your reading and writing skills?* P: *Do you think Prosodiya helped your child to improve their reading and writing skills?*	☐ yes ☐ don't know ☐ no
	Only Children	
Q5	C: *How did you like the graphics?*	awful ○○○○○ great
Q6	C: *How do you like our Kuggelichter?*	not at all ○○○○○ very much
Q7	C: *How did you like the tutorials?*	not at all ○○○○○ very much
Q8	C: *How difficult was Prosodiya?*	very hard ○○○○○ very easy
Q9	C: *Was the task's objective always clear?*	☐ yes ☐ sometimes ☐ no

2.1 Evaluation of a Pilot Study

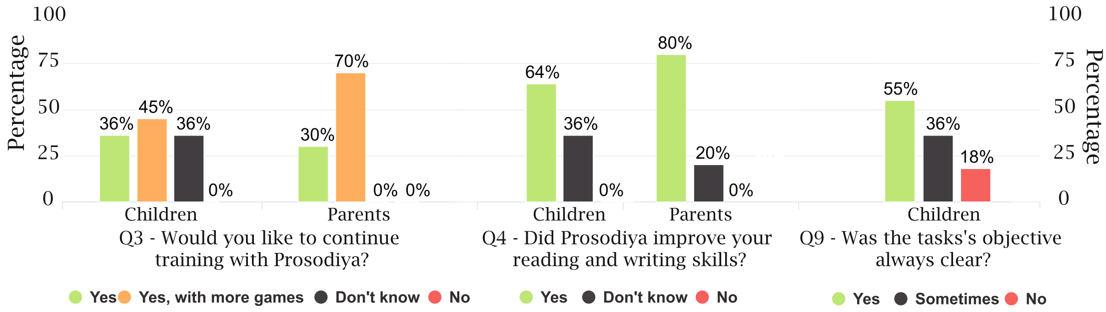
We conducted a pilot study with a preliminary version of Prosodiya in winter of 2015 with 11 dyslexic children from 2nd to 7th grade ($M = 4$, $SD = 1$) aged 7–13 ($M = 9.57$, $SD = 1.63$). Nine of the children were boys and two were girls. The study version contained the first two chapters and a restricted version of the third chapter (different orthographic markers were introduced in one and the same game – not in separate games as in the current version). In total, it consisted of 29 levels and covered 220 words. The children spent an average time of 192.6 min ($SD = 70.69$) training in-game during a period of six weeks.

We used quantitative spelling (DRT 2–5, i.a [26]) and reading tests (SLRT II [25], SLS [23]) to evaluate changes in literacy competence. Two questionnaires (see Table 1) were given to the families after the post-test to assess effects on motivation, enjoyment, and self-efficacy. One was answered by the children and one by their parents. The parents of one child didn't answer the questionnaire.

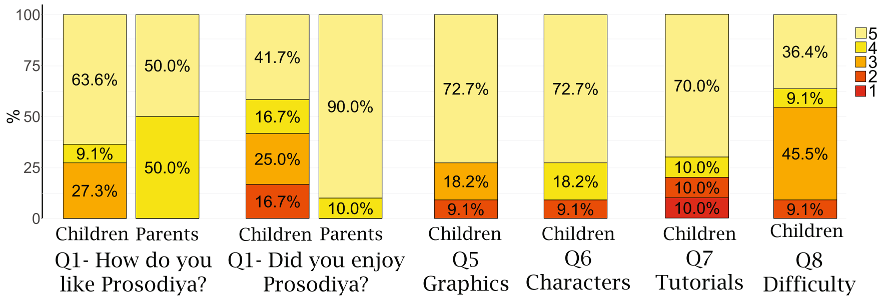
Results

Results of the questionnaires are listed in Figs. 4a and 4b.

Children and parents rated Prosodiya altogether (Q1) with an average of 4.36 and 4.5 out of 5 points. Children enjoyed playing Prosodiya (Q2, $M = 4.0$)



(a) Answers given to the multiple choice questions Q3, Q4, and Q9.



(b) Answers given to the Likert scale questions Q1, Q2, Q5, Q6, Q7, and Q8.

Fig. 4. Results of the questionnaires.

which is in line with the impression their parents reported ($M = 4.5$). Children rated the graphical appearance of the game with an average of 4.36 points (Q5) and its main characters with an average of 4.55 points (Q6). The majority of both children (72%) and parents (100%) would like to continue using Prosodiya (Q3), especially if more games are added (45%, 70%, respectively). Also the majority of both children (64%) and parents (80%) reported that they perceive self-efficacy considering reading and writing skills (Q4).

Children rated the overall difficulty of Prosodiya (Q8) with an average of 3.73 points ($SD = 1.1$) of a scale ranging from 1 (very hard) to 5 (very easy).

The majority of the children (55%) reported that the task's objective was always clear (Q9), 36% reported that it sometimes took some time to figure out the objective and 18% reported that they often had to guess what the task's objective was. One child answered this questions with both of the two latter options resulting in a total response above 100%. The tutorials (Q7) were rated with an average score of 3.82 points ($SD = 0.73$). Two children rated them very low (1 and 2 points) whereas the remaining rated them on average with 4.87 points.

In spelling tests, six children improved their performance in post-tests (DRT 2–5, i.a [26]), only two did not change in performance and one child performed worse. In speeded single word reading post-tests (SLRT II [25]), three children improved, four children did not change in performance and two children performed worse. In speeded reading comprehension tests (SLS [23]), the best improvements were obtained. Almost all children improved with up to 13 points on the fluency scale (a competence level encompasses 9 points). Only one child's performance did not change.

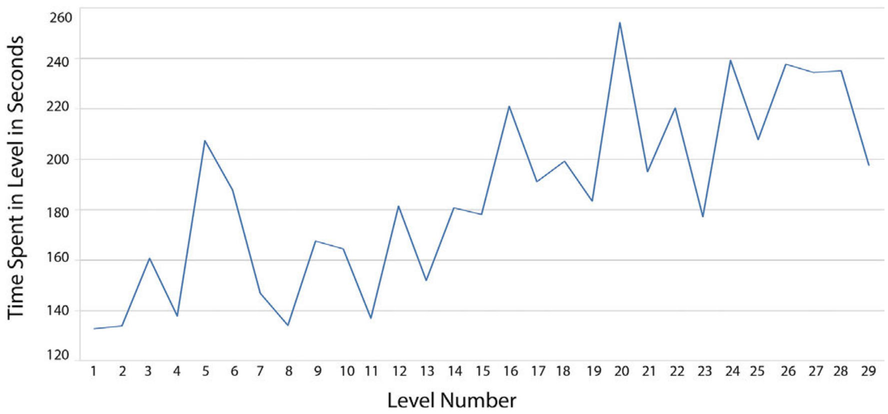


Fig. 5. Average time spent in the different levels. The curve resembles the flow channel used in game design.

Discussion

The results of the questionnaires suggest that both children and parents like Prosodiya (Q1), enjoy spending time playing it (Q2), and are waiting for more content (Q3). The graphical elements of Prosodiya seem to be appealing and match the taste of primary-school children (Q5), which is crucial for serious games [3]. Also the main characters seem to be well received (Q6).

According to the results of Q8, the perceived level of difficulty seems to be adequately between boredom (too easy) and frustration (too hard). Our goal was to keep the children within the flow-channel, a narrow band between boredom and anxiety [9]. In-game measurements of average time spent in a specific level (Fig. 5) indicate the success of our approach. As the number of tasks for each level was constant, the difference of time spent can be attributed to response times for single tasks - which in turn can be used to estimate the proficiency level of a learner [31]. Thus, the in-game metrics indicate that sequences of tension (more time spent, caused by increased difficulty) and relaxation (less time spent, caused by a higher increase of learner proficiency compared to difficulty adjustment) keep the children in a state of flow. This in turn can have positive impact on learning and player's attitude [19].

The reports to Q4 are important to the area of self-efficacy and self-esteem. Boosting both is a central aim of therapeutic intervention [2]. The perceived high self-efficacy reported by children and parents is related to self-awareness of skill increase and actual skill increase [8]. As pointed out by [3], the absence of or little self-awareness of skill increase must be avoided.

We have analysed the reports regarding the questions whether the games were self-explanatory to a satisfactory degree (Q8) and how the tutorials were rated (Q7). Although the majority answered both questions quite positively, the reports also indicate room for improvement. The reasons for negative reports can be that the tutorials didn't communicate the objective of a game well enough or the fact that sometimes the difficulty affecting game mechanics or objectives was slightly increased between levels without informing the children. We therefore derived and applied three changes: First, we refined the tutorials by splitting complex domains and addressing the respective domain in much more detail and including more examples. Second, whenever something changes that could have an impact on the children's answers and behaviour, a brief information is displayed and informs the children of the change and its consequences. Third, we developed in-game tool-tips for each game that explain the game's objective(s).

Overall, in spelling as well as in reading, approximately half of the children could improve their performance, some of them significantly. About a third did not change in performance and two children performed worse. Overall the best improvements were obtained in speeded reading comprehension tests. These results are promising, considering that our pilot's duration was only six weeks. A meta-analysis [13] has shown that interventions with a maximum duration of 12 weeks have only small effect sizes and interventions that last more than 12 weeks have a higher mean effect size. The main objective of the present pilot was primarily to investigate user experience and playability, therefore, duration was set to six weeks. But for future work, focus will be set on therapeutical efficacy and intervention phases will be prolonged.

All in all, the evaluation of the pilot study suggests that the version of Prosodiya at that time reaches a satisfactory level in terms of game and therapeutic design. However, it also highlights the limitations of that version, especially in terms of effects of the game on reading and writing performance. The evaluation of the questionnaires provide evidence that our proposed ICCGD is a valid approach for designing serious games for children. It also highlights the drawbacks of the study version that we addressed in the current version.

Limitations

The pilot study had two major limitations with respect to training effects. First, the children spent significantly less time training with Prosodiya than is recommended by [13]. To find empirical evidence for the effects on reading and writing improvement, the intervention period must be extended and intensified. Second, the games were very limited. We didn't include spelling practice. However, training phonological awareness is assumed only to have an effect on writing performance if combined with spelling exercises (cf. [30]). The crucial core component of spelling was added to the current version of Prosodoiya.

3 Conclusion and Future Work

In summary, we presented a mobile serious game for dyslexic children and the results of a pilot study of a preliminary version of the game. Prosodiya introduces various novelties in this area of research. The main contribution of Prosodiya is its unique focus on syllable stress that we derived reasonably. This focus led to the development of innovative exercises based on empirical findings. The pilot study was conducted with 11 dyslexic children. A central feedback in questionnaires was that the children liked playing the game and that they wanted more levels to keep on playing. Overall, they felt that they could improve their literacy skills. This self-assessment was not represented in quantitative reading and spelling tests in which almost half of the children improved after a period of six weeks, the other half did not show any change in competence, and two children performed worse.

To evaluate the effects of this therapy approach in an improved and most recent version of Prosodiya, a randomized control trial with a waiting control group design is planned starting January 2018. Both a group of dyslexic and unimpaired children from 2nd to 4th grade will practice in total eight weeks with Prosodiya. In the intervention phase, children should practice 20 min per day, 5 days per week. This will result in approximately 800 min of training. Besides the effect of the therapeutic approach on reading and writing, newly introduced elements of gamification will also be investigated.

References

1. Beddington, J., Cooper, C.L., Goswami, U., Huppert, F.A., Jenkins, R., Jones, H.S., Tom, B.L., Sahakian, B.J., Thomas, M., Field, J., Goswami, U., Huppert, F.A., Jenkins, R., Jones, H.S., Kirkwood, T.B.L., Sahakian, B.J., Thomas, S.M.: The mental wealth of nations. Nature **455**(7216), 1057–1060 (2008)
2. Bender, F., Brandelik, K., Jeske, K., Lipka, M., Löffler, C., Mannhaupt, G., Naumann, C.L., Nolte, M., Ricken, G., Rosin, H., Scheerer-Neumann, G., von Aster, M., von Orloff, M.: Die integrative Lerntherapie. Lernen und Lernstörungen **6**(2), 65–73 (2017)
3. Berkling, K., Faller, H., Piertzik, M.: Avoiding failure in modern game design with academic content - a recipe, an anti-pattern and applications thereof. In: Proceedings of the 9th International Conference on Computer Supported Education, pp. 25–36 (2017)
4. Berkling, K., Pflaumer, N., Lavalley, R.: German phonics game using speech synthesis - a longitudinal study about the effect on orthography skills. In: Workshop on Spoken Language Technology for Education (SLaTE), pp. 167–172 (2015)
5. Boyle, E.A., Hainey, T., Connolly, T.M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., Pereira, J.: An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. Comput. Educ. **94**, 178–192 (2016)
6. Bradley, L., Bryant, P.E.: Categorizing sounds and learning to read - a causal connection. Nature **301**(3), 419–421 (1983)
7. Brandelik, K.: Sprachrhythmische Fähigkeiten im Schriftspracherwerb. Südwestdeutscher Verlag für Hochschulschriften (2014)
8. Cleary, T.J., Velardi, B., Schnaidman, B.: Effects of the Self-Regulation Empowerment Program (SREP) on middle school students' strategic skills, self-efficacy, and mathematics achievement. J. Sch. Psychol. **64**, 28–42 (2017)
9. Csikszentmihalyi, M.: Beyond Boredom and Anxiety. The Jossey-Bass Publishers, San Francisco (1975)
10. Daniel, S.S., Walsh, A.K., Goldston, D.B., Arnold, E.M., Reboussin, B.A., Wood, F.B.: Suicidality, school dropout, and reading problems among adolescents. J. Learn. Disabil. **39**(6), 507–514 (2006)
11. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: defining "Gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011, pp. 9–15. ACM, New York (2011)
12. Fullerton, T.: Game Design Workshop: A Playcentric Approach to Creating Innovative Games, 2nd edn. Morgan Kaufmann, Heidelberg (2008)

13. Galuschka, K., Ise, E., Krick, K., Schulte-Körne, G.: Effectiveness of treatment approaches for children and adolescents with reading disabilities: a meta-analysis of randomized controlled trials. PLoS ONE **9**(2), 1–12 (2014)
14. Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., Leong, V.: Impaired perception of syllable stress in children with dyslexia: a longitudinal study. J. Mem. Lang. **69**(1), 1–17 (2013)
15. Holz, H., Beuttler, B., Brandelik, K., Brandelik, J.: Prosodiya - ein Lernspiel zur Förderung des Sprachrhythmus bei Kindern mit LRS. In: Burghardt, M., Wimmer, R., Wolff, C., Womser-Hacker, C. (eds.) Mensch und Computer 2017 - Tagungsband, pp. 395–398. Gesellschaft für Informatik e.V, Regensburg (2017)
16. Holz, H., Beuttler, B., Kirsch, A.: Bewegungserkennung mit Wearables für Embodied Trainings in Serious Games. In: Burghardt, M., Wimmer, R., Wolff, C., Womser-Hacker, C. (eds.) Mensch und Computer 2017 - Tagungsband, pp. 259–262. Gesellschaft für Informatik e.V, Regensburg (2017)
17. Ise, E., Schulte-Körne, G.: Spelling deficits in dyslexia: evaluation of an orthographic spelling training. Ann. Dyslexia **60**(1), 18–39 (2010)
18. Kast, M., Baschera, G.M., Gross, M., Jäncke, L., Meyer, M.: Computer-based learning of spelling skills in children with and without dyslexia. Ann. Dyslexia **61**(2), 177–200 (2011)
19. Kiili, K., de Freitas, S., Arnab, S., Lainema, T.: The design principles for flow experience in educational games. Proc. Comput. Sci. **15**, 78–91 (2012)
20. Klicpera, C., Schabmann, A., Gasteiger-Klicpera, B.: Legasthenie - LRS, 4th edn. Reinhardt, München (2013)
21. Landerl, K.: Categorization of vowel length in German poor spellers: an orthographically relevant phonological distinction. Appl. Psycholinguist. **24**(4), 523–538 (2003)
22. Liebal, J., Exner, M.: Usability für Kids. Vieweg+Teubner, Wiesbaden (2011)
23. Mayringer, H., Wimmer, H.: SLS 1–4 - Salzburger Lesescreening für die Klassenstufen 1–4. Verlag Hans Huber, Bern (2003)
24. Moll, K., Landerl, K.: Double dissociation between reading and spelling deficits. Sci. Stud. Read. **13**(5), 359–382 (2009)
25. Moll, K., Landerl, K.: SLRT II - Lese- und Rechtschreibtest. Verlag Hans Huber, Bern (2010)
26. Müller, R.: Diagnostischer Rechtschreibtest für 2. Klassen (DRT 2). Beltz, Göttingen (2003)
27. Reuter-Liehr, C.: Behandlung der Lese-Rechtschreibschwäche nach der Grundschulzeit: Anwendung und Überprüfung eines Konzeptes. Zeitschrift für Kinder- und Jugendpsychiatrie **21**(3), 135–147 (1993)
28. Sauter, K., Heller, J., Landerl, K.: Sprachrhythmus und Schriftspracherwerb. Lernen und Lernstörungen **1**(4), 225–239 (2012)
29. Schulte-Körne, G., Remschmidt, H.: Legasthenie - Symptomatik, Diagnostik, Ursachen, Verlauf und Behandlung. Dtsch Arztebl International **100**(7), 396–406 (2003)
30. Schulte-Körne, G., Galuschka, K.: Diagnostik und Behandlung von Kindern und Jugendlichen mit Lese- und / oder Rechtschreibstörung (2015)
31. Sense, F., Behrens, F., Meijer, R.R., van Rijn, H.: An individual's rate of forgetting is stable over time but differs across materials. Top. Cogn. Sci. **8**(1), 305–321 (2016)
32. Snowling, M.J.: Phonological processing and developmental dyslexia. J. Res. Reading **18**(2), 132–138 (1995)
33. Staffeldt, S.: Einführung in die Phonetik, Phonologie und Graphematik des Deutschen Ein Leitfaden für den akademischen Unterricht. Stauffenburg (2010)

How to train your syllable stress awareness – A digital game-based intervention for German dyslexic children

Heiko Holz[1,2] (✉), Katharina Brandelik[2], Benedikt Beuttler[2,3],
Jochen Brandelik[2], Manuel Ninaus[1,4]
[1]*LEAD Graduate School & Research Network, University of Tuebingen, Germany*
heiko.holz@uni-tuebingen.de
[2]*Tübinger Institut für Lerntherapie, Germany*
{k.brandelik, b.beuttler, j.brandelik}@til-lerntherapie.de
[3]*Department of Psychology, University of Tuebingen, Germany*
[4]*Leibniz-Institut für Wissensmedien, Tuebingen, Germany*
m.ninaus@iwm-tuebingen.de

Abstract

Developmental dyslexia is one of the most frequent learning disorders and affects 4-10% of the German population. The learning disorder affects educational, personal, and social development of children in a negative way. We examine three different approaches to treat learning disorders. That is, therapeutic, computer-based, and digital game-based interventions. We reflect on the advantages and disadvantages of these approaches that have been shown to be supportive for dyslexic primary-school children. Our literature review shows that there is a lack of digital game-based interventions for the treatment of spelling disorders. To fill this gap, we propose such a mobile serious game which uses evidence-based trainings and introduces novel features in order to help dyslexic children to improve their reading and spelling performance. We propose an intervention to train awareness of syllable stress and explore the innovative use of mouth motor activities and embodied trainings. To conclude, we suggest that, in addition to traditional approaches, digital-game based approaches should be used supplementary to (re-) engage and motivate learners.

Keywords: *Digital game-based learning, Dyslexia, Spelling, Syllable stress awareness;*

1 Introduction

Approximately 4-10% of the German population suffer from developmental dyslexia [1–3]. Dyslexia is therefore one of the most frequent learning disorders [2]. It affects about 50.000 German children of each birth cohort.

These children acquire reading and writing skills in a much slower pace and not as proficient compared to their classmates [4] and they suffer massively from their impaired literacy acquisition. Usually, they lose their motivation for the learning process as well as the faith of being able to develop a comprehension of literacy language [5]. Appropriate treatments and interventions are necessary to prevent negative consequences, such as poor performance in school and higher chances to drop out from school, resulting in few employment prospects or unemployment [6]. Overall, the learning disorder has a negative impact not only on mental health but also on social and cultural participation [7].

On the one hand, it has been shown empirically that computer-based interventions can help dyslexic children to significantly improve reading and spelling skills (e.g., [8–12]). Interventions primarily address the symptoms of reading and writing disorders. Further, by the usage of game elements, such as narratives or rewards, they explicitly address issues related to frustration, demotivation or boredom [13] and support successful learning outcomes (cf. [14]). Mobile serious games can additionally help children to overcome their learning disorder outside of learning therapy and classrooms – independent of location and time. Importantly, mobile serious games for dyslexic children have been proven to have positive effects on the process of literacy acquisition (e.g., [15–17]).

However, there is a lack of evidence-based and empirically evaluated digital game-based interventions for the treatment of spelling disorders in German dyslexic children.

In this article, we present such a mobile serious game called "Prosodiya". The evidence-based trainings of Prosodiya contribute three innovative features to the state of the art of computer-based interventions. First, it is the first digital therapy approach that focuses on improving the awareness of syllable stress and associates the stressed syllable's linguistic features to orthographic principles of the German orthography. It is thought that dyslexic children lack in accurate stress perception and thus fail to identify orthographic markers, which are particularly important in stress-timed languages such as German. Consequently, adequate processing verbal stress might support children acquiring complex spelling rules in German orthography (cf. Section 2.1). Second, mouth motor activities are introduced to support children that struggle to perceive phonological features. Lastly, the use of embodied training for mobile serious games using sensor-based gesture recognition are explored.

This article is structured as follows: First, we reflect the state of research of the relation between syllable stress and reading and spelling skills. This is followed by an overview of the different areas of reading and spelling acquisition and evidence-based trainings to address these areas. We then reflect advantages and disadvantages of three different approaches used for treatment of reading and spelling disorder, i.e. therapeutic, computer-based, and digital game-based interventions. In the fourth part of this article, we describe the current game and therapeutic design of Prosodiya. We conclude this article by reflecting the contributions of Prosodiya to the state of the art and by providing an outlook for the project's future progress with special focus on the opportunities of embodied trainings.

2 Related work

2.1 Syllable stress and literacy acquisition

Deficient phonological awareness is, among others, one major cause of dyslexia [18]. Phonological awareness refers to the ability to deal with the sound system of a language and to detect, distinguish, and manipulate segments of a language, such as syllables, rimes, or even single sounds (i.e. phonemic awareness).

Phonological awareness also includes the perception of prosodic features. A shortcoming in the perception of prosodic features is a strong predictor for developmental dyslexia [19–21]. One of these features is syllable stress, an important characteristic of German speech rhythm. German belongs to the category of stress-timed languages [22]. In stress-timed languages, speech rhythm is generated by the interval between two consecutive stressed syllables, which is perceived to be a fairly constant amount of time. Stressed syllables have larger rise times (the time required to reach peak signal intensity) in the amplitude envelope and are generally perceived to be louder. In contrast, unstressed syllables are compressed and reduced to fit the rhythm. Other typical stress-timed languages are, among others, English, Russian, Germanic languages, and European Portuguese. The rhythm of stress-timed languages differs to, among others, syllable-typed languages, such as Italian, French, or Spanish, in which the duration of every syllable is approximately constant in time.

Stress perception has been shown to be impaired for developmental dyslexics [19], [20, 23] and highly correlates with reading and writing skills [21, 24]. For German dyslexic children, one explanation is thought to be found in the association between stress and German orthographic markers. Orthographic markers – graphemes that mark vowel lengths – generally occur within stressed syllables [markers for long vowels, such as B*ie*-ne (bee)] or in conjunction with stressed syllables [markers for short vowels, such as Ge-wi*t*-*t*er (thunderstorm)] [25]. Since the vowels of unstressed syllables are reduced, such as the schwa sound (/ə/) in /ˈmaːlən/ (malen, to paint), they are generally not marked orthographically.

The orthographic marking of short and long vowels, which underlies complex orthographic rules, is a major challenge for German dyslexic [26, 27]. Therefore, processing verbal stress adequately might help children to acquire the complex spelling rules that underlie vowel length spelling in German orthography.

Although research suggest that rhythmic trainings for poor English readers using, among others, exercises to match the correct syllabic stress pattern to words to be beneficial for the development of literacy and phonological awareness [28, 29], we are not aware of any (mobile) serious game that focusses on improving the awareness of syllable stress or associates the stressed syllable's linguistic features to orthographic principles of the German orthography.

2.2 Evidence-based treatment of reading and spelling disorders

According to the clinical guideline of diagnosis and treatment of German dyslexic children [30], the treatment of reading and spelling disorders can be divided into different areas. The authors infer recommendations for evidence-based trainings for each of these areas from the results of a meta-analysis [31, 32] of randomized controlled field trials.

The areas of treatment of reading disorders comprise: i) syllabic and phonemic awareness, i.e. awareness of syllable and sounds, ii) reading accuracy, iii) reading fluency, and iv) reading and text comprehension. For the first stage, they recommend trainings to identify, categorize, segment, delete, or discriminate syllables and sounds within words. The second stage includes systematic instructions of letter-sound correspondences and exercises of phoneme synthesis. In this regard, derivational synthesis refers to blending (pulling together) individual parts of a language within words, e.g. blending individual sounds or syllables to words. Reading fluency should be trained with systematic exercises of phoneme, syllable, and morpheme synthesis. Lastly, reading comprehension includes interventions that comprise tasks in which participants learn to extract textual information, summarize it, and relate it to existing knowledge.

The areas of treatment of spelling disorders are: i) syllabic and phonemic awareness, ii) phoneme-grapheme correspondence, iii) grapheme memory entries, and iv) knowledge of rules and morphemes. The first stage of spelling acquisition equals to the first stage of reading promotion and thus provides the same interventional recommendations. For the second stage, phonics instruction (systematic instructions of letter-sound correspondences and exercises of phoneme analysis at the (sub-) lexical level) are recommended. For the latter, analysis refers to segmenting words into respective parts, e.g. phonemes or syllables. For the third stage, systematic exercises for storing and remembering frequent sequences of graphemes are recommended. Finally, exercises to acquire orthographic and morphemic regularities are recommended for the fourth stage of the treatment of spelling disorder.

The guideline [30] concludes that reading skills can be most effectively improved with systematic instruction of letter-sound correspondences and phoneme, syllable, and morpheme synthesis. Spelling performance can most effectively be improved by using systematic instructions of letter-sound correspondences and exercises analyzing sounds, syllables, and morphemes as well as trainings enabling the acquisition and generalization of orthographic regularities.

In the remainder of this article, we will mainly focus on interventions and trainings for German dyslexic primary school children. This is due to the fact that differences in orthographic consistency between languages, i.e. differences in regularity and consistency of letter-sound relations, result in differences in reading and spelling difficulties of dyslexic children and consequently the correct therapeutic approach [33–35]. German, due to regular letter-sound correspondences, has a high transparency of the writing system compared to languages with opaque letter-sound relations, such as English.

3 From therapeutic interventions to digital game-based learning

Treatment of reading and spelling disorders yields different interventional approaches that can be applied. In this chapter, we take a closer look at therapeutic, computer-based, and digital game-based interventions for dyslexic children. By highlighting advantages and disadvantages of respective approaches, we deduce our general recommendation that therapeutic and digital-game based interventions should be used in the following way:
Dyslexic children should receive evidence-based therapeutic interventions as soon as possible and as long as necessary until their reading and spelling proficiency enables them to participate in social, cultural, and educational life on an age-appropriate level. In addition, they should use evidence-based digital game-based interventions to maximize engagement, motivation, and learning.

3.1 Therapeutic interventions

The common approach to treat children suffering from learning disorders are therapeutic interventions (TI) administered in individual or group sessions. Therapeutic interventions are carried out by trained practitioners – such as teachers or learning therapists – in learning facilities, usually outside of school time (see Table 1).

If children participate in therapeutic interventions implementing aforementioned evidence-based approaches at least weekly for several months, substantial improvements in reading (e.g., [36–38]) and spelling (e.g., [9, 38, 39]) performance can be achieved.

Although children with reading and spelling disorders may show significant improvements after several weeks or months, they should receive support until their ability to read and spell reaches a level that enables them to participate in public life in an age-appropriate way [30]. This means in most cases several years of intense support and treatment. However, this may often not be provided if the healthcare system has no provision for funding [30]. This is the case in Germany – learning therapy is not covered by health insurances but by the youth welfare office. Applying for financing and reimbursement of learning therapy can be a tedious process disadvantaging families who cannot afford to pay for learning therapy privately. Thus, affected children may not receive appropriate treatment timely, sustainably, or long enough.

Importantly, it is strongly recommended that interventions should be implemented by experts in reading and spelling development and its promotion [30] rather than by peers, parents, or university students – which might come to mind to affected families as a cost-effective alternative but whose effectiveness could not be confirmed unequivocally [31, 40].

To conclude, evidence-based therapeutic interventions are reliable and recommended treatments for dyslexic children when administered by experts but are cost-intensive and might not be provided timely or long enough.

3.2 Computer-based interventions

Computer-based interventions (CBI) and the use of information and communication technology (ICT) have been shown to benefit learning-impaired children. By complementing therapeutic interventions and traditional teaching, they address the aforementioned disadvantages of TI and offer new opportunities to engage the learner (se also Table 1). Based on literature reviews, we conclude that CBI and ICT can have positive effects on dyslexic children's learning development [41] and that they facilitate literacy acquisition for dyslexic children by minimizing difficulties in learning to read and write [42].

First, one of the major advantages of CBIs is that they can be used independent of time and place, outside of class or learning therapy.

Table 1. *Advantages and disadvantages of and exemplary programs for therapeutic, computer-based, and digital game based interventions for German dyslexics.*

		Advantage	Disadvantage	Recommendation	Evidence-based interventions: Focus	Program	Eval.
Human tutor	**TI** Therapeutic and traditional interventions	• Reliable and effective • Usually covers all processes of learning or spelling acquisition	• Requires experts of reading and spelling development and promotion as tutors • Cost-intensive • May not be provided in a timely manner or long enough • Effectiveness depends on tutor • Dependent of time and location	• Timely and long-running treatment until learner can participate in social, cultural, and educational life on an age-appropriate level • Incorporate methods of (digital) game-based learning	Reading	Kieler Leseaufbau [37]	+
						Flüssig lesen lernen [36]	+
						PHONIT [43]	(+)
					Spelling	Lautgetreue Lese-Rechtschreibförderung [38]	+
						Marburger Rechtschreibtraining [39]	+
Digital	**CBI** Computer-based interventions	• Independent of time and place • Can be used autonomously by children • Offers more interactive experiences • Increases engagement and motivation • Continuous assessment • Adaptivity w.r.t. individual needs	• Motivation and engagement of learners may not last • Single programs usually do not cover all processes of literacy acquisition • Effectiveness depends on adaptivity	• Try to use DGBI if available • Use complementary to TI	Reading and Spelling	Lautarium [10]	+
					Spelling	Morpheus [9]	+
	DGBI Digital game-based interventions	• Same advantages as CBIs • Motivate and engage learners over long periods • Promote engagement of learners with special needs • Reengage learners who have disengaged from learning (in the respective domain)	• Single programs usually do not cover all processes of literacy acquisition • Effectiveness depends on adaptivity • Game design and learning objectives must be carefully balanced	• Use complementary to TI • Use to reengage children who lost their interest or motivation in learning	Reading	‘ Lernspiele mit Elfe und Mathis’ [12]	+
					Spelling	Phontasia [15]	(+)
					Reading + (Spelling)	Meister Cody – Namagi [44]	–
					Spelling	Prosodiya [45]	(+)

Second, if designed properly, learners can use CBIs autonomously for homework complementary to school and learning therapy without a real tutor. In fact, dyslexic children may respond to explicit computer instructions in writing as well as they have with human teachers [46].

Third, research has shown that children may concentrate better while engaged in CBI than in traditional school tasks [47]

Fourth, nowadays, ICT can provide interactive experiences which can motivate children at an early age and attenuate the impact of their own difficulties in the daily acquisition of reading and spelling skills [42].

Fifth, CBIs may offer continuous and more frequent assessment of proficiencies and knowledge [48, 49] compared to TI and save time in administration and evaluation of tests.

Through continuous assessment, CBIs are able to recognize individual needs of dyslexic children, which has been emphasized to be important for automatically adapting a CBI [50] – adaptivity is a common requirement posed to CBIs.

Lastly, gamification plays a major role in CBI. Gamification as defined by Deterding et al. [13] is *the use of game design elements in non-game contexts*, such as narratives, scores, or rewards. The role of gamification is primarily in invoking the same psychological experiences as games generally do [51]. Based on the literature review of Hamari et al. [52], gamification in educational and learning contexts mostly positively affects learning. Most importantly, it increases motivation, engagement in, and enjoyment of learning tasks [52].

Lautarium [10] is an adaptive computer-based intervention program for primary-school. It includes training to improve phonological awareness in the narrow sense (e.g., phonemic awareness, phoneme analysis and synthesis, phoneme classification) as well as letter-sound correspondence to foster pronunciation-true reading and spelling. The authors of Lautarium could show empirically in a waiting control group design its efficacy on phonemic and phonological awareness and pronunciation-true reading and spelling [10].

Morpheus [9] is a computer-based spelling training that focuses on improving morphemic awareness, i.e. the ability to recognize, understand, and use morphemes. Morpheus trains orthographic regularities for the categories of consonant doubling, silent h, and German vowel doubling *ie*. It additionally trains orthographic phenomena that may not be deduced from phonetic principles. It has been shown empirically in a control-group design that Morpheus significantly improves morphemic spelling strategies [9].

3.3 Digital game-based interventions

Digital game-based interventions (DGBI) are the top tier of digital interventions for children with learning disorders. DGBI are extensions to CBI in that they include all benefits of CBIs (as they belong to the same technology, namely digital) but address CBIs' drawback of not fully exploiting the engaging and motivational potential of digital games (see Table 1).

Crucially, we differentiate between gamified CBI and DGBI as follows: gamified interventions merely incorporate elements of games while game-based learning or serious games describe the design of full-fledged games for educational purposes [13] that focus on designing activities as playful tasks [53]. We additionally refer to the definition of Wouters et al. [54] of digital games to be interactive, based on a set of agreed rules and constraints, directed toward a clear goal that is often set by a challenge, and constantly providing feedback either as a score or changes in the game world to enabled self-monitoring of progress towards the goal.

As gamified CBIs are often advertised as educational *games*, it is important to highlight that this claim may result in negative consequences due to the expectations posed to educational games: Parents and children consider enjoyment as one of the central principles important in educational software that is used in the home environment [55] and sometimes even prioritize enjoyment above educational benefits. As DGBIs are expected to be engaging and fun and thus naturally motivating [55, 56], the motivational design of DGBIs used in the home environment is of crucial importance, whereas learning effectiveness is referred to the most essential aspect at school [56]. Therefore, labeling CBIs as games may fail to live up to these expectations if the use of game is limited to gamification.

3.3.1 The advantages of digital game-based interventions

The benefits of game playing as a learning process are widely acknowledged [57, 58]. DGBIs have been found to be effective or even outperform conventional instruction methods in terms of learning and retention, such as lectures, reading, drill and practice, or hypertext learning environments [54]. In fact, this is particularly the case for language learning as highlighted by the meta-analysis of Wouters et al. [54]. DGBIs are able to promote engagement and learning for children with special learning needs [59] and may boost children's engagement with literacy activities, foster skill reinforcement, and enhance the perception of reading progress [60] – DGBIs may even reengage learners who disengage from learning, i.e. learners who lost interest, motivation, and engagement in learning and cannot be engaged with other methods [61, 62] (see Table 1). DGBIs are also especially suited to foster learning through embodied cognition, i.e. mapping of gestures or movement to key features of the content to be learned [53]. For example, in a Kinect-based literacy game, using gestures and movements in in-game activities enhanced literacy outcomes of children compared to a group without these activities [63]. The concept and implication of motion-based trainings is discussed more thoroughly in Section 5 .

Importantly, due to fact that DGBIs focus on defined learning outcomes, balancing educational effectiveness and quality of learning with game play is a corollary to the design process of DGBIs [53, 59, 64, 65]. If this is accounted for, DGBIs are able to engage learners on an affective, behavioral, cognitive, and sociocultural level in ways few other learning environments are able to do – as emphasized by the integrated design framework of game-based and playful learning by Plass et al. [53].

Phontasia [66] is a game-based German phonics trainer available for iPads. It is derived from the phonics method for English spelling acquisition and is adjusted for German. Children learn orthographic regularities by changing a defined set of graphemes to form a word. By providing artificial text-to-speech synthesis, children can listen to their solutions. This way, misspellings and correct spellings can be made audible for each of the children's solutions and children may reflect on their answer. Preliminary results of two evaluations indicate that Phontasia has positive effects on the spelling of consonant doublings and the German vowel doubling *ie* [15, 67].

'Lernspiele mit Elfe und Mathis' [12] is a computer-based game for reading acquisition. Its content is categorized in linguistic levels of phonemes and syllables, words, sentences, and texts. Various exercises cover, among others, letter-sound correspondence, syllable synthesis, syllable analysis, and reading comprehension. Elfe was evaluated empirically in a randomized waiting control group design [12]. The authors could show significant effects of the training on reading performance on the level of words, sentences, and texts.

Meister Cody – Namagi [44] is an evidence-based DGBI for reading acquisition. It offers exercises to improve phonemic awareness (e.g., vowel-length distinction, recognition of initial and final sounds), phonological awareness (e.g., syllable analysis and synthesis), letter-sound correspondences, and semantic linking of words with pictures. Exercises to improve spelling are currently under development. Namagi has not yet been empirically evaluated.

3.4 Categorization of computer- and digital game-based interventions for German dyslexic primary school children

In addition to Table 1, we categorized the evidence-based computer- and digital game-based interventions discussed in the previous section with regards to the areas of the treatment of reading and spelling disorders in Table 2. An intervention receives a '+' for a specific area if it contains explicit instructions and exercises, '(+)' if exercises may implicitly support the area, and '–' if the intervention does not include exercises to promote a specific area of reading or spelling. Combining Table 1 and Table 2, parents and practitioners may select interventions according to a child's need.

Table 2. *Categorization of popular evidence-based interventions according to the areas of the treatment of reading and spelling disorders (derived from [30]).*

App	***Treatment of reading disorder***				***Treatment of spelling disorder***				***Eval.***
	Awareness of syllables and sounds	Reading accuracy	Reading fluency	Reading/text comprehens.	Awareness of syllables and sounds	Letter-sound correspond.	Memory retrievals	Knowledge of rules and morphemes	
Computer-based interventions									
Lautarium [10]	(+)	+	+	–	+	+	–	(+)	+
Morpheus [9]	–	(+)	(+)	–	–	–	+	+	(+)
Digital game-based interventions									
Elfe [12]	+	+	+	+	+	(+)	+	(+)	+
Phontasia [15]	(+)	–	+	–	(+)	+	+	+	(+)
Namagi [44]	+	+	+	–	+	+	–	-	–
Prosodiya	+	(+)	(+)	–	+	+	+	+	(+)

3.5 Intermediate conclusion

Therapeutic, computer-based, and digital game-based interventions play a major role in the treatment of reading and spelling disorders. Each approach comes with advantages and disadvantage as discussed above. This chapter is summarized in Table 1 and Table 2 by providing a list of advantages, disadvantages, exemplary programs that incorporate evidence-based approaches for the treatment of reading and spelling disorders, and practical recommendations.

It is important to state that CBIs and DGBIs should be used as a supplementary tool within or outside class in assisting dyslexic children but may not replace traditional teaching and learning strategies or therapeutic interventions [11, 68].

To conclude this chapter, we recommend to treat dyslexic children with evidence-based therapeutic and traditional interventions as soon as possible and as long as necessary until their reading and spelling proficiency enables them to participate in social, cultural, and educational life on an age-appropriate level. Additionally, we recommend the integration of DGBIs in TIs and the use of DGBIs complementary to TIs, outside of class, to (re-) engage and motivate learners in ways withheld to TIs.

As can be drawn from Table 1, evidence-based and empirically evaluated DGBIs for the treatment of spelling disorders are missing for German dyslexic children. With Prosodiya, we intend to fill this gap.

4 Prosodiya

We define Prosodiya as a DGBI based on the definition of Plass et al. [53] that it's a game for an educational purpose and focusses on designing activities as playful tasks, i.e. carefully concerning the balance of the design of learning objectives and game play. The most important arguments for DGBIs as described by Plass et al. [53] are *motivation*, *engagement*, *adaptivity*, and *graceful failure*. To address these arguments researchers mostly agree on the following building blocks of DGBIs: game mechanics, visual aesthetic design, narrative design, incentive system, and content and skills [53].

Accordingly, we use these buildings blocks to analyze and describe our game in the following sections after providing a general description of Prosodiya's rationale.
Prosodiya is based on recent empirical findings and on evidence-based interventions (e.g., [39, 69]). For example, a main component is training phonological awareness in the broader sense (syllable awareness) as well as in the narrow sense (phonemic awareness) as children with dyslexia often struggle with this basic skill [18, 70].

Prosodiya builds on the research reflected in Section 2.1 and primarily aims at training detection and production of syllable stress – a feature that is not found in other serious games. First, this might improve a child's ability to segment words into relevant components. Second, children learn to focus on relevant areas in words. Prosodiya is played on tablets with a touchscreen interface. It aims at clarifying the association between syllable stress and orthographic marking. It helps children to focus on the stressed syllable and to learn how such syllables are spelled. In doing so, it finally leads to a rule-based orthographic spelling training. Prosodiya includes six chapters, each corresponding to a different linguistic or orthographic challenge.

The focus of Prosodiya is primarily on spelling acquisition by training the awareness of linguistic features related to syllable stress and by linking these features to orthographic regularities of German orthography. However, most exercises also cover skills that contribute to promotion of reading. Thus, our pedagogical approach enabling acquisition and application of orthographic knowledge is in line with empirical research emphasizing that acquisition and generalization of orthographic regularities is an essential part of effective methods to improve spelling performance [31, 39, 71, 72].

The game's overall narrative revolves around little inhabitants called "Kugellichter" (spherical lights), which seek for help of the children: A mysterious fog has conquered the land of Prosodiya which causes the inhabitants to live in sadness, see Figure 2a. Only the children, with the help of the Kugellichter, can disperse the suppressing fog by mastering linguistic challenges. Each time the children make progress within the course of the game, parts of Prosodiya are saved and new regions with new challenges call for their help.

The difficulty of Prosodiya increases on three individual levels: First, the game's chapters cover different linguistic or orthographic skills that range from syllable stress awareness to applying spelling rules. Second, subchapters within a chapter target different linguistic or orthographic sub-competences, which is explained in the following sections. Third, the levels increase in difficulty in that the target words' structures become more complex and objective(s) of tasks become more challenging by continuously decreasing the support provided to children to solve an exercise. By automatically increasing and decreasing the difficulty on individual levels, Prosodiya accounts for one of the four arguments of DGBI, namely adaptivity.

In the following, we give a detailed description of Prosodiya's game mechanics, content and skills to be learned, visual aesthetic design, narrative design, and its incentive system.
Short video clips of Prosodiya are listed on youtube.prosodiya.com. A general overview of the therapeutic structure of the first module of Prosodiya is displayed in Figure 1.

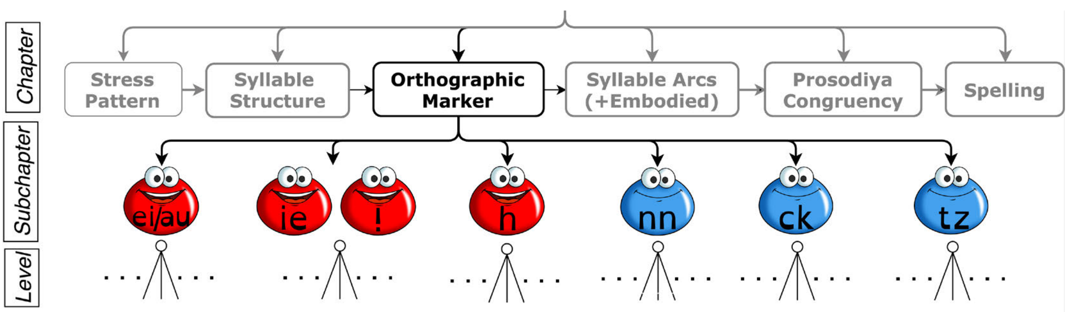
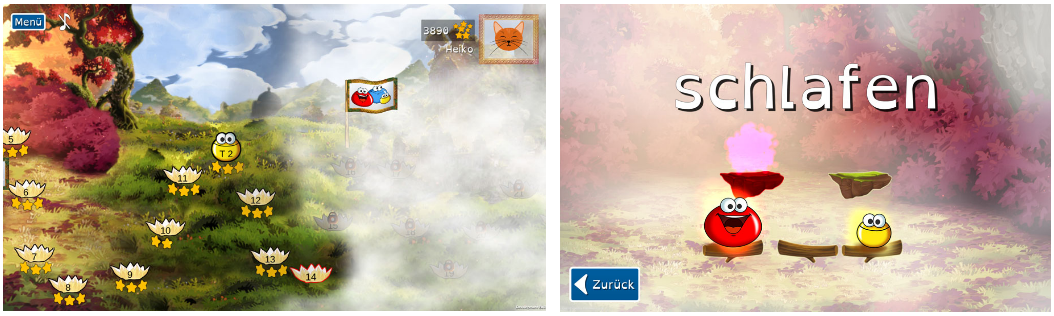


Figure 1. *Overview of the interventional structure of the first module of Prosodiya.*

As the first module of Prosodiya attributes orthographic regularities to the basic (uninflected) form of words, morphologic inflection, i.e. conjugation and declension, is not yet covered. Morphologic training, i.e. to learn to deduce that gewinnt (he wins) is spelled with an ambisyllabic consonant doubling as it's derived from gewinnen (to win), is part of the second module of Prosodiya currently being developed.

4.1 *Game mechanics and content and skills*

We combine two game design elements, i.e. game mechanics and content and skills, in one section as the latter game design element is significantly responsible for the design of game mechanics. As Prosodiya's primary learning goal is to improve spelling, the game serves the function to teach new knowledge and skills as well as practice and reinforce existing knowledge and skills. The game mechanics have been carefully designed and refined based on the results of usability tests to ensure high quality of learning and gaming experience. All game mechanics are explained in interactive tutorials, see Figure 2b.



a. *In-game map of Prosodiya.* b. *Tutorial of "Syllable Structure".*

Figure 2. *In-game map of Prosodiya and exemplary tutorial of the game "Syllable Structure".*

4.1.1 Stress pattern

The game starts by introducing the children to the concept of syllable stress. Characterizations of stressed (larger rise-times and amplitude, non-reduced vowels) and unstressed syllables (reduced vowels) are explained in tutorials. Children are taught to apply this knowledge by rebuilding stress patterns of words to develop awareness for relevant phonological features. They do so by dragging and dropping cartoon blobs (big green blob for stressed syllables, small yellow blobs for unstressed syllables) onto their respective platforms, see Figure 3. This first task aims at drawing a child's attention to the stressed syllable and to the area of the word in which orthographic markers occur, respectively. This is the most innovative training exercise of Prosodiya with respect to state of the art interventions, and it specifically addresses the shortcomings of dyslexic children. That is, improving the perception and production of syllable stress, which is not part of any empirically evaluated or computer-based intervention for German dyslexic children. We have opted to use the interaction of dragging and dropping the blobs instead of other gaming activities, such as a multiple choice task, as this best directs finger and eye movement of children to the area of interest, i.e. stressed syllables.

During the course of Prosodiya's first chapter, we adjust the difficulties of different parameters to explicitly address syllabic and phonemic awareness as well as syllable synthesis, analysis, and segmentation. Parameter combinations that lead to a training of a specific linguistic competence are listed in Table 3.

The game begins with the easiest configuration displayed in Figure 3a: words are presented in spoken form, the number of syllables is exposed by only providing as many platforms as the word's syllable count, and the word is displayed syllabified in written form. Thus, this configuration provides training of syllable synthesis (blending the presented syllables into words) and syllable and phonemic analysis (identifying the stressed syllable). By displaying words non-syllabified, syllable segmentation is addressed additionally.

Table 3. *Linguistic competences trained in the game "Stress Patterns". Different parameter settings address different competences in different degrees.*

Parameter / Linguistic competence	Word is spoken		Visual presentation			Number of syllables	
			written		image		
	yes	*no*	*syllabified*	*non-syllabified*		*exposed*	*not exposed*
Awareness of syllables and sounds	+	+	+	+	+	+	+
Syllable synthesis	–	+	+	+	+	+	+
Syllable segmentation	+	+	–	+	+	–	+

In case the word is not presented in spoken form (Figure 3b), we further aim at training syllable synthesis. By providing more platforms than the word has syllables, children are also required to count syllables autonomously (Figure 3c). As the level of difficulty and complexity changes during the course of these exercises, we provide mini-tutorials and tooltips that clarify explicitly what the children need to do and what they need to take into account in order to solve an exercise. An exemplary tooltip can be seen in Figure 4b.

We provide three different sound files for each word that increase with regard to the intensity of intonation. If a wrong answer is submitted or children request help, the word is spoken in the next stronger intonation level to give scaffolding feedback.

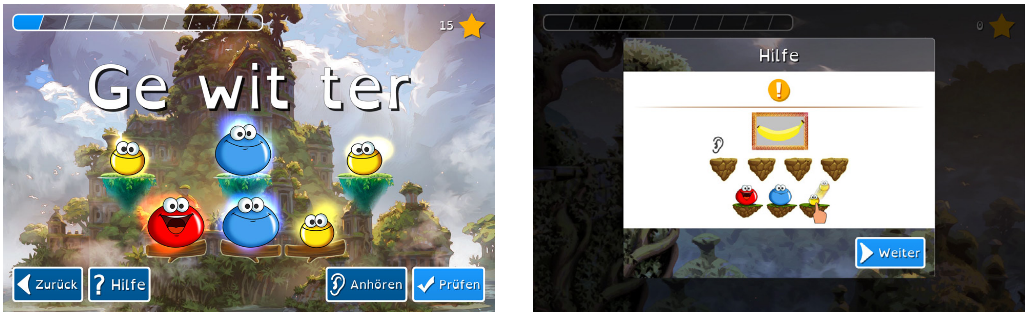
a. *Easy difficulty level.* b. *Medium difficulty level.* c. *Hard difficulty level.*

Figure 3. *Different difficulty levels of Prosodiya's first chapter "Stress Pattern" for the word Bie-ne (bee). a) Syllabification as well as number of syllables are exposed to the learner. b) The word is no longer presented in written form but number of syllables is exposed. c). Neither number of syllables nor syllabification are exposed to the learner, requiring syllable segmentation to solve the exercise.*

4.1.2 Syllable structure

The second type of games provided in Prosodiya focusses on syllable structure and builds upon the competence of stress pattern recognition. First, children need to detect the stressed syllable. Then, they have to decide whether the stressed syllable is open (ends with a long vowel) or closed (ends with a consonant). Here, we introduce awareness of mouth motor activity. Children learn that at the end of open syllables, the mouth is open: they can lengthen the vowel, thus keeping the mouth open. At the end of closed syllables however, the mouth is closed: a consonant "stops" the vowel and "squeezes" it, the mouth is closed at the lips, the teeth, or by the tongue. Depending on these structures, there are different orthographic markers [25]. Therefore, processing the structure of the stressed syllable is a base for acquiring the complex spelling rules that underlie spelling of long vs. short vowels in German orthography. Figure 4a displays a sample exercise of the game's second chapter.

Trainings to distinguish vowel length are commonly used in empirically evaluated interventions of reading and spelling acquisition and is not by itself a novelty. However, we introduce two novel features to computer-based treatments that have not been addressed before: First, we link the linguistic characteristics of syllable stress to vowel length. Second, we use the mouth features of the blobs and the terms of open vs. closed syllables instead of long vs. short vowels in order to support mouth motor activity.



a. *“Syllable Structure”: Gewitter (thunderstorm).*

b. *Exemplary tooltip of “Syllable Structure”.*

Figure 4. *“Syllable Structure”. a) Red blobs with mouths open refer to open stressed syllables (i.e., end with vowels) and blue blobs with mouths closed refer to closed stressed syllables (i.e., end with consonants). b) Shows a tooltip for this game. Children are told to i) find the open/closed syllable, ii) that the word is read out and represented by an image but not written, and iii) that they have to count the syllables autonomously.*

Children with dyslexia have difficulties permeating the sound level of a language in order to improve letter-sound correspondence on the segmental level [73]. Mouth motor activity, which is intact for dyslexic children [4], can be used to facilitate learning of letter-sound correspondence [74]. By teaching the children to pay attention to their mouth motor activities at the end of the stressed syllables, we aim at facilitating the process of vowel length distinction.

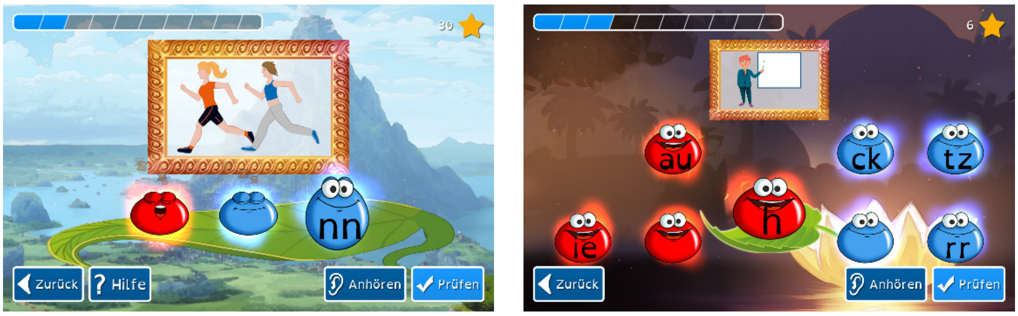
We also provide sound files of minimal pairs for each word. These minimal pairs consist of the word itself and the pseudoword counterpart for which the vowel length of the stressed syllable was changed to the contrary, i.e. long vowels are spoken as short vowels (e.g., Bie-ne vs. Bin-*ne*) and short vowels are spoken as long vowels (e.g., Ge-wit-*ter* vs. Ge-wie-ter). These minimal pairs are provided when wrong answers have been submitted or when help is requested. This way, we explicitly train vowel length perception and distinction as well as provide scaffolding feedback.

The difficulty as well as the target linguistic competences are defined by the same adjustments of parameters as in the game’s first chapter.

4.1.3 Orthographic markers

After acquiring the knowledge about the structures of stressed syllables, children are exposed to the spelling of these structures in the game’s third chapter. In interactive tutorials, they acquire metalinguistic knowledge about the rules that underlie the spelling of open and closed syllables. Thus, this type of game focuses on learning orthographic and morphemic regularities and letter-sound correspondences. Children train the recognition of orthographic markers for open and closed syllables, which are listed in Figure 5, in various minigames. The vowel of an open syllable can either be i) unmarked (e.g., ma-len, to paint), ii) marked with a vowel doubling such as diphthongs (e.g., Dau-men, thumb) or the German long i (e.g., Bie-ne, bee), which also has unmarked exceptions (e.g., Ti-ger), or iii) marked with the silent h (e.g., feh-len, to miss). Closed syllables with simple structures are i) unmarked (e.g., Fel-sen, rock). Closed syllables with the more complex structure of ambisyllabic consonants are ii) marked with doublings (e.g., ge-win-*n*en, to win). Additionally, German consonant doubling comes with two special subtypes: iii) *ck* instead of *kk* as in Ha-*ck*e (pick) and *tz* instead of *zz* as in Hit-*z*e (heat).

Blobs represent each of the above mentioned orthographic realizations. The children need to select the Blob that represents the vowel spelling of the stressed syllable (e.g., Figure 5b). This chapter contains in total six subchapters, each dealing with one of the special orthographic rules. The characteristics of each of the orthographic markers are explained in individual tutorials.



a. *Ambisyllabic consonant doubling: rennen (to run).*

b. *Consolidation task: Lehrer (teacher).*

Figure 5. *"Orthographic Markers". a) Children need to recognize if the displayed word ren-nen (to run) contains a long or short vowel and for the latter case whether the short vowel is marked orthographically with an ambisyllabic consonant doubling. b) Consolidation task with all orthographic markers. Displayed is the word Leh-rer (teacher), whose long vowel is marked orthographically with a silent h.*

For example, ambisyllabic consonant doublings (e.g., nn, ck, tz) are explained to generally appear if the vowel of the stressed syllable is followed by a single consonant phoneme before the next vowel is perceived or the word ends [e.g., ren-nen (to run) vs. fin-den (to find)].

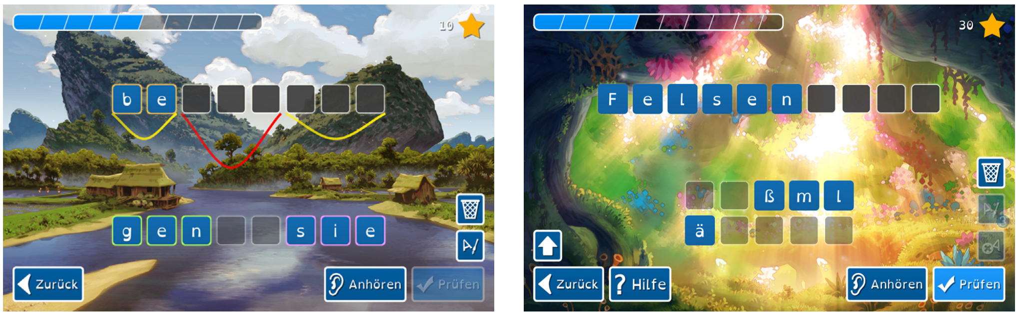
This chapter is of special importance as training to recognize orthographic markers is crucial for spelling [26, 31]. Algorithms of spelling rules to detect and apply orthographic markers are used in various evidence-based treatments (e.g., [39, 75], cf. [31]). However, to the best of our knowledge, the algorithms to determine orthographic marking of vowel length have not been related to syllable stress in any computer-based intervention before.

Step by step, the number of blobs the children can choose from increases. Additionally, the similarity of distracting blobs varies. Exemplary tasks can be seen in Figure 5.

4.1.4 Syllable arcs

Prosodiya's fourth chapter is about syllable segmentation. Children learn to segment written words into relevant components and thus foster their orthographic representations (their knowledge about written words in long-term memory). The game is a digitized version of the task "draw syllable arc", commonly used in learning therapy to improve syllable analysis [30]. Again, we enhanced the commonly used version with emphasis on syllable stress – the syllable arc ought to be drawn deeper for stressed than for unstressed syllables, as can be seen in Figure 6.

Figure 6. *"Syllable Arcs". Children need to draw syllable arcs for the word ver-su-chen (to try) to mark syllable boundaries. Deeper arcs are drawn for stressed syllables.*



a. *Easy difficulty level: besiegen (to win).* b. *Hard difficulty level: Felsen (rock).*

Figure 7. *Spelling. a) The word be-<u>sie</u>-gen (to win) needs to be spelled. Syllable arcs indicate syllable stress and vowel length. Syllables are indicated by same-color border of letters. b) Consolidation task for the word <u>Fel</u>-sen (rock). No syllable arcs are drawn, number of letters is not exposed and letters leading to homophonic misspellings are added as distractors.*

4.1.5 Prosodiya congruency

In chapter five of Prosodiya, children are taught that the rules they have learned in the previous chapters do not account for all word forms. To learn that the rules are only applicable to the basic (uninflected) form of a word, children have to recognize which words are in a 'Prosodiya-ish' form. This exercises bridges the gap from Prosodiya's first module to the second module that focuses on morphologic training, i.e. to learn to deduce that *gewinnt* (he wins) is spelled with an ambisyllabic consonant doubling as it's derived from *gewinnen* (to win).

4.1.6 Spelling

The first module of Prosodiya ends with a chapter about spelling. In this game, the children learn to apply and foster their previously learned skills by spelling words. They do so by dragging-and-dropping letters from a predefined set to form the word.

This exercise comes with several difficulties which are provided by adjusting the parameters 'letters' and 'syllable arcs'. The set of letters to spell a given word may i) only consist of the exact letters of the words, resulting in a letter-arrangement task as can be seen in Figure 7a, ii) also contain letters that do not share phonological similarities to any letter of the word, resulting in a letter discrimination task, or iii) contain distracting letters that would lead to homophonic (words that sound alike) misspellings (e.g., <u>Fel</u>-*l*sen instead of <u>Fel</u>-sen, see Figure 7b). To help link the awareness of orthographic markers to the stressed syllables, syllable arcs are drawn underneath the target input field in some conditions, depending on the individual difficulty. The color of the syllable arcs refer to syllable stress and vowel length (yellow for unstressed syllables, red for open stressed syllables, blue for closed stressed syllables).

To support scaffolding feedback, individual letters can be solved or distractor letters can be deleted after the children misspell a word.

It is our intention that, at the end of the first module of Prosodiya, children should have improved: the perception and production of syllable stress, vowel length distinction, identification of syllable boundaries, recognition of orthographic markers, and, finally, spelling words in basic forms.

4.2 Visual aesthetic design

The visual aesthetic design of Prosodiya follows three design principles to be appealing, consistent, and simple [76]. To address the first principle, we are collaborating with a renowned comic artist and licensed images from the comic "The Wormworld Saga" (https://wormworldsaga.com) as background images for our game. We adjusted the high quality images to fit our needs regarding story, atmosphere, and mechanics of our game.

We have chosen a fantasy-themed setting for Prosodiya due to the fact that embedding learning activities in fantasy contexts has proven to be beneficial for motivation, involvement, and learning (e.g., [77]). Based on the background images, we also crafted a map of Prosodiya that is used as level selection and to show overall progress within the game, see Figure 2. The background images change during game play continuously and represent places of the world map.

Our pedagogical agents, the Kugellichter, were also carefully designed to engage children and facilitate learning. Their round shapes may induce positive emotions that facilitate learning and improve comprehension [78]. The agents have three distinguishable properties, i.e. color, size, and mouth, to facilitate memorizing their linguistic property in game play.

The pictures representing words (e.g., Figure 5) have also been designed iteratively to fit in the game's overall look and feel and to describe the word as best as possible.

We use OpenDyslexic [79] as font because it has been specifically designed for dyslexics in appropriate sizes. Buttons were kept as simple as possible using icons supplementary to text in order to minimize the effort to read unrelated information, with which young or dyslexic children may struggle.

To keep it simple, we have limited the game screen to game elements required by the respective activity and forego additional elements that might distract the learner or hinder learning.

Overall, the visual appearance of Prosodiya is intended to increase the perception of positive affect and immersion and to highlight the content to be learned.

4.3 Narrative design

The narrative of Prosodiya provides information on the content to be learned and game mechanics as well as to be an incentive and to increase self-perception of progression.

The narrative of Prosodiya reflects worries and needs of families of children with dyslexia: Prosodiya is haunted by a mysterious fog that has covered all of the peaceful land, causing the inhabitants to live a live full of worries and sorrows. This relates to affected children who often experience the difficulties of literacy acquisition as an impenetrable fog. In a way that the story deals with real-life struggles of affected children, we wanted to positively affect children's lives beyond the game's world. The game aims at helping children to "see clearly again" and feel comfortable within the world of reading and writing. Progressing through the world of Prosodiya and deliberating the world's places from the mischievous should ideally reflect the progress of learner's literacy acquisition.

In the current version of Prosodiya, the active continuation of the storyline is limited to a prologue in form of a cutscene and the interactive tutorials to continue the storyline. However, the map of Prosodiya and a change of background images in game activities implicitly continue the storyline. In future development, additional cutscenes will be included to actively tell the full story in order to contribute to the game's stickiness.

4.4 Incentive system

Prosodiya uses intrinsic and extrinsic motivational elements to encourage learners to continue playing with Prosodiya.

Regarding extrinsic rewards, children can collect points (see top right corner of the game screen) and receive encouraging responses from the pedagogical agents in the game activities when answering correctly. More points are given when they solve a task at the first go to avoid trial-and-error behavior.

Depending on children's performance, subsequent game content may be unlocked. To provide high replay value and to increase training effects, we use a 1-3 star rating (i.e. more stars for higher performance) for each level displayed underneath each level on the world map.

Intrinsic rewards cover increased knowledge of the children, increasing complexity of tasks and words to make for new challenges, and the story progression.

For future development, we are planning that children may redeem collected points to buy a dragon egg, incubate it, and raise the hatched dragon to be their companion within the world of Prosodiya and to play bonus games with. With this, we aim to support long-term motivation and give children a place within the game where they can relax and ease their mind.

4.5 Musical score

The musical score of Prosodiya consists of ambient music in the level selection, feedback used to acknowledge correct and incorrect responses, as well as voices of the pedagogical agents.

Regarding feedback to tasks, the game responds to correct answers by playing pleasant sounds and praising the children with short expressions spoken by the pedagogical agents. A different, more sophisticated sound is played if the task was solved at the first go. If wrong answers are logged in, children are encouraged to try again.

When present, the pedagogical agents give spoken feedback. Pedagogical agents also explain game mechanics and learning goals in tutorials. Professional speakers have lent their voices to the pedagogical agents with a special focus on positive affect and emphasizing individual characteristics relating to the linguistic property represented by the agent. This affective encouragement may also positively affect children's performance (cf. [80]).

4.6 Results of a pilot study

We conducted a pilot study with a preliminary version of Prosodiya during a period of six weeks in winter 2014/2015 with 11 dyslexic children to primarily investigate feasibility and user experience. The preliminary version included the first two games ("Stress Pattern", "Syllable Structure") and a limited version of the third game ("Orthographic Markers"). A limited vocabulary of 220 words was used. We reported the results in the corresponding conference paper in detail [45] and will summarize the findings:

The results of questionnaires handed out to children and parents indicate: children enjoyed playing the game, the children's motivation could be maintained during the period, children and parents would continue to use Prosodiya and request for more exercises, children and parents perceived self-efficacy, the graphical appearance of Prosodiya is appealing, and the pedagogical agents are well perceived. However, the results also made the following drawbacks apparent: tutorials of the preliminary versions were perceived by some of the children to be too complicated and too long, and some children reported that the objective of an exercise was not always obvious. We have addressed these issues by facilitating and shortening the tutorials and introduced tooltips for each type of task (Figure 4b).

Overall, in spelling as well as in reading, approximately half of the children of the current sample improved their performance. About a third did not change in performance, and two children performed worse. Overall, the best improvements were obtained in speed reading comprehension tests. Due to the limitations of the study of no control group and not including explicit spelling training, implications must be drawn with caution.

5 Conclusion and outlook

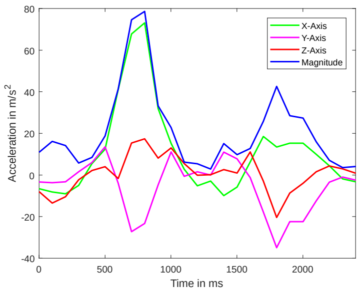
In this article, we have reflected the state of research of the relation of syllable stress and literacy acquisition. We then have examined advantages and disadvantages of therapeutic (TI), computer-based (CBI), and digital game-based interventions (DGBI) to treat reading and spelling disorders. We conclude that dyslexic children should receive evidence-based

TIs as soon as possible and as long as necessary and in addition use evidence-based DGBIs to maximize engagement, motivation, and learning until their reading and spelling proficiencies enables them to participate in social, cultural, and educational life on an age-appropriate level. The disadvantages of TIs of being cost-intensive, dependent on time, location, and tutor may be addressed by DGBIs that can be used autonomously by children within or outside of class while offering unique opportunities to (re-) engage and motivate learners. As our literature review revealed, there is a lack of DGBIs for the treatment of spelling disorders in German dyslexic children.

Given the potential of DGBIs, we proposed our mobile serious game “Prosodiya” for German dyslexic primary school children. By utilizing the building blocks of game design for learning proposed by Plass et al. [53], i.e. game mechanics, visual aesthetic design, narrative design, an incentive system, musical score, and content and skills, we examined how Prosodiya contributes to the state of the art of computer-based interventions for German dyslexic children. Prosodiya focusses on teaching the relation between syllable stress, vowel length, and orthographic markers to improve reading and spelling. We examined the various evidence-based trainings included in Prosodiya. To sum up, the contributions of Prosodiya to date are: i) teaching the perception and production of syllable stress, ii) relating the stressed syllable’s linguistic characteristics to vowel length and orthographic marking, iii) including mouth motor activities to facilitate the improvement of phonological awareness, while iv) balancing aforementioned learning objectives with the design of game play to engage and motivate learners and maximize learning. Lastly, we aim to explore and integrate approaches of embodied trainings to support and facilitate learning. DGBIs offer unique opportunities to incorporate sensor-based embodied trainings in a way that children practice on their own, without the need of human instructors, engaging learners on a behavioral and cognitive level [53].

More specifically, our first approach of using embodied trainings will address the games of “syllable arcs” (see Section 4.1.2). While we already tested a preliminary version of the game called “swing the stress pattern”, it is still under development. The embodied game consists of two parts. At first, children swing with customary fitness trackers on their hand the stress pattern of the word (see Figure 8a). That is, instead of drawing a syllable arc on the touchscreen, children are required to use their (whole) body to execute deep or shallow swings. Spelling the syllables is the second part of the game. Children do so by placing the letters in their corresponding syllable arcs, similar to Section 4.1.4. The first part is considered an embodied training, a novelty in the area of mobile serious games. Embodied trainings – interventions mapping (whole) body movements and gestures to key features of learning content – can be used to additionally support learning-impaired children in learning therapy. The concept of embodied trainings is based on the theory of embodied cognition [81]. Influencing the learning process positively with body movement and embodied experience, the aims of such physical interventions are the development of embodied memory aids to foster learning, to ease mental effort and demands of working memory, and to be more fun. The positive effect of gesturing on learning was shown for various areas, for example gesturing while learning a new arithmetical strategy resulted in a better retention of knowledge [82] or acting out movements or gestures of animals increased the learning of their names in a foreign language [83]. Embodied trainings are already part of many analogous interventions for learning disorders, for example embodied number-line estimation tasks for children with dyscalculia [48, 84] or the so-called “swing the syllables” [69, 85] for dyslexic children. Besides analogous interventions, research on computer-based embodied trainings provides evidence for the positive impact on learning. For example, [86] have shown that computer-based embodied trainings helped children to foster the acquisition of basic numerical competencies significantly better than counterparts without (or with less) embodiment. We digitized and enhanced the embodied training “swing the syllables”. By “dancing” the speech rhythm, children learn to segment words into syllables and identify their stress pattern.

a. *Stress pattern swung for a disyllabic trochee (e.g., Bie-ne, bee).*



b. *Acceleration recorded with wearable.*

Figure 8. *a) Embodied training "Swing the Stress Pattern". Children speak each syllable clearly and loudly, simultaneously do a sidestep, and swing their writing hand from their left to their right shoulder. Deeper and more emphasized swing is performed for stressed syllables. b) Acceleration values recorded with a fitness tracker worn on children's wrist.*

They are instructed to speak each syllable clearly and loudly, simultaneously do a sidestep, and swing their writing hand from their left to their right shoulder, see Figure 8a. A deeper and more emphasized swing is performed for stressed syllables. Swinging the speech rhythm of words, children are expected to improve their reading and writing performance by systematically fostering syllable awareness. In contrast to the original intervention, we again distinguish between stressed and unstressed syllables.

To include this embodied training into Prosodiya, we are developing a system that recognizes the movements of the children and classifies the stress pattern they have swung [87]. We use built-in accelerometers of customary fitness trackers to record movement and to classify syllable stress. Our user tests throughout the development highlight three important challenges that arise when sensor-based embodied trainings are developed for children: First, fitness trackers should be used due to size and comfort. Smartphones are too big for children's hands causing restricted movement, even when the Smartphone is attached securely with a wrist wrap. Second, systems have to be adjusted to the children's behavior and needs. Prominent features that were used to classify the gestures with adults are less reliable when performed by children and need to be refined or changed. Lastly, performing the gestures correctly cannot be expected but must be learned and trained extensively. Stressed syllables were often swung weaker or hardly distinguishable from unstressed ones. It is not clear whether this is caused by the lack of the awareness of syllable stress or by missing feedback about swing intensity. We therefore will integrate an intensive learning phase where a character of the game's world will explain extensively the principle of syllable stress and the intended use of the gestures and give qualitative feedback on the performed gestures. Overall, our results and systematic literature research encourage the approach to develop embodied trainings with simple, body-based interactions using built-in accelerometers of Smartphones and fitness trackers. This is, to the best of our knowledge, the first embodied training in a mobile serious game that evaluates movement and gestures. Other games already include the instruction to perform embodied trainings but do not evaluate the children's response (e.g., [88]). In contrast, we display the classified stress pattern on the game screen with corresponding syllable arcs.

To evaluate the therapeutic approach of Prosodiya empirically and validate the positive indications of our pilot study, we are conducting a randomized controlled field trial with a waiting control group design in spring/summer of 2018. One hundred and twenty-nine primary-school children are participating and are practicing with Prosodiya for a duration of 8-10 weeks. Approximately 70 children are poor readers and/or spellers. According to the training plan, children are asked to practice with Prosodiya five days per week, 20 minutes per day.

Acknowledgements

This research was funded by the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments. Heiko Holz is a doctoral student at the LEAD Graduate School & Research Network.

References

[1] Moll, K., Kunze, S., Neuhoff, N., Bruder, J., Schulte-Körne, G., Specific Learning Disorder: Prevalence and Gender Differences, PLoS One, Vol. 9, Nr. 7, 2014. https://doi.org/10.1371/journal.pone.0103537
[2] Moll, K., Landerl, K., Double Dissociation Between Reading and Spelling Deficits, Sci. Stud. Read., Vol. 13, Nr. 5, 2009. https://doi.org/10.1080/10888430903162878
[3] Katusic, S.K., Colligan, R.C., Barbaresi, W.J., Schaid, D.J., Jacobsen, S.J., Incidence of reading disability in a population-based birth cohort, 1976-1982, Rochester, Minn, Mayo Clin. Proc., Vol. 76, Nr. 11, 2001. https://doi.org/10.4065/76.11.1081
[4] Schulte-Körne, G., Remschmidt, H., Legasthenie - Symptomatik, Diagnostik, Ursachen, Verlauf und Behandlung, Dtsch Arztebl Int., Vol. 100, Nr. 7, 2003.
[5] Bender, F., Brandelik, K., Jeske, K., Lipka, M., Löffler, C., Mannhaupt, G., Naumann, C.L., Nolte, M., Ricken, G., Rosin, H., Scheerer-Neumann, G., von Aster, M., von Orloff, M., Die integrative Lerntherapie, Lernen und Lernstörungen, Vol. 6, Nr. 2, 2017. https://doi.org/10.1024/2235-0977/a000167
[6] Daniel, S.S., Walsh, A.K., Goldston, D.B., Arnold, E.M., Reboussin, B. a, Wood, F.B., Suicidality, School Dropout, and Reading Problems Among Adolescents, J. Learn. Disabil., Vol. 39, Nr. 6, 2006. https://doi.org/10.1177/00222194060390060301
[7] Beddington, J., Cooper, C.L., Field, J., Goswami, U., Huppert, F.A., Jenkins, R., Jones, H.S., Kirkwood, T.B.L., Sahakian, B.J., Thomas, S.M., The mental wealth of nations, Nature, Vol. 455, Nr. 7216, 2008. https://doi.org/10.1038/4551057a
[8] Abrami, P., Borohkovski, E., Lysenko, L., The Effects of ABRACADABRA on Reading Outcomes: A Meta-Analysis of Applied Field Research, J. Interact. Learn. Res., Vol. 26, Nr. 4, 2015.
[9] Kargl, R., Purgstaller, C., Weiss, S., Fink, A., Effektivitätsüberprüfung eines morphemorientierten Grundwortschatz-Segementierungstrainings (MORPHEUS) bei Kindern und Jugendlichen, Heilpädagogische Forsch., Vol. 34, Nr. 3, 2008.
[10] Klatte, M., Bergström, K., Steinbrink, C., Konerding, M., Lachmann, T., "Effects of the Computer-Based Training Program Lautarium on Phonological Awareness and Reading and Spelling Abilities in German Second-Graders", In Lachmann, T. and Weis, T. (Eds.), Reading and Dyslexia: From Basic Functions to Higher Order Cognition, Springer International Publishing, Cham, 2018. https://doi.org/10.1007/978-3-319-90805-2_15
[11] Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., Goswami, U., 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, Nr. 1, 2013. https://doi.org/10.1002/rrq
[12] [Lenhard, W., Lenhard, A., Evidenzbasierte Förderung schulischer Fertigkeiten am Computer: Lernspiele mit Elfe und Mathis, Förderprogramme für Vor. und Grundschule, Vol. 2006, 2016.
[13] Deterding, S., Dixon, D., Khaled, R., Nacke, L., "From Game Design Elements to Gamefulness: Defining 'Gamification'", In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, ACM, New York, NY, USA, 2011. https://doi.org/10.1145/2181037.2181040
[14] Boyle, E.A., Hainey, T., Connolly, T.M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., Pereira, J., An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games, Comput. Educ., Vol. 94, Nr. Supplement C, 2016. https://doi.org/10.1016/j.compedu.2015.11.003
[15] Berkling, K., "Phontasia: A phonics game for German and its effect on orthographic skills–first corpus explorations", In WOCCI 2017: 6th International Workshop on Child Computer Interaction, ISCA, ISCA, 2017. https://doi.org/10.21437/WOCCI.2017-2

[16] Kast, M., Baschera, G.-M., Gross, M., Jäncke, L., Meyer, M., Computer-based learning of spelling skills in children with and without dyslexia, Ann. Dyslexia, Vol. 61, Nr. 2, 2011. https://doi.org/10.1007/s11881-011-0052-2
[17] Rello, L., Bayarri, C., Otal, Y., Pielot, M., "A computer-based method to improve the spelling of children with dyslexia", In Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility - ASSETS '14, ACM Press, New York, New York, USA, 2014. https://doi.org/10.1145/2661334.2661373
[18] Bradley, L., Bryant, P.E., Categorizing sounds and learning to read—a causal connection, Nature, Vol. 301, Nr. 5899, 1983. https://doi.org/10.1038/301419a0
[19] Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., Leong, V., Impaired perception of syllable stress in children with dyslexia: A longitudinal study, J. Mem. Lang., Vol. 69, Nr. 1, 2013. https://doi.org/10.1016/j.jml.2013.03.001
[20] Leong, V., Hämäläinen, J., Soltész, F., Goswami, U., Rise time perception and detection of syllable stress in adults with developmental dyslexia, J. Mem. Lang., Vol. 64, Nr. 1, 2011. https://doi.org/10.1016/j.jml.2010.09.003
[21] Sauter, K., Heller, J., Landerl, K., Sprachrhythmus und Schriftspracherwerb, Lernen und Lernstörungen, Vol. 1, Nr. 4, 2012. https://doi.org/10.1024/2235-0977/a000023
[22] Kohler, K.J., "Invariance and variability in speech timing: From utterance to segment in German", In Perkell, J.S. and Klatt, D.H. (Eds.), Invariance and Variability in Speech Processes, Erlbaum, Hillsdale, 1986.
[23] Jiménez-Fernández, G., Gutiérrez-Palma, N., Defior, S., Impaired stress awareness in Spanish children with developmental dyslexia, Res. Dev. Disabil., Vol. 37, 2015. https://doi.org/10.1016/j.ridd.2014.11.002
[24] Brandelik, K., Sprachrhythmische Fähigkeiten im Schriftspracherwerb, Südwestdeutscher Verlag für Hochschulschriften, 2014.
[25] Staffeldt, S., Einführung in die Phonetik, Phonologie und Graphematik des Deutschen. Ein Leitfaden für den akademischen Unterricht, Stauffenburg, Tübingen, 2010.
[26] Landerl, K., Categorization of vowel length in German poor spellers: An orthographically relevant phonological distinction, Appl. Psycholinguist., Vol. 24, Nr. 4, 2003. https://doi.org/10.1017/S0142716403000262
[27] Klicpera, C., Gasteiger Klicpera, B., Sind Rechtschreibschwierigkeiten Ausdruck einer phonologischen Störung?, Z. Entwicklungspsychol. Padagog. Psychol., Vol. 32, Nr. 3, 2000. https://doi.org/10.1026//0049-8637.32.3.134
[28] Bhide, A., Power, A., Goswami, U., A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention, Mind, Brain, Educ., Vol. 7, Nr. 2, 2013. https://doi.org/10.1111/mbe.12016
[29] Thomson, J.M., Leong, V., Goswami, U., Auditory processing interventions and developmental dyslexia: a comparison of phonemic and rhythmic approaches, Read. Writ., Vol. 26, Nr. 2, 2013. https://doi.org/10.1007/s11145-012-9359-6
[30] Galuschka, K., Schulte-Körne, G., Clinical practice guideline: The diagnosis and treatment of reading and/or spelling disorders in children and adolescents, Dtsch Arztebl Int., Vol. 113, Nr. 16, 2016. https://doi.org/10.3238/arztebl.2016.0279
[31] Galuschka, K., Ise, E., Krick, K., Schulte-Körne, G., Effectiveness of Treatment Approaches for Children and Adolescents with Reading Disabilities: A Meta-Analysis of Randomized Controlled Trials, PLoS One, Vol. 9, Nr. 2, 2014. https://doi.org/10.1371/journal.pone.0089900
[32] McArthur, G., Eve, P.M., Jones, K., Banales, E., Kohnen, S., Anandakumar, T., Larsen, L., Marinus, E., Wang, H.-C., Castles, A., Phonics training for English-speaking poor readers, Cochrane Database Syst. Rev., Nr. 12, 2012. https://doi.org/10.1002/14651858.CD009115.pub2
[33] Aro, M., Learning to read: the effect of orthography, 2004.
[34] Landerl, K., Wimmer, H., Frith, U., The impact of orthographic consistency on dyslexia: A German-English comparison, Cognition, Vol. 63, Nr. 3, 1997. https://doi.org/https://doi.org/10.1016/S0010-0277(97)00005-X
[35] Rau, A.K., Moll, K., Snowling, M.J., Landerl, K., Effects of orthographic consistency on eye movement behavior: German and English children and adults process the same words differently, J. Exp. Child Psychol., Vol. 130, 2015. https://doi.org/10.1016/j.jecp.2014.09.012
[36] Tacke, G., Evaluation eines Lesetrainings zur Förderung lese-rechtschreibschwacher Grundschüler der zweiten Klasse, Psychol. Erziehung und Unterr., Vol. 52, Nr. 3, 2005.
[37] Groth, K., Hasko, S., Bruder, J., Kunze, S., Schulte-Körne, G., Interventionseffekte bei Lese-Rechtschreibstörung: Evaluation von zwei Förderkonzepten unter besonderer Betrachtung

methodischer Aspekte, Lernen und Lernstörungen, Vol. 2, Nr. 3, 2013. https://doi.org/10.1024/2235-0977/a000038
[38] Klicpera, C., Weiss, J., Gasteiger-Klicpera, B., Erfolg einer schulischen Legasthenikер-Förderung für Kinder der 3. und 4. Schulstufe nach dem Programm von Reuter-Liehr, Heilpädagogische Forsch., Vol. 39, Nr. 2, 2013.
[39] Ise, E., Schulte-Körne, G., Spelling deficits in dyslexia: Evaluation of an orthographic spelling training, Ann. Dyslexia, Vol. 60, Nr. 1, 2010. https://doi.org/10.1007/s11881-010-0035-8
[40] Ise, E., Engel, R.R., Schulte-Körne, G., Was hilft bei der Lese-Rechtschreibstörung? Ergebnisse einer Metaanalyse zur Wirksamkeit deutschsprachiger Förderansätze, Kindheit und Entwicklung, Vol. 21, Nr. 2, 2012. https://doi.org/10.1026/0942-5403/a000077
[41] Drigas, A., Batziaka, E., Dyslexia and ICTs, Assessment and Early Intervention in Kindergarten, Int. J. Emerg. Technol. Learn., Vol. 11, Nr. 02, 2016. https://doi.org/10.3991/ijet.v11i02.5193
[42] Cidrim, L., Madeiro, F., Information and Communication Technology (ICT) applied to dyslexia: literature review, Rev. CEFAC, Vol. 19, Nr. 1, 2017. https://doi.org/10.1590/1982-021620171917916
[43] Stock, C., Schneider, W., PHONIT: Ein Trainingsprogramm zur Verbesserung der phonologischen Bewusstheit und Rechtschreibleistung im Grundschulalter, Hogrefe, Göttingen, 2011.
[44] Kaasa health GmbH, Meister Cody - Namagi, https://www.meistercody.com/produkte/namagi-legasthenie/
[45] Holz, H., Brandelik, K., Brandelik, J., Beuttler, B., Kirsch, A., Heller, J., Meurers, D., "Prosodiya – A Mobile Game for German Dyslexic Children", In Dias, J., Santos, P.A., and Veltkamp, R.C. (Eds.), Games and Learning Alliance: 6th International Conference, GALA 2017, Lisbon, Portugal, December 5--7, 2017, Proceedings, 2017. https://doi.org/10.1007/978-3-319-71940-5_7
[46] Berninger, V.W., Nagy, W., Tanimoto, S., Thompson, R., Abbott, R.D., Computer instruction in handwriting, spelling, and composing for students with specific learning disabilities in grades 4–9, Comput. Educ., Vol. 81, 2015. https://doi.org/10.1016/j.compedu.2014.10.005
[47] Ronimus, M., Kujala, J., Tolvanen, A., Lyytinen, H., Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge, Comput. Educ., Vol. 71, 2014. https://doi.org/10.1016/j.compedu.2013.10.008
[48] Ninaus, M., Kiili, K., Mcmullen, J., Moeller, K., Assessing fraction knowledge by a digital game, Comput. Human Behav., Vol. 70, 2017. https://doi.org/10.1016/j.chb.2017.01.004
[49] Sense, F., Behrens, F., Meijer, R.R., van Rijn, H., "Stability of Individual Parameters in a Model of Optimal Fact Learning", In Proceedings of the 13th International Conference on Cognitive Modeling, 2015.
[50] Björklund, M., Dyslexic Students: Success Factors for Support in a Learning Environment, J. Acad. Librariansh., Vol. 37, Nr. 5, 2011. https://doi.org/10.1016/j.acalib.2011.06.006
[51] Huotari, K., Hamari, J., "Defining Gamification: A Service Marketing Perspective", In Proceeding of the 16th International Academic MindTrek Conference, ACM, New York, NY, USA, 2012. https://doi.org/10.1145/2393132.2393137
[52] Hamari, J., Koivisto, J., Sarsa, H., "Does Gamification Work? – A Literature Review of Empirical Studies on Gamification", In 2014 47th Hawaii International Conference on System Sciences, IEEE, 2014. https://doi.org/10.1109/HICSS.2014.377
[53] Plass, J.L., Homer, B.D., Kinzer, C.K., Foundations of Game-Based Learning, Educ. Psychol., Vol. 50, Nr. 4, 2015. https://doi.org/10.1080/00461520.2015.1122533
[54] Wouters, P., van Nimwegen, C., van Oostendorp, H., van der Spek, E.D., A meta-analysis of the cognitive and motivational effects of serious games., J. Educ. Psychol., Vol. 105, Nr. 2, 2013. https://doi.org/10.1037/a0031311
[55] Kerawalla, L., Crook, C., From promises to practices: The fate of educational software in the home, Technol. Pedagog. Educ., Vol. 14, Nr. 1, 2005. https://doi.org/10.1080/14759390500200195
[56] Ronimus, M., Lyytinen, H., Is School a Better Environment Than Home for Digital Game-Based Learning? The Case of GraphoGame, Hum. Technol. An Interdiscip. J. Humans ICT Environ., Vol. 11, Nr. 2, 2015. https://doi.org/10.17011/ht/urn.201511113637
[57] Gee, J.P., What video games have to teach us about learning and literacy, Comput. Entertain., Vol. 1, Nr. 1, 2003. https://doi.org/10.1145/950566.950595
[58] Prensky, M., Digital game-based learning, Comput. Entertain., Vol. 1, Nr. 1, 2003. https://doi.org/10.1145/950566.950596

[59] Ke, F., Abras, T., Games for engaged learning of middle school children with special learning needs, Br. J. Educ. Technol., Vol. 44, Nr. 2, 2013. https://doi.org/10.1111/j.1467-8535.2012.01326.x

[60] Holmes, W., Using game-based learning to support struggling readers at home, Learn. Media Technol., Vol. 36, Nr. 1, 2011. https://doi.org/10.1080/17439884.2010.531023

[61] Griffiths, M., The educational benefits of videogames, Educ. Heal., Vol. 20, Nr. 3, 2002.

[62] Squire, K.D., Video Games and Education: Designing Learning Systems for an Interactive Age, Educ. Technol., Vol. 48, Nr. 2, 2008.

[63] Homer, B.D., Kinzer, C.K., Plass, J.L., Letourneau, S.M., Hoffman, D., Bromley, M., Hayward, E.O., Turkay, S., Kornak, Y., Moved to learn: The effects of interactivity in a Kinect-based literacy game for beginning readers, Comput. Educ., Vol. 74, 2014. https://doi.org/10.1016/j.compedu.2014.01.007

[64] Arnab, S., Lim, T., Carvalho, M.B., Bellotti, F., de Freitas, S., Louchart, S., Suttie, N., Berta, R., De Gloria, A., Mapping learning and game mechanics for serious games analysis, Br. J. Educ. Technol., Vol. 46, Nr. 2, 2015. https://doi.org/10.1111/bjet.12113

[65] Quinn, C.N., Engaging learning: designing e-learning simulation games, Jossey-Bass, Quinn2005, 2005.

[66] Berkling, K., Pflaumer, N., Phontasia - a Phonics Trainer for German Spelling in Primary Education, Proc. 4th Work. Child Comput. Interact. (WOCCI 2014), Nr. Wocci, 2014.

[67] Berkling, K., Pflaumer, N., Lavalley, R., "German Phonics Game using Speech Synthesis - A Longitudinal Study about the Effect on Orthography Skills", In Workshop on Spoken Language Technology for Education (SLaTE), 2015.

[68] Reid, G., Strnadová, I., Cumming, T., Expanding horizons for students with dyslexia in the 21st century: Universal design and mobile technology, J. Res. Spec. Educ. Needs, Vol. 13, Nr. 3, 2013. https://doi.org/10.1111/1471-3802.12013

[69] Reuter-Liehr, C., Behandlung der Lese-Rechtschreibschwäche nach der Grundschulzeit: Anwendung und Überprüfung eines Konzeptes, Z. Kinder. Jugendpsychiatr., Vol. 21, Nr. 3, 1993.

[70] Snowling, M.J., Phonological processing and developmental dyslexia, J. Res. Read., Vol. 18, Nr. 2, 1995. https://doi.org/10.1111/j.1467-9817.1995.tb00079.x

[71] Berninger, V.W., Winn, W.D., Stock, P., Abbott, R.D., Eschen, K., Lin, S.J., Garcia, N., Anderson-Youngstrom, M., Murphy, H., Lovitt, D., Trivedi, P., Jones, J., Amtmann, D., Nagy, W., Tier 3 specialized writing instruction for students with dyslexia, Read. Writ., Vol. 21, Nr. 1–2, 2008. https://doi.org/10.1007/s11145-007-9066-x

[72] Kirk, C., Gillon, G.T., Integrated Morphological Awareness Intervention as a Tool for Improving Literacy, Lang. Speech Hear. Serv. Sch., Vol. 40, Nr. 3, 2009. https://doi.org/10.1044/0161-1461(2008/08-0009)

[73] Moll, K., Wallner, R., Landerl, K., Kognitive Korrelate der Lese-, Leserechtschreib- und der Rechtschreibstörung, Lernen und Lernstörungen, Vol. 1, Nr. 1, 2012. https://doi.org/10.1024/2235-0977/a000002

[74] Boyer, N., Ehri, L.C., Contribution of Phonemic Segmentation Instruction With Letters and Articulation Pictures to Word Reading and Spelling in Beginners, Sci. Stud. Read., Vol. 15, Nr. 5, 2011. https://doi.org/10.1080/10888438.2010.520778

[75] Kargl, R., Purgstaller, C., Morphemunterstütztes Grundwortschatz- Segmentierungstraining, Hogrefe, Göttingen, 2010.

[76] Berkling, K., Faller, H., Piertzik, M., "Avoiding Failure in Modern Game Design with Academic Content - A Recipe, an Anti-Pattern and Applications Thereof", In Proceedings of the 9th International Conference on Computer Supported Education, 2017. https://doi.org/10.5220/0006281800250036

[77] Parker, L.E., Lepper, M.R., Bartholomew, K., Cordova, D., Mayer, J., Effects of fantasy contexts on children's learning and motivation: Making learning more fun., J. Pers. Soc. Psychol., Vol. 62, Nr. 4, 1992. https://doi.org/10.1037/0022-3514.62.4.625

[78] Plass, J.L., Heidig, S., Hayward, E.O., Homer, B.D., Um, E., Emotional design in multimedia learning: Effects of shape and color on affect and learning, Learn. Instr., Vol. 29, 2014. https://doi.org/10.1016/j.learninstruc.2013.02.006

[79] Gonzalez, A., OpenDyslexic, https://www.opendyslexic.org/.

[80] Schmitt, K.L., Hurwitz, L.B., Duel, L.S., Linebarger, D.L.N., Nichols Linebarger, D.L., Learning through play: The impact of web-based games on early literacy development, Comput. Human Behav., Vol. 81, 2018. https://doi.org/10.1016/j.chb.2017.12.036

[81] Glenberg, A., Witt, J., Metcalfe, J., From the Revolution to Embodiment 25 Years of Cognitive Psychology, Perspect. Psychol. Sci., Vol. 8, Nr. 5, 2013. https://doi.org/10.1177/1745691613498098
[82] Cook, S.W., Mitchell, Z., Goldin-Meadow, S., Gesturing makes learning last, Cognition, Vol. 106, Nr. 2, 2008. https://doi.org/10.1016/j.cognition.2007.04.010
[83] Toumpaniari, K., Loyens, S., Mavilidi, M.-F., Paas, F., Preschool Children's Foreign Language Vocabulary Learning by Embodying Words Through Physical Activity and Gesturing, Educ. Psychol. Rev., Vol. 27, Nr. 3, 2015. https://doi.org/10.1007/s10648-015-9316-4
[84] Link, T., Moeller, K., Huber, S., Fischer, U., Nuerk, H.C., Walk the number line - An embodied training of numerical concepts, Trends Neurosci. Educ., Vol. 2, Nr. 2, 2013. https://doi.org/10.1016/j.tine.2013.06.005
[85] Michel, H.-J., Freiburger Rechtschreibschule. Grundlagen – Diagnosemöglichkeiten – praktische Übungen zum Thema LRS, AOL-Verlag, Buxtehude, 2006.
[86] Dackermann, T., Fischer, U., Nuerk, H., Cress, U., Moeller, K., Applying embodied cognition: from useful interventions and their theoretical underpinnings to practical applications, ZDM, 2017. https://doi.org/10.1007/s11858-017-0850-z
[87] Holz, H., Beuttler, B., Kirsch, A., "Bewegungserkennung mit Wearables für Embodied Trainings in Serious Games", In Burghardt, M., Wimmer, R., Wolff, C., and Womser-Hacker, C. (Eds.), Mensch und Computer 2017 - Tagungsband, Gesellschaft für Informatik e.V., Regensburg, 2017. https://doi.org/https://doi.org/10.18420/muc2017-mci-0368
[88] Ernst Klett Verlag GmbH, Deutsch Klasse 2 mit Zebra, https://www.klett.de/produkt/isbn/ECN20001APA99.

Design Rationales of a Mobile Game-Based Intervention for German Dyslexic Children

Heiko Holz
LEAD Graduate School &
Research Network
Unviersity of Tuebingen
Tuebingen, Germany
heiko.holz@uni-tuebingen.de

Benedikt Beuttler
University of Tuebingen
Tuebingen, Germany
benedikt.beuttler@uni-tuebingen.de

Manuel Ninaus
Leibniz Institut für
Wissensmedien (IWM)
Tuebingen, Germany
m.ninaus@iwm-tuebingen.de

CHI PLAY'18 Extended Abstracts, Oct. 28–31, 2018, Melbourne, Australia.

ACM ISBN 978-1-4503-5968-9/18/10 ...$15.00.
http://dx.doi.org/10.1145/3270316.3272053

Abstract
Educational games have been shown to provide support for children with learning disabilities to overcome their challenges. Gaming and feedback elements can benefit the learning process while keeping children engaged and help them to regain motivation for learning. We present the design of a mobile serious game for German dyslexic primary school children that incorporates gaming elements such as narrative, pedagogical agents, tutorials, feedback, and reward mechanisms. We derive our game design decisions and specify the rationales behind with special focus on the needs and demands of the target group. We evaluate the gaming elements based on the results of 63 children who played the game at home during a period of 9–10 weeks. Results indicate overall positive perception of the game elements. Children were immersed in the fantasy-themed world, liked the pedagogical agents, and indicated that the interactive tutorials gave an easy start into the game, and emphasized the special importance of praise.

CCS Concepts
•Applied computing → Computer games; Interactive learning environments; •Human-centered computing → *User centered design;* Empirical studies in HCI;

Author Keywords
Game Design; Digital Game-Based Learning; Dyslexia

Figure 1: Game 1 "Stress Pattern". Children need to rebuild the stress pattern for the word *rennen* (to run) using big green blobs for stressed and small yellow blobs for unstressed syllables.

Figure 2: Game 2 "Open and Closed Syllables". Children need to decide whether the stressed syllable is open (ends with a vowel, red) or closed (ends with a consonant, blue) for the word *rennen* (to run).

Introduction

Research on digital game-based learning has gained attention and became more popular in recent years (for an overview see [6, 21]). Game-based learning has been shown to be effective or even outperform conventional instruction methods, especially for language learning [48]. Specifically for learning disorders, such as dyslexia and dyscalculia, serious games have proven to support children in addition to school support and learning therapy (e.g., [1, 4, 7, 11, 34]). Educational digital games for children contain various gaming elements, e.g. feedback, rewards, or storytelling, that – on their own and via interactions – influence learning positively [48]. These elements play a crucial role to achieve learning goals [6] and address negative feelings such as frustration, demotivation, or boredom [10]. Consequently, game elements are an important mechanism for therapy, cognitive training, and educational interventions due to their ability to keep players motivated to play and to interact with the application [15, 35] and to promote engagement and learning for children with special needs [28].

We present a mobile serious game for German dyslexic primary school children that has been developed iteratively with a special focus on the target group. We explicitly describe the design rationales behind the game elements, i.e., graphics, narrative, pedagogical agents, interactive tutorials, feedback, and rewarding mechanics. To evaluate end refine the game elements, the results of a pilot study and of 63 German primary school children, who played the game at home during a period of 9–10 weeks within the scope of a randomized controlled field trial, are reported.

Requirements

Prosodiya targets German dyslexic children aged 5–12. Due to this target group, specific requirements, which were derived from interviews with game experts, practitioners, and learning therapists, had to be met. We concluded that the game should i) incorporate engagement that integrates with educational effectiveness, ii) be easy to understand and to use by the target group and support unsupervised training at home, iii) engage children aged between 5–12 with poor reading and/or spelling skills throughout a period of several months, and iv) deliver high quality regarding visual appearance and game experience.

Iterative Children-Centered Game Design

Digital game-based interventions need to deliver high user and gaming experience while being properly designed in terms of educational effectiveness and learning [3, 28, 40, 41]. Although children have their own preferences and needs and may not be seen as "just short adults" [13], developers often approach parents or teachers to ask about their children needs rather than asking the children directly [12]. However, if design decisions are not adapted to the target group, it may pose barriers and make the game less accessible, particularly to children with special educational needs [14]. Thus, including dyslexic primary school children throughout the whole design and development process was necessary. According to [13], children can take the role of the user, tester, informant, and design partner. Therefore, we propose and utilized a design and development approach called iterative children-centered game design (ICCGD) to always focus on the target group. The ICCGD combines the two familiar and successful approaches of user-centered design [2, 36] (UCD) and iterative game design [16] (IGD). Accordingly, we followed the ICCGD for each part of the game.

The ICCGD consists of an initial requirement and context analysis based on expert interviews and observations of paper-based prototypes tested in learning therapy. The result is a first concept of the game. After that, two main

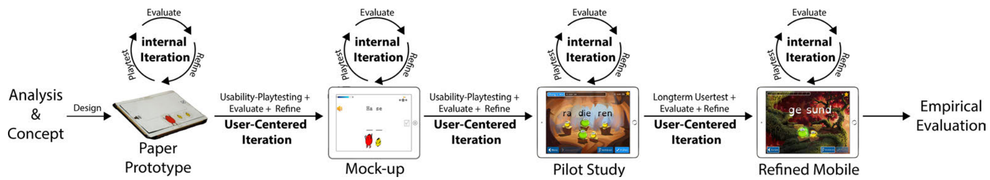


Figure 3: Exemplary Iterative Children-Centered Game Design Process for the game “stress pattern”.

Figure 4: Game 3 “Orthographic markers”. Children need to recognize the orthographic marking responsible for the spelling of the word *rennen* (to run).

Figure 5: Game 4 “Spelling”. Children spell the word *rennen* (to run) by dragging and dropping respective letters.

phases in different development stages are constantly repeated until the game reaches release. These two main phases again consist of three constantly repeating sub-main phases – playtesting, evaluation, and refinement. In the first main phase, the prototype is tested and evaluated internally with team members and experts following the IGD (“internal iteration”). The second main phase is referred to as “user-centered iteration”, in which the internally refined prototype is tested by the children in a session of usability-playtesting or unsupervised user tests over a longer period of time. An example of the ICCGD can be seen in Figure 3.

Following the ICCGD, we can meet the requirements and learning principles, but also optimize game principles, usability, as well as game and user experience regarding the needs, skills, and expectations of (dyslexic) primary school children.

Luckily, we could regularly playtest, evaluate, and refine our game with children enrolled in a center for learning therapy. Based on the results of usability-playtestings and pilot studies conducted with prototypes of different development phases (see Figure 3), we strongly encourage the use of ICCGD for the development of mobile serious games for children with special educational needs.

Prosodiya

Prosodiya [22] is a mobile serious game for German dyslexic primary school children aged 5–12. It is based on evidence-based interventions (e.g., [25, 43]) and recent empirical findings. Prosodiya differs from other game-based interventions in its its special emphasis on syllable stress awareness, which has been shown to be impaired in dyslexic children [20, 26, 32] and highly correlates with reading and spelling skills [44]. Consequently, one explanation for poor spelling performance in dyslexics is seen in the association between syllable stress and orthographic markers. Orthographic markers, i.e. graphemes marking vowel lengths such as consonant and vowel doubling, generally occur in stressed syllables [47]. Mastering the complex orthographic rules to mark long and short vowels is a major difficulty for German children [29, 31].

The presented game aims first at improving the awareness of the stressed syllable’s linguistic features. In a second step, children acquire metalinguistic knowledge about the association between these features and German orthographic rules. In total, four mini-games have been developed that cover the domains of stress pattern recognition and reproduction (see Figure 1), vowel length distinction in form of recognizing open and closed syllables (see Figure 2), metalinguistic knowledge about how such syllables are spelled by recognizing orthographic markers (see Figure 4), and spelling exercises (see Figure 5). Videos of the games can be accessed on prosodiya.com/levels.

Design of Game Elements

As summarized by Plass et al. [40], game mechanics, visual aesthetics, narrative, incentives, musical score, and content and skills are generally agreed as the building blocks of games. Hereafter, we address Prosodiya’s visual aesthetic design, narrative design, and its incentive system.

Figure 7: Progression through the chapter *The Great River*.

Figure 6: In-game map of Prosodiya. All regions except for the final chapter – the *Magic Forest* – have been freed from the fog.

Additionally, we explain the game's pedagogical agents, tutorials, and feedback, which belong to the aforementioned building blocks.

We refer to [22] for a comprehensive description of Prosodiya's game mechanics and pedagogical objectives (content and skills). Detailed evaluation of the game's general validity and player experience are reported in [24].

Visual Aesthetic Design

As concluded by [4], "Learning is fun if it increases without boundaries in difficulty. As long as the graphics are good." we tried to make the game's graphical appearance appealing, consistent, and simple. To address the requirement of having high quality graphics, we collaborated with a renowned comic artists and licensed images from the comic "The Wormworld Saga" (https://wormworldsaga.com) and adjusted them for our needs to fit the story, atmosphere, and mechanics of our game. The characters (see Section *Pedagogical Agents*) and interface elements were designed by different artists. We used OpenDyslexic [19] as the font, which has been specifically designed for dyslexics.

Narrative, Environment, and Progress

As research has shown that embedding learning activities in appealing and appropriate fantasy contexts to be beneficial for motivation, involvement, and learning [9, 38], we developed a fantasy-themed setting consisting of story and

Figure 8: Journey through the world of Prosodiya as reflected by the game's story and progression. Each image represents an exemplary background image for its eponymous chapter.

environmental elements: The eponymous fantasy themed world is haunted by a mysterious fog that has covered all of the peaceful land. The inhabitants of Prosodiya – among others trolls, fairies, and other mythical creatures – are saddened and live a life full of sorrows. Little inhabitants called "Kuggellichter" ("spherical lights"), kindred to will-o-wisps, seek the children's help as they themselves are too weak to help their homeland. The children, guided by the Kugellichter through the world of syllables and orthography, can join their journey and disperse the suppressing fog and free Prosodiya from its dreadful destiny. In order to decipher the mysteries of German orthography and to obtain the *wisdom of words*, they need to understand and apply the *power of*

(a) Small yellow agent representing unstressed syllables.

(b) Big green agent representing stressed syllables.

(c) Big red agent with open mouth representing open syllables (long stressed vowels).

(d) Big blue agent with mouth shut representing closed syllables (short stressed vowels).

Figure 9: The pedagogical agents of Prosodiya

the stressed syllable.

The story of Prosodiya reflects worries and needs of families of children with dyslexia. They often experience the difficulties of literacy acquisition as an impenetrable fog – they feel like there is "no land in sight". The game aims at helping children to clear the blurred vision in order to feel comfortable within the world of reading and writing. It was important for us that the story deals with real-life struggles of affected children to positively affect their life beyond the game's world.

Based on the feedback received in the pilot study [23] and to incorporate the suggestion of [4] to frequently introduce new elements into the game, we implemented a weekly and daily progression system in form of cutscenes, a world map, and changes of environment as well as atmosphere. In the pilot version, we only had one background image for each chapter and no map. Many children and their parents reported that the change of setting had a strong positive effect on their motivation and self-awareness of progression. We addressed their wishes as explained in the following.

Prosodiya's narrative was designed to follow the progression of three individual levels of difficulty: First, different linguistic or orthographic skills are covered in individual chapters. Second, subchapters within a chapter deal with different linguistic or orthographic sub-competences. Lastly, levels within a subchapter increase in difficulty by increasing word structure and word complexity, and by decreasing hints and support provided to children, resulting in more challenging task objectives. Each chapter is embedded in a unique environment and has an eponymous milestone that needs to be accomplished, which is reflected by the map and by level-based environments of subchapters, see Figure 6 and Figure 8. The children's journey starts at the *Waterfall* – the source of the stressed syllable's force – before it takes them through the *Hovi-Village* to rescue its inhabitants, all the way to the *Glass-Blossom Lake* for its purification. Subsequently, the *Dragon's Stronghold* leads the children to higher grounds, past the *East Mountain* and across *The Great River* before the journey ends in the *Magic Forest*. The world map, that also features ambiend music and lighting elements, reflects level selection and is considered the main "scene" of the game.

Each time children progress through the game, corresponding regions on the map are redeemed from the fog and adjacent areas call for their help, awaiting them with new challenges. The world's exploration is also reflected in the different environments used in levels. Each subchapter has a unique environment that continues the journey through a chapter, see Figure 7. This way, we wanted children to always be aware of their progress. As children will not master a subchapter each day, we included a daily progression. We embedded a fog and lighting system within levels (cf. prosodiya.com/level-run). For each task solved, the fog lightens up, glowing inhabitants of Prosodiya show up (e.g., fireflies), and other light sources like torches, lanterns, or sun rays appear to brighten the atmosphere. This low-level progress is also emphasized by a progress bar. We have chosen this multilevel progress, which also implicitly tells the story by progressing through the level's backgrounds, to increase the children's self-perception of progression, their perception of positive affect and immersion, and to maintain motivation over a longer period of time.

To date, we have implemented the prologue of the story to raise the children's interest, which can be accessed at prosodiya.com/story. All of the story's cutscenes are scripted and currently in the pipeline of development.

Pedagogical Agents

A commonly used approach to engage children in games are social interactions via companions [33]. They can help

Figure 10: Tutorial for the game "Stress Pattern". The big green Kugellicht explains the principle of syllable stress and that it – in this case – belongs to the first syllable.

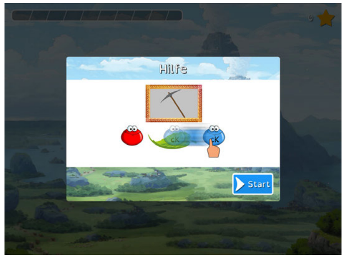


Figure 11: Exemplary tooltip for the game "Orthographic Markers".

players to understand game mechanics, teach game interactions, and resemble story-driven elements. In serious games, those companions are often referred to as pedagogical agents. Despite companions, agents are more focused on giving contextual feedback and explanations customized for the children in need [18, 37]. In Prosodiya, the Kugellichter are the pedagogical agents and serve different purposes in order to support and accompany the player.
Firstly, the agents introduce themselves as inhabitants of Prosodiya and establish a connection between the player and the world, explaining where they are and what is happening. They take the role of typical companions by the side of the children. In cut-scenes, the Kugellichter narrate the story and continue the storyline.
Secondly, the pedagogical agents take over the role as linguistic tutors. In each chapter, children get introduced to upcoming linguistic challenges and receive rule-based explanations by the agents on how to solve those challenges, see Section *Tutorials and Tooltips*.
Thirdly, the agents are also responsible for tutoring linguistic knowledge and provide scaffolding feedback to the input of the user, see Section *Feedback*.

To satisfy all needs, we designed the Kugellichter to be unisex (as advised by [30]), child friendly regarding color and shape, and to facilitate learning. We chose round shapes as this may induce positive emotions that facilitate learning and improve comprehension [39]. To ensure that color blindness or other color impairments do not affect learning, color was not the only unique feature of the pedagogical agents. The agents also differ in their size and in the shape of their mouth. Each agent provides a unique feature that links to its linguistic characteristics and supports clear distinction, (see Figure 9). There are four types of agents representing different linguistic characteristics. As mentioned before and communicated in the first tutorial, the yellow Kugellicht is smaller in size and represents unstressed syllables (see Figure 9a). Its counterparts are the bigger green (Figure 9b), blue (Figure 9d), and red (Figure 9c) Kugellichter who represent different stressed syllables. The green and blue Kugellicht look alike leaving out their colors. However, they do not occur simultaneously (see Section *Prosodiya*) and, thus, green-blue deficiencies do not adversely affect game play.

Tutorials and Tooltips
We developed interactive tutorials for each game with the Kugellichter as pedagogical agents and narrators. The tutorials serve different purposes. First and foremost, they introduce and explain game mechanics, linguistic competencies and challenges, and impart linguistic knowledge. This instructional support, particularly when focusing on learning new skills and selecting relevant (new) information, has been shown to improve learning [48]. In addition, the tutorials continue the storyline.
As the tutorials also start new sections of the game, they can be seen as on-boarding phases. The benefits of successful on-boarding phases are crucial for a long-term success of the game (cf. [17]) and thus are of special importance. Reports from previous pilot studies indicated that the tutorials were too long and too complicated. Therefore, we cut them short, kept them simple and fun, and tried to make sure the children understand game mechanics and the linguistic background. To proceed within a tutorial, children are frequently asked to actively solve the current step following the instructions of the Kugellichter, see Figure 10. We focused on a high level of interactivity to increase the children's participation and to ensure that they understand new game mechanics and linguistic principles. Multiple gameplay iterations showed that, in addition to linguistic explanations, it is important to teach and practice the interactions within the game before practicing new content.

Figure 12: 5-point Smileyometer.

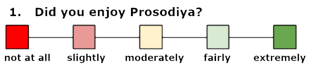


Figure 13: 5-point color and word coded rating scale. Rating ranges from *not at all* (red) to *extremely* (dark green)

Due to feedback of our focus group and that a chapter can cover broad linguistic categories as well as that game mechanics within a chapter may vary, just one detailed and comprehensive tutorial in the beginning of a chapter was not enough. Therefore, so called "Tooltips" - short and spot-on task explanations - were implemented at the beginning of each level, see Figure 11. Tooltips automatically show up at the start of a new level and can additionally be invoked manually by a button. Unlike the tutorials and due to limited time, the tooltips don not include animated pedagogical agents. The spot-on content is delivered auditive via the voice of the yellow Kugellicht and visually by displaying a simple image of the level's objective and challenges. Depending on the degree of difficulty, the player might also get additional hints on what has changed in the gameplay or what to pay attention to. General explanations of the Tooltip can be skipped while challenges specific to the level are mandatory to listen to, cf. prosodiya.com/level-run.
Lastly, tutorials and tooltips were included to make the game easy to start and stop playing, even if the children take longer breaks.

Feedback
Feedback in an educational context is considered to be crucial for knowledge improvement and skill acquisition and might affect motivation of learners (cf. [46]). Our game uses scaffolding and so-called knowledge of correct response (KRC) feedback. Scaffolding feedback may help dyslexic children to solve exercises faster [27] and KCR feedback has been shown to support memorization and deeper learning (e.g., [8, 15]). The feedback depends on the children's answer and is as follows: if the answer given to a task is correct, a positive sound is played, stars are collected and added to the current score, the progress bar is adjusted, and game elements respond positively, e.g. Kugellichter happily bounce up and down. A different, more sophisticated sound is played if the task is solved at the first go. In case of wrong answers, children are encouraged to try again. Affective encouragement may also positively affect their performance (cf. [45]). In addition, scaffolding feedback facilitates the task, e.g. replaying the word with increasingly emphasized intonation or in case of the spelling exercise, children may delete distracting letters, i.e. letters not present in the target word, or get single letters solved automatically. If they are not able to solve a word within three trials, the solution is displayed. When present, the pedagogical agents give spoken feedback as their empathic responses may positively impact learning [40]. The feedback mechanism for the first game can be watched at prosodiya.com/scaffolding-feedback.

Rewards
We designed different rewards for Prosodiya. Children can collect points when answering correctly. They are rewarded with more points solving a task at the first go to avoid trial-and-error behavior. Upon finishing a level, children are rewarded with a summary. Depending on their performance, the level might have been successfully mastered, unlocking subsequent game content. To account for poorer performing children and to avoid frustration, subsequent content is also unlocked after dynamically adapted number of level repetitions. To provide a high replay value and to increase training effects, we use a 1-3 star rating (i.e. more stars for higher performance) for each level displayed underneath each level on the world map. To date, collected points cannot be redeemed and only reflect in-game achievement.

Evaluation

Participants
In total, 137 German primary school children from second to fourth grade (age range 7–10 yrs.) took part in a randomized controlled field trial with a waiting control group

	Story
S1	Was the story of Prosodiya interesting?
	Kugellichter (Ped. Agents)
K1	How did you like the Kugellichter?
K2	Did you enjoy playing together with the Kugellichter?
K3	Did you feel like the Kugellichter are your friends?
K4	Do you think the Kugellichter are stupid?
	Map and Fog
M1	How did you like Prosodiya's map?
M2	Did the map make you feel you were inside the world of Prosodiya?
M3	How did you like the fog on the map?
M4	Did you enjoy dispelling the fog off the map?
M5	Did the fog on the map bother you?
	Environment and Fog
E1	How did you like the background images (environments of the levels)?
E2	How did you like the fog in the levels/exercises?
E3	Did you enjoy dispelling the fog in the levels to free the background image?
E4	Did you think the fog in the levels was stupid?
	Tutorials
T1	How well did the Kugellichter explain the secret of words?
T2	Did you enjoy the tutorials?
T3	Did the tutorials help you to understand the secret of words?
T4	Do you think the tutorials were too long?
	Tooltips
TT1	Did you understand the tooltips / short explanations?
TT2	Did the tooltips help you to better understand and solve the exercises?
TT3	Did the tooltips prior to the start of each level annoy you?

Table 1: Questions about the different game elements.

design. We recruited children with (suspected) dyslexia or very low reading and/or spelling proficiencies. Additionally, any interested second grader was encouraged to sign up to participate in a chronological age-matched control group regarding literacy skills to further evaluate game experience. We pseudo randomly assigned 69 children to the first intervention group and 68 children to the waiting control group, based on literacy skills assessed in pre-tests.

Here, we evaluate the data of the first intervention group who finished their training in May 2018. Six children were excluded from data analysis due to not finishing the intervention or data loss. The remaining 63 children were composed of 23 second (*m=13, f=10*), 29 third (*m=20, f=9*), and 11 fourth graders (*m=8, f=3*).

Materials

Game. A version of the game with a mainly linear course of play was used due to our research questions. The levels included all games described above with different levels of difficulty. During the intervention phase (9–10 weeks), children were given a tablet and asked to play the game at home 5 days per week, 20 minutes per day. We additionally handed out the training plan in form of a sticker book with a set of 40 stickers to keep the children on track.

Game Elements. To investigate the game elements described above, we evaluated a subset of 21 self-constructed questions from a questionnaire of 69 questions, see Table 1. We used 5-point Smileyometer (cf. [42]), Figure 12) and 5-point word (*not at all* to *extremely*) and color coded (*dark red* to *dark green*) rating scales, see Figure 13. Additionally, we evaluated answers from two free text fields, which gave the children the chance to leave qualitative feedback and tell us what they specifically liked about Prosodiya, what could be improved, or what was missing. Due to the scope of the randomized controlled field trial, the questionnaire was quite long. In future, shorter questionnaires will be used.

Procedure

After the children from the first intervention group played the game for 9–10 weeks, they answered the questionnaire after 45 minutes classroom testing of literacy skills: We first explained the rating scales to the children and gave explicit examples for positive and negative items with mock-up questions (e.g, "I like chocolate" vs. "I hate gummy bears"). Moreover, preliminary tests indicated that children had problems reading and understanding the questions due to their lack of proficient reading skills and unfamiliarity with such questionnaires. Thus, we read aloud each question individually and clarified posed questions to ensure that everyone understood the items. We continued with succeeding questions after every child had answered the preceding one. It took approximately 20 minutes to answer the questionnaire. Upon completion, children were rewarded with small toy dinosaurs and flexible pencils.

Results

We transformed the answers into values 1 to 5. Answers with no clearly selected option were excluded. We transformed answers into floating point values where the middle of two adjacent options was marked. The results are summarized as box plots in Figures 14–18.

Narrative and Environment. Children reported to be interested in the story (*S1*, M = 3.93, SD = 1.12), that they liked the map of Prosodiya (*M1*, M = 4.43, SD = 0.77), and that they felt immersed into the fantasy world (*M2*, M = 3.62, SD = 1.54). They also reported that they liked the fog that covered the map (*M3*, M = 3.89, SD = 1.32), that they enjoyed to dispel the fog (*M4*, M = 4.12, SD = 1.08),

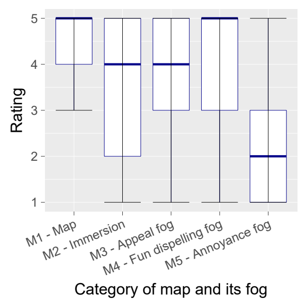


Figure 14: Ratings of Prosodiya's map and the fog covering it.

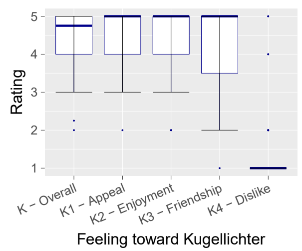


Figure 15: Ratings of our pedagogical agents.

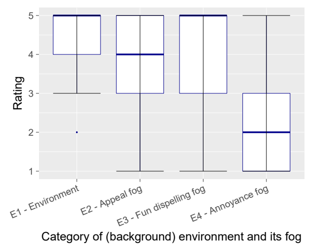


Figure 16: Ratings of background environments and fog that was dispelled within a level.

and that they weren't bothered by the fog (*M5*, *M* = 2.43, *SD* = 0.54). Results of the map and its fog are summarized in Figure 14.

Regarding the environments of subchapters and levels, children reported that they liked the background images used (*E1*, *M* = 3.89, *SD* = 1.32) and the fog that they needed to fight back during a level (*E2*, *M* = 3.86, *SD* = 1.39), see Figure 16. They enjoyed clearing the environment of its fog (*E3*, *M* = 4.23, *SD* = 1.07) while not being annoyed by the fog (*E4*, *M* = 2.36, *SD* = 1.15).

In qualitative feedback, children emphasized that they really liked the graphics of Prosodiya, that the map was "like a real world" and one of Prosodiya's highlights, and that "the mission to dispel the fog was great". As the study version of the game did not have a final sequence marking the end of the game, children proposed possible endings, such as a "castle in which the children can interact with the inhabitants" or a cut-scene with the sunrise of Prosodiya that was used as a background image in one of the levels.

Pedagogical Agents. The ratings of our Kugellichter, pedagogical agents and narrators of Prosodiya, are displayed in Figure 15. We combined the four questions about the agents *K1–K4* into a single value *K*, inverting the response to item *K3* (dislike). Children reported a positive overall impression of the agents (*K*, *M* = 4.44, *SD* = 0.71, Cronbach's α = .78), high appeal (*K1*, *M* = 4.41, *SD* = 0.75), and that they enjoyed to play with them together (*K2*, *M* = 4.51, *SD* = 0.82). They also reported to consider them as their friends (*K3*, *M* = 4.15, *SD* = 1.15) and that they experienced very low dislike towards the Kugellichter (*K4*, *M* = 1.32, *SD* = 0.89).

Children qualitatively reported about the Kugellichter that they were one of the best aspects of Prosodiya, that they would like to have them as cuddly toys, and that they especially liked the yellow Kugellicht responsible for spoken feedback. The uniqueness of each Kugellicht was also emphasized positively.

Tutorials and Tooltips. Results of the tutorials are summarized in Figure 17. Children reported that the pedagogical agents explained the game mechanics and linguistic knowledge well (*T1*, *M* = 4.27, *SD* = 0.87), that the tutorials were kind of fun (*T2*, *M* = 3.42, *SD* = 1.19), and that they were very useful in order to learn the "secret of words" (*T3*, *M* = 4.11, *SD* = 1.16). The duration of the tutorials were perceived to be not too long (*T4*, *M* = 2.46, *SD* = 1.44).

The tooltips were rated to be very comprehensible (*TT1*, *M* = 4.35, *SD* = 0.88) and to be very useful (*TT2*, *M* = 3.72, *SD* = 1.27) while not annoying the children (*TT3*, *M* = 2.5, *SD* = 1.37), see Figure 18.

Feedback and Rewards. Although we did not ask questions about feedback or rewards, children used the comment fields of the questionnaire to talk about it. Regarding feedback, they reported the praising words spoken by the yellow Kugellicht as one of Prosodiya's highlights. Other children complained that they were missing the joyful feedback in the later course of the game, in which the yelllow Kugellicht was no longer a game element. This was especially the case when exercises became harder and one of the children's main challenges – spelling – was dealt with. Regarding rewards, we investigated the times children additionally practiced levels after their completion and identified that children eagerly tried to gather three stars for each level. In the beginning of the intervention, we defined that a flawless level grants the third star. We realized that poorer performing children had major problems solving all exercises in a level at the first go and hence were stuck in the game, trying to get the third star. In response, we lowered the threshold to 90% and implemented a fall-back system

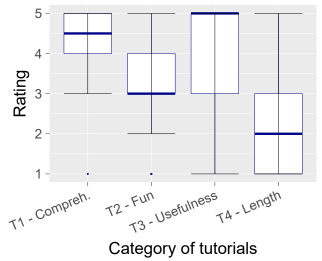


Figure 17: Ratings of the tutorials.

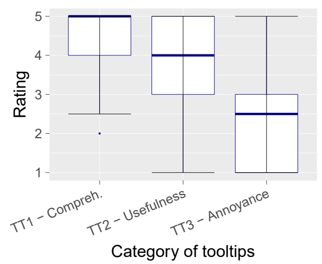


Figure 18: Ratings of the tooltips.

Acknowledgments

This research was funded by the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments. Heiko Holz is a doctoral student at the LEAD Graduate School & Research Network.

granting the third star after a level has been practiced for a defined number of times.

Discussion

The results indicate overall very positive reactions to the game elements embedded in Prosodiya. The graphical appearance of Prosodiya, which includes the design of the pedagogical agents, the world map, and the background environments of levels, was rated very high by the children. This indicates that we appealed the children's taste and that they felt comfortable in the world of Prosodiya. The story that was introduced with a prologue and continued implicitly in tutorials, the world map, and by changing level environments, has raised the children's interest. Also, the fog was received well, which children needed to dispel from the game's world map and its levels to increase their self-perception of progression and to reward them. Qualitative feedback of the children about the map and suggestions of how the story might end imply involvement and that they thought about it.

The Kugellichter, our pedagogical agents and narrators, were very popular with children and were even regarded as their friends. It also seemed that children really appreciated the positive feedback from the yellow Kugellicht.

The refined tutorials and newly introduced tooltips served their purpose to explain game mechanics and to convey linguistic knowledge. Children reported to enjoy the tutorials and did not perceive them as too long. We also infer that the tooltips popping up prior to each level were good solutions as a regular reminder of task objectives and linguistic properties to pay attention to without being annoying. Although feedback was not directly addressed in questions, children reported qualitatively that they really appreciated the positive feedback spoken by the yellow Kugellicht and missed it in later course of play, emphasizing its necessity. Finally, the number of times children practiced old levels, even when new content had been unlocked, revealed that the stars awarded to level completions promoted competition with self and supported replayability of levels.

Conclusion and Outlook

In this article, we presented the design and evaluation of the game elements of a mobile game for German dyslexic children. These elements include graphics, narrative, a world map, environment, pedagogical agents, tutorials, as well as feedback and rewards. Due to the specific target group, we put special emphasis on designing the elements to appeal to primary school boys and girls alike. Participating children indicated very positive perception of the game's graphics and elements. We conclude that the story and map supported a sense of immersion and self-perception of progression, that the pedagogical agents the kids have befriended during game-play were very popular, and that tutorials and tooltips provided easy introductions into games. Qualitative feedback by children implied that they engaged with the game and have been thinking about it and its elements. Lastly, the importance of positive feedback and praise was emphasized especially in later course of play that dealt with spelling. To conclude, our results encourage the use of the aforementioned game elements in digital game-based learning environments.

We plan to complete the integration of the story and to develop an extrinsic reward systems to foster long-term motivation that features a dragon's stronghold and cave in the future: children may redeem the points earned in levels to buy a dragon egg, incubate it, and raise the hatched dragon to be their companion and play bonus games with. Additionally, we plan to attune the different used graphic styles to each other and also consider customization of the pedagogical agents, as this has been shown to increase engagement [5].

REFERENCES

1. Philip Abrami, Eugene Borohkovski, and Larysa Lysenko. 2015. The Effects of ABRACADABRA on Reading Outcomes: A Meta-Analysis of Applied Field Research. *Journal of Interactive Learning Research* 26, 4 (2015), 337–367.
2. Chadia Abras, Diane Maloney-Krichmar, and Jenny Preece. 2004. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications* 37, 4 (2004), 445–456. DOI: http://dx.doi.org/10.3233/WOR-2010-1109
3. Sylvester Arnab, Theodore Lim, Maira B. Carvalho, Francesco Bellotti, Sara de Freitas, Sandy Louchart, Neil Suttie, Riccardo Berta, and Alessandro De Gloria. 2015. Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology* 46, 2 (mar 2015), 391–411. DOI: http://dx.doi.org/10.1111/bjet.12113
4. Kay Berkling. 2017. Phontasia: A phonics game for German and its effect on orthographic skills–first corpus explorations. In *WOCCI 2017: 6th International Workshop on Child Computer Interaction*. ISCA, ISCA, 7–16. DOI: http://dx.doi.org/10.21437/WOCCI.2017-2
5. Max V. Birk, Cheralyn Atkins, Jason T. Bowey, and Regan L. Mandryk. 2016. Fostering Intrinsic Motivation through Avatar Identification in Digital Games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, New York, New York, USA, 2982–2995. DOI: http://dx.doi.org/10.1145/2858036.2858062
6. Elizabeth A. Boyle, Thomas Hainey, Thomas M. Connolly, Grant Gray, Jeffrey Earp, Michela Ott, Theodore Lim, Manuel Ninaus, Claudia Ribeiro, and João Pereira. 2016. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education* 94, Supplement C (2016), 178–192. DOI: http://dx.doi.org/10.1016/j.compedu.2015.11.003
7. Luciana Cidrim and Francisco Madeiro. 2017. Information and Communication Technology (ICT) applied to dyslexia: literature review. *Revista CEFAC* 19, 1 (feb 2017), 99–108. DOI: http://dx.doi.org/10.1590/1982-021620171917916
8. Gemma Corbalan, Liesbeth Kester, and Jeroen J.G. van Merriënboer. 2009. Dynamic task selection: Effects of feedback and learner control on efficiency and motivation. *Learning and Instruction* 19, 6 (dec 2009), 455–465. DOI: http://dx.doi.org/10.1016/j.learninstruc.2008.07.002
9. Diana I. Cordova and Mark R. Lepper. 1996. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology* 88, 4 (1996), 715–730. DOI: http://dx.doi.org/10.1037/0022-0663.88.4.715
10. Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining "Gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11)*. ACM, New York, NY, USA, 9–15. DOI: http://dx.doi.org/10.1145/2181037.2181040

11. Athanasios Drigas and Elektra Batziaka. 2016. Dyslexia and ICTs, Assessment and Early Intervention in Kindergarten. *International Journal of Emerging Technologies in Learning (iJET)* 11, 02 (feb 2016), 53. DOI:http://dx.doi.org/10.3991/ijet.v11i02.5193
12. Allison Druin. 1999. Cooperative Inquiry: Developing New Technologies for Children with Children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM Press, New York, New York, USA, 592–599. DOI: http://dx.doi.org/10.1145/302979.303166
13. Allison Druin. 2002. The role of children in the design of new technology. *Behaviour & Information Technology* 21, 1 (jan 2002), 1–25. DOI: http://dx.doi.org/10.1080/01449290110108659
14. Kevin Durkin, James Boyle, Simon Hunter, and Gina Conti-Ramsden. 2013. Video Games for Children and Adolescents With Special Educational Needs. *Zeitschrift für Psychologie* 221, 2 (jan 2013), 79–89. DOI:http://dx.doi.org/10.1027/2151-2604/a000138
15. Severine Erhel and Eric Jamet. 2013. Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education* 67, September (sep 2013), 156–167. DOI: http://dx.doi.org/10.1016/j.compedu.2013.02.019
16. Tracy Fullerton. 2008. *Game Design Workshop: A Playcentric Approach to Creating Innovative Games* (2nd ed.). Morgan Kaufmann, Heidelberg. https://books.google.de/books?id=OjIYWtqWxtAC
17. Gabe Zichermann and Christopher Cunningham. 2011. *Gamification by Design-Implementing Game Mechanics in Web and Mobile Apps*. Vol. 27. Cambridge University Press, Cambridge. 208 pages. http://ebooks.cambridge.org/ref/id/CBO9781107415324A009
18. Benjamin Goldberg and Janis Cannon-Bowers. 2015. Feedback source modality effects on training outcomes in a serious game: Pedagogical agents make a difference. *Computers in Human Behavior* 52, 1-2 (nov 2015), 1–11. DOI: http://dx.doi.org/10.1016/j.chb.2015.05.008
19. Abbie Gonzalez. 2014. OpenDyslexic. (2014). https://www.opendyslexic.org/
20. Usha Goswami, Natasha Mead, Tim Fosker, Martina Huss, Lisa Barnes, and Victoria Leong. 2013. Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language* 69, 1 (2013), 1–17. DOI: http://dx.doi.org/10.1016/j.jml.2013.03.001
21. Thomas Hainey, Thomas M. Connolly, Elizabeth A. Boyle, Amanda Wilson, and Aisya Razak. 2016. A systematic literature review of games-based learning empirical evidence in primary education. *Computers & Education* 102 (2016), 202–223. DOI: http://dx.doi.org/10.1016/j.compedu.2016.09.001
22. Heiko Holz, Katharina Brandelik, Benedikt Beuttler, Jochen Brandelik, and Manuel Ninaus. In press 2018. How to train your syllable stress awareness – A digital game-based approach for German dyslexic children. *International Journal of Serious Games* (In press 2018).

23. Heiko Holz, Katharina Brandelik, Jochen Brandelik, Benedikt Beuttler, Alexandra Kirsch, Jürgen Heller, and Detmar Meurers. 2017. Prosodiya – A Mobile Game for German Dyslexic Children. In *Games and Learning Alliance: 6th International Conference, GALA 2017, Lisbon, Portugal, December 5–7, 2017, Proceedings*, João Dias, Pedro A. Santos, and Remco C. Veltkamp (Eds.). 73–82. DOI: http://dx.doi.org/10.1007/978-3-319-71940-5_7

24. Heiko Holz, Manuel Ninaus, Detmar Meurers, and Alexandra Kirsch. 2018. Validitiy and Player Experience of a Mobile Game for German Dyslexic Children. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18 Extended Abstracts)*. ACM. DOI: http://dx.doi.org/10.1145/3270316.327152

25. Elena Ise and Gerd Schulte-Körne. 2010. Spelling deficits in dyslexia: Evaluation of an orthographic spelling training. *Annals of Dyslexia* 60, 1 (2010), 18–39. DOI: http://dx.doi.org/10.1007/s11881-010-0035-8

26. Gracia Jiménez-Fernández, Nicolás Gutiérrez-Palma, and Sylvia Defior. 2015. Impaired stress awareness in Spanish children with developmental dyslexia. *Research in Developmental Disabilities* 37 (feb 2015), 152–161. DOI: http://dx.doi.org/10.1016/j.ridd.2014.11.002

27. Maria N. Kazakou and Spiros Soulis. 2015. Feedback and the speed of answer of pupils with dyslexia in digital activities. *Procedia - Procedia Computer Science* 67 (2015), 204–212. DOI: http://dx.doi.org/10.1016/j.procs.2015.09.264

28. Fengfeng Ke and Tatiana Abras. 2013. Games for engaged learning of middle school children with special learning needs. *British Journal of Educational Technology* 44, 2 (2013), 225–242. DOI:http://dx.doi.org/10.1111/j.1467-8535.2012.01326.x

29. Christian Klicpera and Barbara Gasteiger Klicpera. 2000. Sind Rechtschreibschwierigkeiten Ausdruck einer phonologischen Störung? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 32, 3 (jul 2000), 134–142. DOI: http://dx.doi.org/10.1026//0049-8637.32.3.134

30. Jonna Koivisto and Juho Hamari. 2014. Demographic differences in perceived benefits from gamification. *Computers in Human Behavior* 35 (jun 2014), 179–188. DOI: http://dx.doi.org/10.1016/j.chb.2014.03.007

31. Karin Landerl. 2003. Categorization of vowel length in German poor spellers: An orthographically relevant phonological distinction. *Applied Psycholinguistics* 24, 4 (2003), 523–538. DOI: http://dx.doi.org/10.1017/S0142716403000262

32. Victoria Leong, Jarmo Hämäläinen, Fruzsina Soltész, and Usha Goswami. 2011. Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language* 64, 1 (2011), 59–73. DOI: http://dx.doi.org/10.1016/j.jml.2010.09.003

33. Sohye Lim and Byron Reeves. 2010. Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player. *International Journal of Human-Computer Studies* 68, 1-2 (jan 2010), 57–68. DOI: http://dx.doi.org/10.1016/j.ijhcs.2009.09.008

34. Manuel Ninaus, Kristian Kiili, Jake McMullen, and Korbinian Moeller. 2016. A Game-Based Approach to Examining Students' Conceptual Knowledge of Fractions. In *Games and Learning Alliance: 5th International Conference, GALA 2016, Utrecht, The Netherlands, December 5–7, 2016, Proceedings*, Rosa Bottino, Johan Jeuring, and Remco C. Veltkamp (Eds.). Springer International Publishing, 37–49. DOI: http://dx.doi.org/10.1007/978-3-319-50182-6_4

35. Manuel Ninaus, Matthias Witte, Silvia E Kober, Elisabeth V C Friedrich, Erwin Hartsuiker, Christa Neuper, Guilherme Wood, Jürgen Kurzmann, Erwin Hartsuiker, Christa Neuper, and Guilherme Wood. 2014. Neurofeedback and Serious Games. In *Psychology, Pedagogy, and Assessment in Serious Games*, Thomas M Connolly, Thomas Hainey, Elizabeth Boyle, Gavin Baxter, and Pablo Moreno-Ger (Eds.). IGI Global, Hershey, PA, USA, 82–110. DOI: http://dx.doi.org/10.4018/978-1-4666-4773-2.ch005

36. Donald A Norman and Stephen W Draper. 1986. User Centered System Design; New Perspectives on Human-Computer Interaction. *Journal educational computing research* 3, 1 (1986), 129–134.

37. Hyacinth Nwana. 1990. Intelligent tutoring systems: an overview. *Artificial Intelligence Review* 4 (1990), 251–277. DOI: http://dx.doi.org/10.1007/BF00168958

38. Louise E. Parker, Mark R. Lepper, Kim Bartholomew, Diana Cordova, and John Mayer. 1992. Effects of fantasy contexts on children's learning and motivation: Making learning more fun. *Journal of personality and social psychology* 62, 4 (1992), 625–633.

39. Jan L. Plass, Steffi Heidig, Elizabeth O. Hayward, Bruce D. Homer, and Enjoon Um. 2014. Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction* 29 (2014), 128–140. DOI: http://dx.doi.org/10.1016/j.learninstruc.2013.02.006

40. Jan L. Plass, Bruce D. Homer, and Charles K. Kinzer. 2015. Foundations of Game-Based Learning. *Educational Psychologist* 50, 4 (oct 2015), 258–283. DOI: http://dx.doi.org/10.1080/00461520.2015.1122533

41. Clark N. Quinn. 2005. *Engaging learning: Designing e-learning simulation games*. John Wiley & Sons.

42. Janet C. Read. 2008. Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. *Cognition, Technology & Work* 10, 2 (apr 2008), 119–128. DOI: http://dx.doi.org/10.1007/s10111-007-0069-9

43. Carola Reuter-Liehr. 1993. Behandlung der Lese-Rechtschreibschwäche nach der Grundschulzeit: Anwendung und Überprüfung eines Konzeptes. *Zeitschrift für Kinder- und Jugendpsychiatrie* 21, 3 (1993), 135–147.

44. Katharina Sauter, Jürgen Heller, and Karin Landerl. 2012. Sprachrhythmus und Schriftspracherwerb. *Lernen und Lernstörungen* 1, 4 (2012), 225–239. DOI: http://dx.doi.org/10.1024/2235-0977/a000023

45. Kelly L. Schmitt, Lisa B. Hurwitz, Laura Sheridan Duel, Deborah L Nichols Linebarger, and Deborah L. Nichols Linebarger. 2018. Learning through play: The impact of web-based games on early literacy development. *Computers in Human Behavior* 81 (2018), 378–389. DOI:http://dx.doi.org/10.1016/j.chb.2017.12.036

46. Valerie J. Shute. 2008. Focus on Formative Feedback. *Review of Educational Research* 78, 1 (mar 2008), 153–189. DOI: http://dx.doi.org/10.3102/0034654307313795
47. Sven Staffeldt. 2010. *Einführung in die Phonetik, Phonologie und Graphematik des Deutschen. Ein Leitfaden für den akademischen Unterricht*. Stauffenburg, Tübingen.
48. Pieter Wouters, Christof van Nimwegen, Herre van Oostendorp, and Erik D van der Spek. 2013. A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology* 105, 2 (2013), 249–265. DOI: http://dx.doi.org/10.1037/a0031311

Validity and Player Experience of a Mobile Game for German Dyslexic Children

Heiko Holz[1,3]
heiko.holz@uni-tuebingen.de

Detmar Meurers[1,2,3]
dm@sfs.uni-tuebingen.de

Manuel Ninaus[1,2]
m.ninaus@iwm-tuebingen.de

Alexandra Kirsch
Independent Scientist
Stuttgart, Germany
kirsch@alexkirsch.de

[1]LEAD Graduate School & Research Network, University of Tuebingen, Germany
[2] Leibniz-Institut für Wissensmedien, Tuebingen, Germany
[3] ICALL Research Group, icall-research.com, University of Tuebingen, Germany

CHI PLAY'18 Extended Abstracts, October 28–31, 2018, Melbourne, VIC, Australia.

ACM ISBN 978-1-4503-5968-9/18/10.
http://dx.doi.org/10.1145/3270316.3271523

Abstract
Approximately 4–10% of the German population suffer from developmental dyslexia, influencing children's educational, personal, and social development negatively. Digital interventions have shown great promise to additionally support dyslexic children outside of school or learning therapy. We present the results of a mobile serious game for German dyslexic children to improve reading and spelling performance with special emphasis on syllable stress awareness. We evaluate player experience and investigate the relationship between real-life literacy skills and in-game data of 63 children who played the game at home for 9–10 weeks within the scope of a randomized controlled field trial. Results indicate positive player experience and a completion rate of 75% indicates the feasibility of unsupervised digital game-based interventions. Moreover, real-life reading and spelling proficiencies correlated significantly with processing times and scores measured in-game, providing first evidence of the game's validity.

CCS Concepts
•Applied computing → Computer games; Interactive learning environments; •Human-centered computing → Empirical studies in HCI;

Author Keywords
Digital Game-Based Learning; Player Experience; Dyslexia

Figure 1: Game 1 “Stress Pattern”.

Figure 2: Game 2 “Open and Closed Syllables”.

Figure 3: Game 3 “Orthographic Marker”.

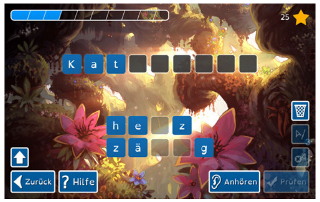


Figure 4: Game 4 “Spelling”.

Introduction

Dyslexia is one of the most frequent learning disorders, affecting 4–10 % of the German population [27, 28]. The learning disorder negatively affects educational, personal, and social development of children [2, 8], thus appropriate interventions are needed to prevent negative consequences in the long run. Mobile and computer-based interventions have shown great promise to support the acquisition of reading and spelling for dyslexic primary school children (e.g., [3, 6, 10, 19, 24, 37]). Game elements used in digital interventions can explicitly address negative feelings such as frustration, demotivation or boredom [9], and support successful learning [5, 42].
Despite the importance of playful digital interventions to deliver high player experience and to balance game play with educational effectiveness and quality of learning [1, 4, 20, 34], these factors have not been studied systematically, i.e. beyond short questionnaires or observations.

In this article, we propose a mobile serious game for German dyslexic children called “Prosodiya”. We investigate player experience as well as the validity of our pedagogical approach based on the results of 63 primary school children who played the game at home during a period of 9–10 weeks within the scope of a randomized controlled field trial. Consequently, we specifically address i) if the proposed game delivers high player experience, ii) if such playful digital interventions are feasible for use at home, and iii) whether we can find preliminary evidence of our pedagogical approach to improve reading and spelling by analyzing in-game times and scores.

This article starts with an introduction to the game, followed by the results and discussion of our study. We conclude with an outlook to future data analyses and game development.

The Game

Prosodiya [15] is a mobile serious game based on recent empirical findings and evidence-based interventions (e.g., [17, 38]). Prosodiya differs from similar games in that it trains syllable stress awareness, which highly correlates with reading and spelling skills [39] and is impaired in dyslexic children [11, 18, 25]. One explanation is thought to be found in the association between stress and German orthographic markers. Vowel length markers, i.e. graphemes marking long or short vowels, generally occur in stressed syllables [41]. Mastering the complex orthographic rules to mark long and short vowels is a major difficulty for German children [22, 23].

The focus of Prosodiya is primarily on spelling acquisition by training the awareness of linguistic features related to syllable stress and linking these features to orthographic regularities of German orthography. By shifting children’s attention to relevant areas of words, Prosodiya aims to clarify the association between syllable stress and orthographic marking, such as vowel or consonant doubling, to learn how such syllables are spelled. Figure 1 – Figure 4 display the four games used in the intervention for the word *Katze* (cat), whose short vowel is marked with the graphemes *tz*. In the first game, children rebuild stress patterns of words by dragging and dropping cartoon blobs onto platforms, a big green blob for stressed and small yellow blobs for unstressed syllables. The second game is a novel variant of vowel length distinction tasks. Children additionally need to decide whether the stressed syllable is open (ends with a long vowel, big red blob with an open mouth) or closed (vowel is closed by a consonant, big blue blob with its mouth shut), see Figure 2. The recognition of orthographic markers, i.e. spelling of long and short vowels, is the subject of the third game (Figure 3). In the fourth game

Figure 5: In-game map of the game. Glass blossoms are used as level symbols.

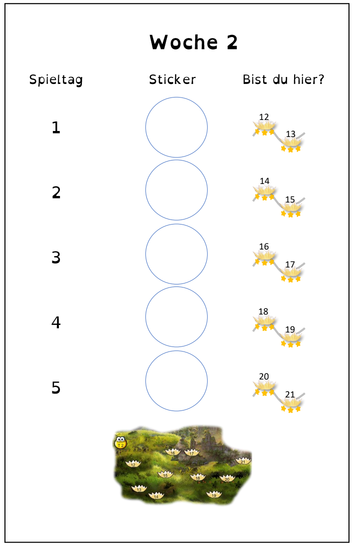


Figure 6: One week of the training plan in the sticker book.

(Figure 4), children finally spell words and thereby foster their previously acquired knowledge.

Prosodiya's overall narrative is about the deliverance of the eponymous world from a mysterious fog that has arisen (Figure 5). Little inhabitants called "Kugellichter" (spherical lights), the game's protagonists and pedagogical agents, call for the children's help. To redeem the inhabitants from their sorrowful lives, only children, accompanied by the Kugellichter, can disperse the suppressing fog by mastering linguistic challenges. Progressing through the course of the game, parts of Prosodiya are saved, and new regions await the children with challenges to be mastered.

Evaluation

In this article, we evaluate the general validity and player experience of Prosodiya. A detailed evaluation of individual game elements is reported in [14].

Participants

In total, 137 German primary school children from second to fourth grade (age range 7–10 yrs.) took part in a randomized controlled field trial with a waiting control group design. We recruited children with (suspected) dyslexia or very low reading and/or spelling proficiencies. Additionally, any interested second grader was encouraged to sign up to participate in a chronological age-matched control group regarding literacy skills to further evaluate player experience. Based on spelling and reading proficiency assessed in pre-tests, we pseudo randomly assigned 69 children to the first intervention group and 68 to the waiting control group.

In this article, we evaluate the data of the first intervention group whose training ended in May 2018. Six children were excluded from data analysis due to not finishing the intervention or data loss. The remaining 63 children was comprised as follows: 23 second (*m=13, f=10*), 29 third (*m=20, f=9*), and 11 fourth graders (*m=8, f=3*). As 13 children answered a preliminary version of the questionnaire without items of the Game Experience Questionnaire (GEQ), respective subscales of game experience are evaluated based on the results of 50 children (18 second [*m=8, f=10*], 23 third (*m=15, f=8*), and 9 fourth graders [*m=7, f=2*]).

Materials

Game and training plan. A version of the game – as described above – with a mainly linear course of play was used due to our research questions. Therefore, adaption was limited to the number of level repetitions and word selection. A training plan of 8 weeks in the form of a sticker book with a set of 40 stickers was used to keep the children on track. The sticker book depicted for each training day and week the levels to be practiced, see Figure 6. Each page corresponded to one training week and was in line with the map used in the game (Figure 5). Due to school holidays during training, we deployed more levels than displayed in the sticker book. We deployed in total 80 levels. The training plan officially ended at level 66, labeling the rest as bonus. To avoid binge-playing and loss of training effect, content of a new training week was unlocked on Monday mornings. During the intervention phase (9–10 weeks), children were given a tablet and were asked to play the game at home 5 days per week, 20 minutes per day following the training plan.

Reading, spelling, and syllable stress awareness. Reading and spelling skills in pre- and post-tests were assessed using standardized classroom tests of spelling [30, 31, 12], reading speed [26] and individually administered standardized tests of reading fluency [29]. Syllable stress awareness was assessed using an individually administered paper version of the game "Stress Pattern" (Figure 1).

Overall Impression

- How much did you like Prosodiya?

Usability

- Did you quickly understand how to play the game?
- Do you think the game is easy to use?
- Did you always know what to do while playing?
- In the different exercises, was it always clear to you what you had to do?

Self-efficacy

- How much did you learn in this game with regard to reading and spelling?
- Did the game help you to learn to read?
- Did the game help you to learn to spell?
- Did the training increase your confidence in German classes?
- How often do you think about the things that you learned in the game when you don't know how to spell a word?

Intention to use

- Would you like to continue playing with Prosodiya?

Likelihood to recommend

- Would you go tell a friend Prosodiya is a good game?

Game or homework

- Do you think Prosodiya is more like homework or more like a game?

Table 1: Questions for additional subscales of player experience.

	items iGEQ	*item+*	α_{iGEQ}	$\alpha_{\text{iGEQ+}}$	α^*
Positive affect	[1,14]	4	.76	.79	.80
Competence	[17,2]	15	.62	.66	.83
Immersion	[3,27]	12	.75	.82	.81
Challenge	[26,33]	11	.70	.74	.74
Negative affect	[16,9]	7	.50	.25	.71
Tension/Annoyance	[29,24]	22	.65	.79	.82

Table 2: Cronbach's alpha of the items used from the iGEQ (α_{iGEQ}) and iGEQ+($\alpha_{\text{iGEQ+}}$). Item number refers to an item's number in the core GEQ [16].

Player experience. Player experience was evaluated based on a subset of a questionnaire of 69 questions.

First, we evaluated 19 questions from the Game Experience Questionnaire [16] (GEQ) using a 5-point word and color coded rating scale (Figure 8). We refer to the 19 questions from the GEQ as the iGEQ_+ that was composed of the in-game GEQ (iGEQ) and 6 additional questions from the GEQ's core module. The GEQ is intended to measure the subscales *Positive affect*, *Competence*, *Sensory & imaginative immersion*,*Challenge*, *Flow*, *Negative affect*, and *Tension/Annoyance*, see Table 2.
The iGEQ_+ adds one additional item to all subscales except flow. The items as well as Cronbach's alpha for the subscales of the iGEQ and iGEQ_+ are shown in Table 2. Due to an increase in Cronbach's alpha, we kept the additional item for all subscales except *Negative affect*, which resulted in a severe decrease of Cronbach's alpha.

In addition to the iGEQ_+, we evaluated 12 self-constructed questions covering the children's *Overall impression* of the game and the subscales *Usability*, *Self-efficacy*, *Intention to use*, *Likelihood to recommend*, and whether Prosodiya feels more like homework or like a game, see Table 1. Likelihood to recommend is inspired by the net promoter score by [36], which is one of the simplest loyalty measures.
We used either a 5-point Smileyometer [35] (Figure 7), a bipolar rating scale, or the same scale used for the iGEQ_+.

Procedure
First, we administered the pre-test (T1) involving tests of spelling, reading speed and fluency, and syllable stress awareness. Second, children of the first intervention group performed 9–10 weeks of training. The waiting control group did not play the game and thus is not considered in the current analysis.Third, the post-test (T2) was administered similarly to T1. In addition, children of the first intervention group answered the questionnaire, for which we explained to the children that they now have the chance to express anonymously what they think about the game with no right or wrong answers. We explained the rating scales of the questionnaire and provided explicit examples for positive and negative items with mock-up questions to identify possible careless responses (e.g, "I like chocolate" vs. "I hate gummy bears"). Moreover, preliminary tests indicated that children had problems reading and understanding the questions due to their lack of proficient reading skills and unfamiliarity with such questionnaires. Thus, we read aloud each question individually and clarified posed questions to ensure that everyone understood the items. We continued with subsequent questions after every child had answered the previous one. Answering the questionnaire took approximately 20 minutes. Upon completion, children were rewarded with flexible pencils and small toy dinosaurs.

Results

Answers of the questionnaire were transformed into values 1 to 5. Answers with no clearly selected options were excluded.If children put marks between two options, we kept and transformed the answer into a floating point value.

Player experience. Mean values of subscales were considered to reflect player experience. We used a conservative approach of analyzing each subscale by conducting one sample Wilcoxon signed rank tests against the middle value of the subscale's 5-point Likert scale (3 = moderately). Descriptive results and inferential statistics are summarized in Figure 9 and 10.

Children's ratings of the game on the subscales *Positive affect*, *Competence* and *Immersion* of the iGEQ_+ were significantly higher than moderately. In contrast, ratings of the *Tension/Annoyance*, *Negative affect*, and *Challenge* sub-

Figure 7: 5-point Smileyometer.

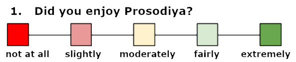


Figure 8: 5-point color and word coded rating scale.

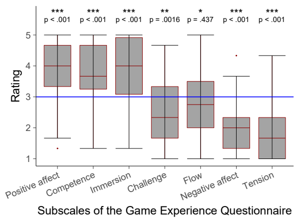


Figure 9: Results of player experience based on the iGEQ+ and one-sample Wilcoxon signed rank test compared to μ=3.

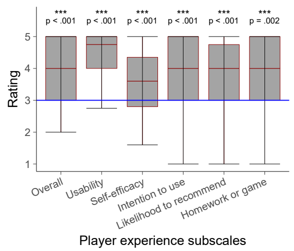


Figure 10: Results of player experience subscales and one-sample Wilcoxon signed rank test compared to μ=3.

scales were significantly lower than moderately, see Figure 9. Ratings of the *Flow* subscale did not differ significantly from moderately, see Figure 9. Children reported a significantly positive *Overall impression*, rated the game's *Usability* to be very good, and reported a feeling of *Self-efficacy*, indicated by ratings significantly higher than moderately, see Figure 10. In addition, children reported that they would likely recommend the game to a friend and continue playing the game themselves. Finally, children rated Prosodiya to be more like a game , as reflected by ratings significantly above "neither homework nor game".

We correlated the variables of literacy skills assessed in pre-tests with the subscales of player experience. We opted for Spearman's rank correlation due to non-normal distribution of the data. Results are reported in Figure 11.

In-game measures. Children on average played on 27.9 (*SD* = 11.5) out of the recommended 40 days, spent on average 494 min. (*SD* = 194) playing, reached on average level 68.6 (*SD* = 17.5), and practiced on average 160.6 levels (*SD* = 47.4). Out of 63 children, 47 (74.6%) fulfilled their training plan and reached level 66 or higher. Children spent on average 3.1 min. (*SD* = 0.8) and scored on average 139.6 (*SD* = 5.4) out of 150 points per level consisting of 10 tasks. They solved on average 8.3 (*SD* = 0.8) out of 10 tasks per level at the first go.

Detailed correlations between literacy proficiencies and in-game data of times and scores are listed in Figure 12.

Discussion

The results indicate an overall positive perception of the proposed game and prove its usability and application as an intervention used at home. Inferred from the results of the iGEQ+, children reported high positive affect and to feel

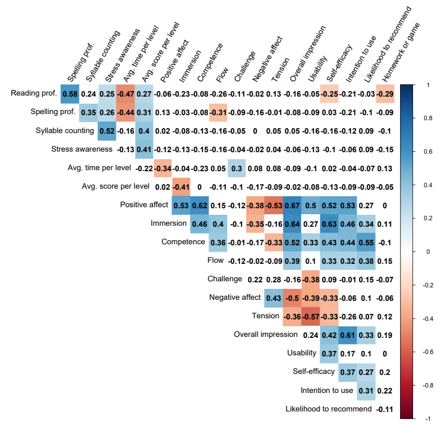


Figure 11: Spearman's Rank correlation between literacy skills and player experience. Correlations significant on $\alpha = .05$ are colored.

competent and immersed while playing. Furthermore, the children did not perceive negative affect nor did they feel tense or annoyed during game play. Positive engagement has been shown to positively affect learning [13].
The results additionally imply that the game did not overstrain the children but also that it might have been too easy. The low value of challenge could be explained by the high percentage of tasks solved at the first go or by the study version of the game that increased the difficulty gradually to keep training similar across participants and to not overextend poor performers or young children. This might have led to an too easy course of play in general.
The game's moderately value of flow implies room for improvement, which in turn can positively affect learning [21].
The negative correlation between reading and spelling proficiencies and flow indicates that children with lower literacy

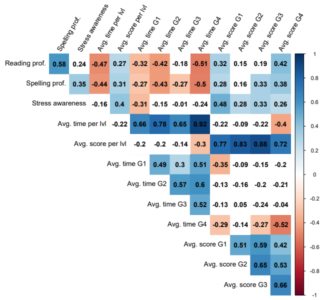


Figure 12: Detailed Spearman's Rank correlations between literacy skills and in-game scores and times per game (G). Correlations significant on $\alpha = .05$ are colored.

Acknowledgments

This research was funded by the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments. Heiko Holz is a doctoral student at the LEAD Graduate School & Research Network.

skills needed to pay more attention to the game and, thus, experienced higher levels of flow.

In addition, the perception of self-efficacy was reported by the children, which is a central aim of therapeutic interventions [2] and is related to self-awareness of and actual skill increase [7]. The positive correlations between self-efficacy and likelihood to recommend and between self-efficacy and intention to use indicate that the more effective the children perceived the game to be, the more likely would they recommend the game to friends and would play it in future. This is supported by positive correlations between the positive subscales of the iGEQ$_+$ and self-efficacy, likelihood to recommend, and intention to use.

Moreover, the results indicate that primary school children can use the game very easily, that they would likely keep on playing it, and recommend the game to friends. Generally, perceived ease of use, among other variables, was reported to be an important predictor of learning success and flow in other game-based trainings (e.g., [33]).

Taken together with the training behavior measured by the number of days and the amount of time children spent with the game, as well as that three-quarter of the children fulfilled their training plan, we infer that the game is feasible as an intervention at home and that it was able to engage children throughout the training.
Significant correlations between literacy skills and in-game data of scores and times (Figure 12) provide support for the game's pedagogical approach, i.e. difficulties of children with poor literacy skills were trained. This is in line with previous research providing evidence that in-game measures such as times (e.g., [40]) or scoring (e.g., [32]) may allow for valid assessment of skills and knowledge.
Finally, Prosodiya was overall rated to be significantly more like a game than like homework.

Conclusion and Outlook

In this article, we presented the preliminary evaluation of a randomized controlled field trial of our proposed mobile serious game for German dyslexic children regarding player experience, feasibility of the digital intervention, and first evidence of our pedagogical approach. The results of 63 children from the first intervention group, who answered a questionnaire after playing the game for 9–10 weeks, are overall very promising. The game was perceived very positively while not invoking negative feelings. The results also indicate that challenge and flow of the game can be improved. We may conclude that our game-based intervention can easily be used by primary school children, is likely considered to be a game, engages and motivates children over a longer time period, and that children are likely to recommend the game to friends and keep on playing it. Thus, we may infer that such interventions can successfully be used in field trials outside the classroom or learning therapy. Significant correlations between in-game data of times and scores and real-life literacy skills provide first evidence that our game addresses difficulties of children with poor reading and spelling skills. Importantly, these results are promising as to the validity of our serious game. Taken together, it seems that our game achieved a balance between game-play and learning objectives.

We plan to further investigate in-game measures and aim at increasing the game's flow and challenge, e.g. by continuing the development of an adaptive user model and integration of daily narratives, to keep player experience high over longer period of time. The efficacy of the game regarding reading, spelling, and syllable stress awareness will be evaluated in the near future after the second intervention group has successfully finished its training.

REFERENCES

1. Sylvester Arnab, Theodore Lim, Maira B. Carvalho, Francesco Bellotti, Sara de Freitas, Sandy Louchart, Neil Suttie, Riccardo Berta, and Alessandro De Gloria. 2015. Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology* 46, 2 (mar 2015), 391–411. DOI: http://dx.doi.org/10.1111/bjet.12113
2. Franziska Bender, Katharina Brandelik, Kerstin Jeske, Marlies Lipka, Cordula Löffler, Gerd Mannhaupt, Carl Ludwig Naumann, Marianne Nolte, Gabriele Ricken, Heinz Rosin, Gerheid Scheerer-Neumann, Michael von Aster, and Maria von Orloff. 2017. Die integrative Lerntherapie. *Lernen und Lernstörungen* 6, 2 (2017), 65–73. DOI: http://dx.doi.org/10.1024/2235-0977/a000167
3. Kay Berkling. 2017. Phontasia: A phonics game for German and its effect on orthographic skills–first corpus explorations. In *WOCCI 2017: 6th International Workshop on Child Computer Interaction*. ISCA, ISCA, 7–16. DOI: http://dx.doi.org/10.21437/WOCCI.2017-2
4. Kay Berkling, Heiko Faller, and Micha Piertzik. 2017. Avoiding Failure in Modern Game Design with Academic Content - A Recipe, an Anti-Pattern and Applications Thereof. In *Proceedings of the 9th International Conference on Computer Supported Education*. 25–36. DOI: http://dx.doi.org/10.5220/0006281800250036
5. Elizabeth A. Boyle, Thomas Hainey, Thomas M. Connolly, Grant Gray, Jeffrey Earp, Michela Ott, Theodore Lim, Manuel Ninaus, Claudia Ribeiro, and João Pereira. 2016. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education* 94, Supplement C (2016), 178–192. DOI: http://dx.doi.org/10.1016/j.compedu.2015.11.003
6. Luciana Cidrim and Francisco Madeiro. 2017. Information and Communication Technology (ICT) applied to dyslexia: literature review. *Revista CEFAC* 19, 1 (feb 2017), 99–108. DOI: http://dx.doi.org/10.1590/1982-021620171917916
7. Timothy J. Cleary, Brittany Velardi, and Bracha Schnaidman. 2017. Effects of the Self-Regulation Empowerment Program (SREP) on middle school students' strategic skills, self-efficacy, and mathematics achievement. *Journal of School Psychology* 64, Supplement C (2017), 28–42. DOI: http://dx.doi.org/10.1016/j.jsp.2017.04.004
8. Stephanie S. Daniel, Adam K. Walsh, David B. Goldston, Elizabeth M. Arnold, Beth A. Reboussin, and Frank B Wood. 2006. Suicidality, School Dropout, and Reading Problems Among Adolescents. *Journal of Learning Disabilities* 39, 6 (2006), 507–514. DOI: http://dx.doi.org/10.1177/00222194060390060301
9. Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining "Gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11)*. ACM, New York, NY, USA, 9–15. DOI: http://dx.doi.org/10.1145/2181037.2181040
10. Athanasios Drigas and Elektra Batziaka. 2016. Dyslexia and ICTs, Assessment and Early Intervention in Kindergarten. *International Journal of Emerging Technologies in Learning (iJET)* 11, 02 (feb 2016), 53. DOI:http://dx.doi.org/10.3991/ijet.v11i02.5193

11. Usha Goswami, Natasha Mead, Tim Fosker, Martina Huss, Lisa Barnes, and Victoria Leong. 2013. Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language* 69, 1 (2013), 1–17. DOI: http://dx.doi.org/10.1016/j.jml.2013.03.001

12. Martin Grund, Rainer Leonhart, and Carl Ludwig Nauman. 2017. *Diagnostischer Rechtschreibtest für 4. Klassen (DRT 4)* (3 ed.). Hogrefe Schultests. https://www.testzentrale.de/shop/diagnostischer-rechtschreibtest-fuer-4-klassen.html

13. Juho Hamari, David J. Shernoff, Elizabeth Rowe, Brianno Coller, Jodi Asbell-Clarke, and Teon Edwards. 2016. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior* 54 (jan 2016), 170–179. DOI: http://dx.doi.org/10.1016/j.chb.2015.07.045

14. Heiko Holz, Benedikt Beuttler, and Manuel Ninaus. In press 2018a. Design Rationales of a Mobile Game-Based Intervention for German Dyslexic Children. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18 Extended Abstracts)*. ACM. DOI: http://dx.doi.org/10.1145/3270316.3271523

15. Heiko Holz, Katharina Brandelik, Benedikt Beuttler, Jochen Brandelik, and Manuel Ninaus. In press 2018b. How to train your syllable stress awareness – A digital game-based approach for German dyslexic children. *International Journal of Serious Games* (In press 2018).

16. W.A. IJsselsteijn, Y.A.W. de Kort, and K. Poels. 2013. *The Game Experience Questionnaire*. Technische Universität Einhofen, Eindhoven. https://pure.tue.nl/ws/files/21666907/Game

17. Elena Ise and Gerd Schulte-Körne. 2010. Spelling deficits in dyslexia: Evaluation of an orthographic spelling training. *Annals of Dyslexia* 60, 1 (2010), 18–39. DOI: http://dx.doi.org/10.1007/s11881-010-0035-8

18. Gracia Jiménez-Fernández, Nicolás Gutiérrez-Palma, and Sylvia Defior. 2015. Impaired stress awareness in Spanish children with developmental dyslexia. *Research in Developmental Disabilities* 37 (feb 2015), 152–161. DOI: http://dx.doi.org/10.1016/j.ridd.2014.11.002

19. Monika Kast, Gian-Marco Baschera, Markus Gross, Lutz Jäncke, and Martin Meyer. 2011. Computer-based learning of spelling skills in children with and without dyslexia. *Annals of Dyslexia* 61, 2 (2011), 177–200. DOI:http://dx.doi.org/10.1007/s11881-011-0052-2

20. Fengfeng Ke and Tatiana Abras. 2013. Games for engaged learning of middle school children with special learning needs. *British Journal of Educational Technology* 44, 2 (2013), 225–242. DOI:http://dx.doi.org/10.1111/j.1467-8535.2012.01326.x

21. Kristian Kiili, Sara de Freitas, Sylvester Arnab, and Timo Lainema. 2012. The Design Principles for Flow Experience in Educational Games. *Procedia Computer Science* 15, Supplement C (2012), 78–91. DOI: http://dx.doi.org/10.1016/j.procs.2012.10.060

22. Christian Klicpera and Barbara Gasteiger Klicpera. 2000. Sind Rechtschreibschwierigkeiten Ausdruck einer phonologischen Störung? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 32, 3 (jul 2000), 134–142. DOI: http://dx.doi.org/10.1026//0049-8637.32.3.134

23. Karin Landerl. 2003. Categorization of vowel length in German poor spellers: An orthographically relevant phonological distinction. *Applied Psycholinguistics* 24, 4 (2003), 523–538. DOI: http://dx.doi.org/10.1017/S0142716403000262

24. Wolfgang Lenhard and Alexandra Lenhard. 2016. Evidenzbasierte Förderung schulischer Fertigkeiten am Computer: Lernspiele mit Elfe und Mathis. *Förderprogramme für Vor- und Grundschule* 2006 (2016), 87–113.

25. Victoria Leong, Jarmo Hämäläinen, Fruzsina Soltész, and Usha Goswami. 2011. Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language* 64, 1 (2011), 59–73. DOI: http://dx.doi.org/10.1016/j.jml.2010.09.003

26. Hans Mayringer and Heinz Wimmer. 2014. *Salzburger Lese-Screening für die Schulstufen 2-9.* Hogrefe.

27. Kristina Moll, Sarah Kunze, Nina Neuhoff, Jennifer Bruder, and Gerd Schulte-Körne. 2014. Specific Learning Disorder: Prevalence and Gender Differences. *PLoS ONE* 9, 7 (jul 2014), e103537. DOI: http://dx.doi.org/10.1371/journal.pone.0103537

28. Kristina Moll and Karin Landerl. 2009. Double Dissociation Between Reading and Spelling Deficits. *Scientific Studies of Reading* 13, 5 (2009), 359–382. DOI:http://dx.doi.org/10.1080/10888430903162878

29. Kristina Moll and Karin Landerl. 2010. *SLRT II – Lese- und Rechtschreibtest. Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT).* Verlag Hans Huber, Bern. https://www.testzentrale.de/shop/lese-und-rechtschreibtest.html

30. Rudolf Müller. 2003a. *Diagnostischer Rechtschreibtest für 2. Klassen (DRT 2)*. Beltz, Göttingen. https://www.testzentrale.de/shop/diagnostischer-rechtschreibtest-fuer-2-klassen.html

31. Rudolf Müller. 2003b. *Diagnostischer Rechtschreibtest für 3. Klassen (DRT 3)*. Beltz, Göttingen. https://www.testzentrale.de/shop/diagnostischer-rechtschreibtest-fuer-3-klassen.html

32. Manuel Ninaus, Kristian Kiili, Jake Mcmullen, and Korbinian Moeller. 2017a. Assessing fraction knowledge by a digital game. *Computers in Human Behavior* 70 (2017), 197–206. DOI: http://dx.doi.org/10.1016/j.chb.2017.01.004

33. Manuel Ninaus, Korbinian Moeller, Jake McMullen, and Kristian Kiili. 2017b. Acceptance of Game-Based Learning and Intrinsic Motivation as Predictors for Learning Success and Flow Experience. *International Journal of Serious Games* 4, 3 (sep 2017), 15–30. DOI:http://dx.doi.org/10.17083/ijsg.v4i3.176

34. Jan L. Plass, Bruce D. Homer, and Charles K. Kinzer. 2015. Foundations of Game-Based Learning. *Educational Psychologist* 50, 4 (oct 2015), 258–283. DOI: http://dx.doi.org/10.1080/00461520.2015.1122533

35. Janet C. Read. 2008. Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. *Cognition, Technology & Work* 10, 2 (apr 2008), 119–128. DOI: http://dx.doi.org/10.1007/s10111-007-0069-9
36. Frederick Reichheld. 2004. The One Number you Need to Grow. *Harvard business review* 81 (2004), 46–54.
37. Luz Rello, Clara Bayarri, Yolanda Otal, and Martin Pielot. 2014. A computer-based method to improve the spelling of children with dyslexia. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility - ASSETS '14*. ACM Press, New York, New York, USA, 153–160. DOI: http://dx.doi.org/10.1145/2661334.2661373
38. Carola Reuter-Liehr. 1993. Behandlung der Lese-Rechtschreibschwäche nach der Grundschulzeit: Anwendung und Überprüfung eines Konzeptes. *Zeitschrift für Kinder- und Jugendpsychiatrie* 21, 3 (1993), 135–147. http://cat.inist.fr/?aModele=afficheN
39. Katharina Sauter, Jürgen Heller, and Karin Landerl. 2012. Sprachrhythmus und Schriftspracherwerb. *Lernen und Lernstörungen* 1, 4 (2012), 225–239. DOI: http://dx.doi.org/10.1024/2235-0977/a000023
40. Florian Sense, Friederike Behrens, Rob R. Meijer, and Hedderik van Rijn. 2016. An Individual's Rate of Forgetting Is Stable Over Time but Differs Across Materials. *Topics in Cognitive Science* 8, 1 (jan 2016), 305–321. DOI: http://dx.doi.org/10.1111/tops.12183
41. Sven Staffeldt. 2010. *Einführung in die Phonetik, Phonologie und Graphematik des Deutschen. Ein Leitfaden für den akademischen Unterricht*. Stauffenburg, Tübingen.
42. Pieter Wouters, Christof van Nimwegen, Herre van Oostendorp, and Erik D. van der Spek. 2013. A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology* 105, 2 (2013), 249–265. DOI: http://dx.doi.org/10.1037/a0031311

COAST – Customizable Online Syllable Enhancement in Texts: A flexible framework for automatically enhancing reading materials

Heiko Holz* **Zarah Weiß*** **Oliver Brehm** **Detmar Meurers**
LEAD Graduate School & Research Network
ICALL Research Group†
University of Tübingen
{heiko.holz,zarah-leonie.weiss}@uni-tuebingen.de,
olibrehm@googlemail.com, dm@sfs.uni-tuebingen.de

Abstract

This paper presents COAST, a web-based application to easily and automatically enhance syllable structure, word stress, and spacing in texts, that was designed in close collaboration with learning therapists to ensure its practical relevance. Such syllable-enhanced texts are commonly used in learning therapy or private tuition to promote the recognition of syllables in order to improve reading and writing skills.

In a state of the art solutions for automatic syllable enhancement, we put special emphasis on syllable stress and support specific marking of the primary syllable stress in words. Core features of our tool are i) a highly customizable text enhancement and template functionality, and ii) a novel crowd-sourcing mechanism that we employ to address the issue of data sparsity in language resources. We successfully tested COAST with real-life practitioners in a series of user tests validating the concept of our framework.

1 Introduction

Reading and writing disabilities are a pressing issue for today's society – approximately 4–8 % of the German population suffer from dyslexia (Moll and Landerl, 2009; Bundesverband Legasthenie und Dyskalkulie e.V, 2014). Research on reading acquisition has shown that phonological awareness is a crucial skill for successful reading and writing acquisition (Röber-Siekmeyer, 2005). Important dimensions of phonological awareness are syllable synthesis and analysis. Syllable synthesis refers to the ability to blend syllables to a whole word, and syllable analysis to the ability of segmenting a word into its syllables. Experimental studies have shown that syllable synthesis and syllable analysis are essential components of evidence-based reading training (Galuschka and Schulte-Körne, 2016; Galuschka et al., 2014). Scheerer-Neumann (1981) have shown that specific training of segmenting words into syllables can improve reading accuracy of impaired German primary-school children significantly. Additionally, computer-based programs for primary-school children that sequentially speak and highlight syllables can facilitate the learning process of reading (Jiménez et al., 2007; Olson and Wise, 1992).

Based on these empirical findings, enhanced texts with custom spacing and syllables alternately displayed in different font colors are commonly used in teaching and learning therapy to support acquisition of reading and writing. This so-called *Silbenmethode* (syllable method) (Mildenberer Verlag, 2018) teaches children to focus on and understand syllables and their structures rather than single characters and is commonly used in Germany, which is reflected by popular reading materials , such as *ABC der Tiere* and *Leselöwe*, and by available tools that facilitate the learning process of reading and writing, such as *Celeco Druckstation* and *ABC Silbengenerator*.

While first language acquisition happens through mere exposure, learning to read and write is a learned skill and thus requires explicit instruction, similar to Second Language Acquisition (SLA). In this regard, insights form SLA research on input enhancement relate to reading and writing acquisition. The well-established *Noticing Hypothesis* (Schmidt, 1990) states that learning requires the exposure to salient linguistic constructions that may be recognized by the learner. To facilitate this recognition of relevant linguistic constructions, *Input Enhancement* (Smith, 1993) has been successfully used, in particular in terms of visual enhancement of texts (e.g. colors, font changes, capitalization, spacing), cf. (Rello and Baeza-Yates, 2017; Zorzi et al., 2012; Meurers et al., 2010).

In response to this, we developed COAST.[1]

* These authors contributed equally to this work.
† http://icall-research.de

[1] www.sfs.uni-tuebingen.de/coast/

Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications, pages 89–100
New Orleans, Louisiana, June 5, 2018.

COAST is a web-based application to easily and automatically enhance syllable structure, word stress, and spacing in texts. Its primary focus is on functionality and practicability. In terms of functionality, COAST offers a high degree of customization for text enhancement, supports management of annotation schemes, and includes syllable stress. The performance of detecting syllable stress strongly predicts dyslexia (e.g., Goswami et al. (2013); Landerl (2003)) and correlates highly with reading and writing skills (Sauter et al., 2012) and, thus, is of special importance for dyslexic children. Trainings to improve the awareness of syllable stress are being developed and evaluated (Holz et al., 2017). We extend the approach of text enhancement that are provided by state of the art tools to make syllable structures and stress more salient for German native (dyslexic) speakers using NLP resources. Enhancing the text with such additional linguistic information might boost children's ability to segment words into relevant components and might help them to learn to focus on relevant areas of words – as major orthographic challenges, such as vowel length markers, mainly occur in (conjunction with) stressed syllables (Staffeldt, 2010). To account for practicability, we implement this functionality by collaborating closely with prospective users and in particular teaching practitioners to meet real-life demands.

The remainder of the article is structured as follows: In Section 2, we report findings of a requirement analysis that we conducted in form of expert interviews prior to the system design to determine the wishes and needs of practitioners and compare COAST to two state of the art tools currently used in learning therapy and reading and writing acquisition. In Section 3, we describe the framework of COAST and explain the two core functionalities crowd-sourcing and text enhancement with real-life use cases. In Section 4, we evaluate the usability and user experience of COAST by means of user tests conducted with learning therapists and validate its practical applicability. We conclude by describing the current state of COAST and providing an outlook for its further development in Section 5.

2 Requirements Analysis

2.1 Expert Interviews

As the primary focus of our work was on the design of a tool that allowed for the immediate practical application by language teachers and learning therapists, we performed a requirement analysis for our system preceding its implementation. We conducted four expert interviews with teaching therapists to establish their wishes and requirements for a text-enhancement tool that would facilitate their work. During this process, we identified a series of concrete requests going beyond the tool's basic text analysis functionality. They were centered around four main issues: i) input/output options, ii) flexible customization settings, iii) user profiles and re-usability of settings, and iv) optional expert/user judgments.

Input/Output Options proved to be of particular interest for prospective users. They emphasized the wish to not only be allowed to upload their own texts, but also to be able to flexibly edit them while seeing the syllable enhancement. Therefore, we provide a text box for users in which they may enter and alter their texts. Regarding output options users expressed interest in being able to choose between the formats HTML, MS Word, and PDF/printing, or simply copying texts with enhancements to the clipboard. All of these were incorporated into our system.

Flexible Customization Settings were, aside from the I/O options, one of the most prominent user concerns. We found that the text representations should be customizable not only in terms of the basic text layout, but also preferably in all aspects of the actual syllable enhancement. Thus, users may freely customize the spacing of lines, words, syllables, and characters, as well as different font sizes. Furthermore, the visual syllable enhancement is customizable in terms of the colors used for stressed and unstressed syllables with the additional options to assign a separate color to secondary unstressed syllables. Colors may either be applied to the background or the font. Users may further decide to additionally highlight stressed syllables with bold font. They may also choose to mark syllable boundaries with a freely selectable delimiter. Finally, users can select certain parts-of-speech to be either i) annotated, ii) marked as unstressed, or iii) ignored. Combined, these parameters allow for a highly customizable text design and visual enhancement, that gives users a high degree of freedom regarding the representation of their texts.

User Profiles and Re-Usability became relevant concerns in the course of our expert interviews: Our flexible customization options give users the freedom to design text representations and visual enhancements that are tailored specifically to their purposes. However, users stressed the importance that they could re-use their elaborate customization across sessions, and that they need to be able to switch between various customized enhancement templates. To allow users to save, manage, and re-use their templates, we created user accounts that allow users to locally save their customization. Users may also save the texts they uploaded under a user-defined title in previous sessions.

Expert/User Judgments proved to be a final, pressing issue for prospective users: The option to adjust the automatic analyses in cases where users disagree with the syllabification or stress annotation performed by the system was crucial to our prospective users. To give them complete authority over their analyses, each of both analyses may be altered by the user on click. Furthermore, they asked us to flag words that were unknown to our system and thus more error prone. To facilitate manual corrections, we offer users to review all unknown words consecutively in a separate view, where they are supported by the syllabification and stress suggestions of our systems. All changes conducted by users are saved in their local syllabification data base and used for future analyses. Users may review and edit these new entries in their account settings. A final suggestion of our expert users was to allow the system to learn from user feedback. We thus include a crowd-sourcing based mechanism for updates to the global data base, which is explained in detail in Section 3.2.

2.2 Related Work

There are two dominant syllable enhancement tools for German whose functionality is centered around the so called *Silbenmethode* ("syllable method"), in which reading is taught by focusing on syllables and their pronunciation rather than single characters: the *Silbengenerator* ("syllable generator") and the *Celeco Druckstation* ("Celeco printing station").[2] Table 1 shows a comparison of the tools with *COAST* based on the characteristics that we identified in our expert interviews and some more general usability considerations.

System Feature	Silbengenerator	Celeco	COAST
Platform Independent	✗	✗	✓
Web-Based	✗	✗	✓
Freely Available	(✓)	✗	✓
Free Text Input	✓	✓	✓
Text Box	✗	✓	✓
Basic Text Layout Customization	✓	✓	✓
Additional Text Layout Customization	✗	(✓)	✓
Customizable Syllable Enhancement	✗	✓	✓
Configuration Templates	✗	n.a.	✓
Stress Annotation	✗	✗	✓
Syllable Arcs	✗	✓	✗
Customizable Analysis	(✓)	(✓)	✓
Crowd-Sourcing	✗	✗	✓
Exercise Generation	✓	✓	✗

Table 1: Comparison of *ABC Silbengenerator*, *Celeco Druckstation*, and *COAST*.

Silbengenerator is a Windows program published by *Mildenberger Verlag (2018)*.[3] It is part of their *ABC der Tiere* ("animal alphabet") series of learning materials based on syllabification as reading aid. Its main functionality is to allow teachers to visually enhance syllables in their reading materials. While the full version has to be purchased, a free demo is freely available for downloads on their web page.The tool allows users to upload own texts for analyses, but not to modify them from within the tool via some form of text box. Supported output formats for enhanced texts are MS Word or PDF/print. The general text layout is adjustable in terms of line spacing, fonts, font sizes, text alignment, line breaks, and background color. However, more advanced changes to the text layout, such as customized syllable, character, or word distances are not supported. The latest customized layout may be re-used upon system restart, but it is not possible to store multiple templates. Syllables are visually enhanced using the conventions of the *ABC der Tiere* materials, which hyphenates syllables and additionally marks alternating syllables with red and blue font. Monosyllabic words default to blue. Word stress is not encoded. To accommodate limited printing capacities, syllables may be enhanced using gray and black instead of red and blue, but further customization is not supported. Users may locally overwrite the syllable boundaries set by the system for individual words by editing a plain text file outside of the program. Changes are applied to all documents upon restart. Changes during run-time or for individual documents are not supported User corrections are not

[2] We are not aware of any tools for the English market that provide any syllable enhancement beyond character-based markings.

[3] For details, see: www.abc-der-tiere.de/index.php?id=388

re-used to improve the system's syllabification performance. The *Silbengenerator* also includes a limited tutoring functionality, which includes two variations of syllable reading exercises as well as capitalization, vowel, and spelling training.

Celeco Druckstation is a Windows program distributed by *Celeco* (Klische, 2007).[4] Since there is no free demo version, we base our review on the elaborate tool description provided on their web page. It should be pointed out that – unlike the *Silbengenerator* and our system – the *Celeco Druckstation* is distributed as a full fledged diagnosis and therapy tool for reading disorders for learning therapists and home tutoring alike. It thus provides a number of tests and exercises for reading and diagnosis, which are generated from texts specified by the user. This also includes a syllable enhancement facility that allows to load texts into the program, visually enhance syllables, and print them. *Celeco Druckstation* offers to adjust the basic text layout in terms of fonts, font size, font color, and background color. It also supports advanced layout modifications in terms of text segmentation: users may choose to put spaces after every syllable or every 3rd, 4th, or 5th character. Syllables are enhanced with two alternating, freely customizable colors, or with syllable arcs. No special encoding of word stress is offered. Users may provide individual syllable analyses of unknown words. These are saved in a local data base. However, the syllabification of known words can – as far as we could determine – not be altered by the user. We could not determine whether enhancement settings may be saved and re-used as templates.

3 Tool

3.1 System Description

We developed COAST as a platform-independent web-based tool that is deployed with Apache on a server hosted on the Amazon Web Services (AWS).[5] The front-end was developed with HTML, CSS, JavaScript, and AngularDart. [6] The back-end was developed with Python using the frameworks Flask,[7] and SQLAlchemy. [8] We use spaCy (Honnibal and Johnson, 2015) for natural language processing (NLP).

[4] www.celeco.de/
[5] www.aws.amazon.com/
[6] www.angulardart.org/
[7] www.flask.pocoo.org/
[8] www.sqlalchemy.org/

Target Users are on the one hand teaching practitioners, but on the other hand any person with an interest in syllabified reading material, such as tutors or parents. We account for this divide with two separate types of user accounts: regular and expert users. Currently, this distinction is relevant for our crowd-sourcing mechanism, which is discussed in Section 3.2.

Analyzing Input Texts is the core functionality of COAST. Figure 1 shows the workflow of automatic text analysis and enhancement. Before users can enhance texts in the front-end as described in Section 3.2, texts need to be processed accordingly: First, spaCy is used for parsing, tokenization, and part-of-speech (PoS) tagging. The letters of a word and its PoS are used as a combined primary key to query the global and local database stored in SQLite. The global database is initialized with the German version of the language corpus CELEX2 of Baayen et al. (1995) and is available to all users. For the approximately 360,000 lemmas and inflected word forms that are included in CELEX2, we infer primary word stress and syllable structure from CELEX2's orthographic and phonetic transcriptions.The local database consists of manually annotated entries and is only available to the specified user. If an entry was found, the syllabification, syllable stress, and lemma of the word are returned. If no entry was found, the word is marked as unknown and must be manually annotated. Manually annotated entries are automatically stored in the local database of the user and forwarded to the crowd-sourcing mechanism explained in detail in Section 3.2. The annotated information is used afterwards to enhance syllables and words of the text as can be seen in Appendix A, Figure 5. Further linguistic information for each enhanced word may be obtained individually, see Appendix A, Figure 6.

3.2 Features

Crowd-Sourcing is one of *COAST*'s most innovative features. We exploit the crowd-knowledge for long-term improvements of our automatic syllabification and word stress analysis. Currently, the crowd is derived from COAST's active users. To reliably identify not only syllable boundaries but also stress patterns is one of the biggest challenges in automatic syllable enhancement due to limitations of the available linguistic resources. This is especially true for languages other than English and

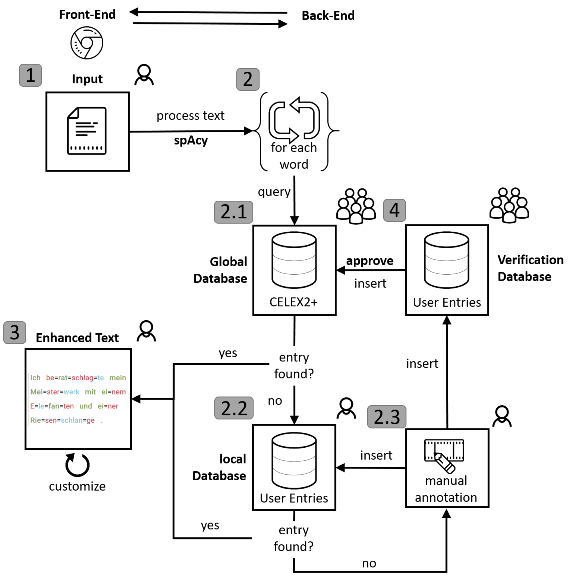


Figure 1: System overview of COAST.

for German this issue is particularly pressing, because irrespective of the size of the underlying data base, morphological composition and derivation are highly productive in German, which makes the occurrence of unknown words more likely. This issue of data sparsity is well-known from other NLP applications such as machine translation or information retrieval, and often addressed in work on compound splitting (Ziering and van der Plas, 2016; Weller et al., 2014). Furthermore, both processes may alter the word stress making the issue more difficult (Féry, 1998).

During the expert interviews it became apparent that prospective users prefer to be alerted to unknown words that may have been mis-analyzed, so they may review and if necessary manually correct them. In this context, we found that users would also prefer the system to learn from their corrections so that they could contribute to making the system more efficient in the long run. Together with our experts we therefore developed a crowd-sourcing mechanism, that would allow local corrections of users to be incorporated to our system's global data base after they have been verified by either two more users or an expert user. We derived this role of an expert user to prioritize the votes of learning therapists and linguists over layman judgments. Upon registration, new users may self-identify as experts or as regular users. A verification of this self-assignment remains for future work.Experts may also revoke crowd-induced updates to the data base. With this combined expertise and additional layer of control through experts, our system may draw from a constantly growing pool of analyses, which ultimately increases its usability and robustness while building a promising resource for future work.

The verification mechanism is located on a separate page that asks users to voluntarily identify syllable boundaries and word stress of words unknown to the system. To facilitate analyses, we provide users with information on how our automatic tools would analyze a word as help as well as with previous analyses of other users. We incorporate the freely accessible MARY-TTS (Schröder and Trouvain, 2003) for automatic suggestions for stress annotation and Pyphen for syllabification. We require users to manually annotate each word unknown to the database due to insufficient performance of automatic stress assignment. This is an extensible framework, which may be expanded with more detailed information in the future.

The following use case illustrates this process: User 1 uploads a text containing two words unknown to the system: *Hitzeschock* ("heat shock") and *Hacken* ("heels"). She is asked to determine the syllable boundaries and stress (marked in bold font) for both words and submits ***Hit**-ze-schock* and ***Hac**-ken*. This syllabification assumes a bisyllabic consonant doubling for both terms. While this is correct for most consonants at syllable boundaries, *ck* is an exception to this rule which is unknown to many laymen. Thus, when our system prompts Users 2 and 3 to verify User 1's analysis, they agree with her and both analyses are updated to the global data base. User 4 uploads another text containing *Hitzeschock*. Afterwards, the word is not flagged as unknown, but analyzed together with all other words that were originally included in the data base. Expert User 5 is asked to review the updates to the data base. She identifies the mistake that has been made and revokes the analysis of *Hacken* to ***Ha**-cken*. The entry is immediately corrected in the global data base and will be displayed correctly for all future analyses.

Text Enhancement is the core functionality of our tool. We enhance syllable boundaries as well as – unlike other systems – stress. For this, we rely on automatic analyses and manual post-hoc corrections by the user for words that are flagged as unknown: Our expert interviews clearly showed that prospective users not only prefer a high degree of customization in the visual representation

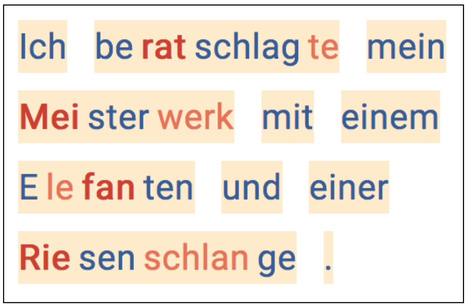


Figure 2: Template inspired by *ABC der Tiere*.

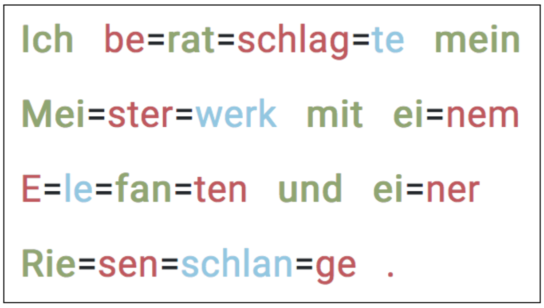


Figure 3: Template inspired by *Leselöwen*.

of their texts, but also want to re-use and switch between templates. Therefore, we not only facilitate advanced customization options for the text and enhancement layout, but also allow users to store various templates, which may be consecutively applied to a text with a simple click.

Our settings feature two main categories: First, they allow to modify the enhancement of syllable stress by allowing users to choose freely the colors assigned to i) stressed syllables, ii) unstressed syllables, and optional iii) the second unstressed syllable. Stressed syllables may be enhanced with bold font. Users can choose whether to apply the color enhancement to the font or the syllables' background. Furthermore, syllable boundaries may be made more salient by using a syllable delimiter character that users may choose freely. Finally, users may specify to which extend certain parts of speech should be analyzed, e.g. they may choose to ignore articles or to default connectives to be enhanced as unstressed. Second, they allow users to customize the text layout independent of the syllable enhancement. This includes basic options such as adjusting font size or line space. However, we also allow to freely choose the distance of words, syllables, and characters. Users may further make word boundaries more salient by choosing a background color for them. The combination of these syllable enhancement and text layout settings may be saved under a descriptive title as a template, which may be re-used and altered at any point across texts or sessions.

The following use case illustrates how this works: User 1 works with children with reading disabilities from two groups: Group A uses the *ABC der Tiere* materials in school. The children are thus used to the blue and red layout, which User 1 wants to alter as little as possible, while still providing her pupils with materials that also mark syllable stress. Therefore, she customizes a template to use the *ABC der Tiere* style for her enhancement. Figure 2 shows the result for the sentence *Ich beratschlagte mein Meisterwerk mit einem Elefanten und einer Riesenschlange* ("I consulted my masterpiece with an elephant and a giant snake").

She sets the marking color of stressed syllables to dark red and of unstressed to blue. In order to make the alternation of syllables more salient, secondary unstressed syllables are also marked in red. To clearly distinguish them from stressed syllables, she additionally uses bold font to mark stress and uses a lighter type of red to mark secondary unstressed syllables. Because *ABC der Tiere* colors monosyllabic words in blue, User 1 further sets typically monosyllabic parts of speech, such as articles and prepositions, from the analysis to be analyzed as unstressed. Finally, she makes syllable boundaries more salient by widening the distance between syllables. To make word boundaries more salient, despite this increased syllable distance, she further widens word distance and assigns a beige background color to words.

Children from Group B do not use the *ABC der Tiere* materials at school, but they are reading syllabified stories at home from the *Leselöwen* ("reading lions") materials by the *Loewe* publisher.[9] These materials use three colors to mark alternating syllables and they do not treat monosyllabic words differently from others. For this group, too, User 1 wants to make stressed syllables more salient in her materials, while otherwise not deviating much from the layout the children are already used to. Thus, she designs a second layout which mimics the *Leselöwen* style. The result of applying this template to the same sentence she used for Group A may be seen in Figure 3.

The colors used by *Leselöwen* are green, red, and blue. She assigns stressed syllables the color

[9] www.loewe-verlag.de/content-1013-1013/leseloewen/

green and again additionally marks them with bold font. Unstressed and secondary unstressed syllables are colored red and blue. Because this style already features three colors, she does not want to use a background color for words. At the same time, she wants to make word as well as syllable boundaries more salient. For this, User 1 chooses to mark syllable boundaries with a delimiter (in this case =) but without additional space between syllables and increases the distance between words. While the initial customization took a couple of minutes, User 1 may re-apply her two templates to any text in the future, reducing the time required for customization to mere seconds. She may also alter the templates at any time or add new ones when required.

4 Evaluation

We conducted user tests to evaluate COAST with both practitioners as well as with non-experts. Prior to these, we performed an internal pilot testing to identify runtime issues that are not directly related to the functionality of COAST.

Five scenarios were defined to evaluate the tool's functionality, usability, and user experience. They cover i) account creation, ii) text analysis and enhancement, iii) generation and use of annotation templates, iv) reuse of previously stored texts, and v) verification of user-generated entries ("crowd-sourcing").

In the first scenario, the users were asked to create an account with given credentials.

The second scenario consisted of four major steps: First, users were asked to log into the recently created account. Secondly, they had to switch to the *Text Analysis view* of the tool and to analyze and enhance a given text. After analyzing the text, they were told to clarify all words unknown to the system, which are flagged and shaded in red. Finally, users were asked to adjust the annotation settings based on their personal preferences.

The third scenario covered the instructed generation and use of annotation templates. Users were asked to rebuild two annotation schemes by adjusting the annotation settings and save them as new templates.

In the fourth scenario, users were asked to store the analyzed text in their account and re-analyze it by selecting the stored text in the *Account view* of the tool.

In the fifth and final scenario, users were required to verify entries added by other users that are unknown to the global database. In order to do so, they were asked to switch to the *Verification view* (see Figure 7) and approve or edit five entries.

The second, third, and fifth scenario are of special importance as they cover the core-functionality of COAST and can be seen in Figure 8.

User Tests were conducted by seven users from two groups: three experts (learning therapists) to receive subject-specific feedback and four laymen to evaluate the general usability of the tool. The three experts were women aged between 40 and 51 ($M = 45$). The laymen aged between 22 and 27 ($M = 25$) included two men and two women with non-educational professions. The user test was carried out equally for both groups. None of the participants had interacted with the system before.

We used the after-scenario questionnaire (ASQ) by Lewis (1995) for quantitative data analysis. They were answered for each scenario directly after its completion. The ASQ consists of three questions covering ease of use, time efficiency, and documentation of the tool:

1. Overall, I am satisfied with the ease of use of completing the tasks in this scenario
2. Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario
3. Overall, I am satisfied with the support and documentation when completing the tasks

We used a five-point Likert scale ranging from *strongly agree* to *strongly disagree*.

For qualitative analysis, the users were explicitly instructed to "think-aloud" (Rauterberg, 1996) while working on a scenario, thus told to accurately comment each of their actions and to express expectations, thoughts, and critics.

The user test was carried out as follows: Users were free to use their preferred browser for the user test. The default browser was Google Chrome. The user test was conducted on the users' personal laptop if possible, to recreate their home or work environment and to mimic a real-life application as close as possible. If no personal laptop was available, users were provided with one. All input devices were configured according to user preferences. After setting up the work place, users were informed and instructed about the procedure of the user test, its purpose and the think-aloud method.

After clarifying all questions, users processed all scenarios consecutively in fixed order. The user tests were concluded with an interview to get general feedback and to assess the usefulness of the tool with respect to the users' professions.

4.1 Results

The results of the second, third, and fifth scenario are explained in detail due to their relevance, results for scenario one and four can be found in the Table 2.

We normalized the options of the ASQ to range from -2 (strongly disagree) to $+2$ (strongly agree) with 0 (neither) being neutral. In the following, we report positive values (i.e. agree, strongly agree) as positive feedback, negative values (i.e. disagree, strongly disagree) as negative feedback. For a more detailed differentiation of the user feedback, please see Figure 4.

The second scenario was successfully completed by all users. The ease of use ($M = 1.57, SD = 0.53$) and documentation ($M = 1.57, SD = 0.53$) of the tool was rated 100 % positively in the ASQ, time efficiency received 86 % positive and 16 % neutral ratings ($M = 1.57, SD = 0.78$). Some users criticized the layout of the *Text Analysis view*, suggesting a more compact representation of the annotation settings.

The third scenario was completed by five users without help, two needed hints from the investigator to complete all tasks. While the first template could be rebuilt by all users, two users required help with the second template. Error source was the confusion about and between the features *Silbe hervorheben* (enhance syllables), with which either the fore- or background color of syllables could be adjusted accordingly, and *Wort Hintergrundfarbe* (word background color), with which the background color of words could be set (see Figure 5). Three users completed this scenario by trial-and-error. The ease of use of this scenario was

Question	Rating				
	-2	-1	0	1	2
			Scenario 1		
ease of use	0 %	0 %	0 %	29.0 %	71.0 %
time efficiency	0 %	0 %	0 %	0 %	100 %
documentation	0 %	0 %	0 %	29.0 %	71.0 %
			Scenario 4		
ease of use	0 %	0 %	0 %	28.6 %	71.4 %
time efficiency	0 %	0 %	0 %	28.6 %	71.4 %
documentation	0 %	0 %	0 %	28.6 %	71.4 %

Table 2: Results of the ASQ for scenario 1 and 4.

Figure 4: Results of the after-scenario questionnaire completed for scenario 2, 3, and 5.

rated 71.4 % positively and 28.57 % negatively ($M = 0.57, SD = 1.27$), the time efficiency 71% positively and 29 % negatively ($M = 0.86, SD = 1.34$), and 57.2 % positively, 14.3% neutral, and 28.6% negatively in terms of documentation and support ($M = 0.57, SD = 1.14$).

The fifth and last scenario was completed by all users successfully. Ease of use, time efficiency, and documentation of this functionality were rated 100 % positively ($M = 1.71, SD = 0.49$ for each item respectively). Users suggested to design this functionality to be more user friendly by displaying and processing multiple entries at once.

The think-aloud and concluding interviews additionally revealed general layout and design flaws of COAST's visual appearance. While this goes beyond the scope of this paper, we list problems, comments, and feature requests directly linked to the core features of our tool: i) the general navigation of the tool was not very intuitive and self-explaining, ii) some features could only be accessed with scrolling, which was not explicitly visible to the users, iii) some features, e.g. background color of words and syllable enhancement, need explicit documentation/tutorials, iv) some users asked for a simple solution to color syllables alternately independently of syllable stress, v) the feature to not enhance monosyllabic words instead of unchecking the annotation of typically monosyllabic parts of speech was requested, vi) fore- and background color of syllables and words should independently be customizable.

5 Conclusion and Outlook

COAST is a highly user-oriented, platform independent, web-based and easily extensible framework for the automatic augmentation of texts with syllable, stress, and word enhancement. It was developed in close collaboration with practitioners and includes a series of features which were explicitly requested by prospective users and that are lacking from currently available, state of the art systems. This paper presents and evaluates its ability to generate appropriate reading materials based on real-life use cases. Additionally, we evaluated the practical applicability of our tool by conducting user tests based on a series of real-life scenarios.

Our exemplary enhanced texts (see Figure 2 and 3) prove that appropriate reading materials can be easily generated automatically, customized, and exported with COAST. The use cases show that the tool meets the requirements deduced from the *a priori* requirement analysis based on our expert interviews. Compared to other tools that support syllable enhancement, COAST offers a higher degree of customization and more features, such as annotating syllable stress, setting spacing of lines, words, syllables, and characters. The automatic analysis of syllable stress and part of speech also make COAST linguistically more informed than other tools. Finally, we carried out user tests with special focus on practical application. These indicate that the majority of users were able to solve the tasks intuitively and time efficiently for each of the scenarios.

We have successfully shown that the current version of COAST allows practitioners to generate enhanced texts as reading materials for their teaching. Being able to save annotation templates and texts has proven to be an especially useful functionality to easily generate new reading materials within the application with little time effort. Furthermore, COAST features a novel crowd-sourcing approach to overcome the pressing issue of limited resources and data sparsity. This is particularly relevant for languages other than English. Currently, our tool illustrates this for the German language. However, the entire framework was designed to be easily extended for any other language for which sufficient resources are available.

Our consultation with prospective users also yielded a series of practical suggestions to optimize user experience further and to include more features. In particular, we aim at including the features discussed in Section 4.1. We also plan to redesign COAST's visual appearance. Furthermore, we intend to elaborate on the current documentation and to provide application-oriented feature tutorials. To improve the reliability of our proposed crowd-sourcing mechanism, we plan to address the verification of user roles, i.e. expert and regular users. In this regard, the need of further user type customization shall be analysed and implemented accordingly. Finally, we intend to carry out user studies to compare COAST's efficiency and efficacy to state of the art tools that support syllable enhancement in texts.

Our ultimate goal is to develop and include a front-end for learners, the *COAST App*. This results in a tutoring system offering reading and spelling exercises optimized for mobile devices. The current *COAST Tool* could be used by practitioners to generate teaching materials to be shared with the *COAST App* and, thus, to supply exercises directly to their pupils.

Acknowledgments

This research was funded by the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments. Heiko Holz is a doctoral student at the LEAD Graduate School & Research Network.

References

R. Harald Baayen, Richard Piepenbrock, and Léon Gulikers. 1995. CELEX2 LDC96L14. *Web Download. Philadelphia: Linguistic Data Consortium.*

Bundesverband Legasthenie und Dyskalkulie e.V. 2014. Der aktuelle Wissensstand über Legasthenie.

Caroline Féry. 1998. German word stress in optimality theory. *Journal of Comparative Germanic Linguistics*, 2(2):101–142.

Katharina Galuschka, Elena Ise, Kathrin Krick, and Gerd Schulte-Körne. 2014. Effectiveness of Treatment Approaches for Children and Adolescents with Reading Disabilities: A Meta-Analysis of Randomized Controlled Trials. *PLoS ONE*, 9(2):e89900.

Katharina Galuschka and Gerd Schulte-Körne. 2016. The Diagnosis and Treatment of Reading and/or

Spelling Disorders in Children and Adolescents. *Dtsch Arztebl Int.*, 113(16):279–86.

Usha Goswami, Natasha Mead, Tim Fosker, Martina Huss, Lisa Barnes, and Victoria Leong. 2013. Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language*, 69(1):1–17.

Heiko Holz, Katharina Brandelik, Jochen Brandelik, Benedikt Beuttler, Alexandra Kirsch, and Detmar Meurers. 2017. Prosodiya–a mobile game for german dyslexic children. In *International Conference on Games and Learning Alliance*, volume 10653 of *Lecture Notes in Computer Science*, pages 73–82, Cham. Springer International Publishing.

Matthew Honnibal and Mark Johnson. 2015. An Improved Non-monotonic Transition System for Dependency Parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1373–1378, Lisbon, Portugal. Association for Computational Linguistics.

Juan E. Jiménez, Isabel Hernández-Valle, Gustave Ramírez, Ma Del Rosario Ortiz, Mercedes Rodrigo, Adelina Estévez, Isabel O'Shanahan, Eduarde García, and María De La Luz Trabaue. 2007. Computer speech-based remediation for reading disabilities: The size of spelling-to-sound unit in a transparent orthography. *Spanish Journal of Psychology*, 10(1):52–67.

Anja Klische. 2007. *Leseschwächen gezielt beheben: Individuelle Diagnose und Therapie mit dem Programm celeco*. Tectum-Verl.

Karin Landerl. 2003. Categorization of vowel length in German poor spellers: An orthographically relevant phonological distinction. *Applied Psycholinguistics*, 24(4):523–538.

James R. Lewis. 1995. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7(1):57–78.

Detmar Meurers, Ramon Ziai, Luiz Amaral, Adriane Boyd, Aleksandar Dimitrov, Vanessa Metcalf, and Niels Ott. 2010. Enhancing authentic web pages for language learners. In *Proceedings of the 5th Workshop on Innovative Use of NLP for Building Educational Applications (BEA-5) at NAACL-HLT 2010*, pages 10–18, Los Angeles.

Mildenberer Verlag. 2018. *Eine Einführung in die Silbenmethode*. Mildenberer Verlag.

Mildenberger Verlag. 2018. ABC der Tiere – Silben-Generator für die Grundschule.

Kristina Moll and Karin Landerl. 2009. Double Dissociation Between Reading and Spelling Deficits. *Scientific Studies of Reading*, 13(5):359–382.

Richard K. Olson and Barbara W. Wise. 1992. Reading on the computer with orthographic and speech feedback. *Reading and Writing*, 4(2):107–144.

Matthias Rauterberg. 1996. *Usability Engineering*. Morgan Kaufmann.

Luz Rello and Ricardo Baeza-Yates. 2017. How to present more readable text for people with dyslexia. *Universal Access in the Information Society*, 16(1):29–49.

Christa Röber-Siekmeyer. 2005. Die berücksichtigung des kindlichen sprachwissens für den schrifterwerb. In Hans-Werner Huneke, editor, *Geschriebene Sprache. Strukturen, Erwerb, didaktische Modellbildungen*, pages 129–144. Mattes Verlag, Heidelberg.

Katharina Sauter, Jürgen Heller, and Karin Landerl. 2012. Sprachrhythmus und Schriftspracherwerb. *Lernen und Lernstörungen*, 1(4):225–239.

Gerheid Scheerer-Neumann. 1981. The utilization of intraword structure in poor readers: Experimental evidence and a training program. *Psychological Research*, 43(2):155–178.

R. W. Schmidt. 1990. The Role of Consciousness in Second Language Learning1. *Applied Linguistics*, 11(2):129–158.

M. Schröder and J. Trouvain. 2003. The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. *International Journal of Speech Technology*, 6:365–377.

Michael Sharwood Smith. 1993. Input Enhancement in Instructed SLA. *Studies in Second Language Acquisition*, 15(02):165.

Sven Staffeldt. 2010. *Einführung in die Phonetik, Phonologie und Graphematik des Deutschen Ein Leitfaden für den akademischen Unterricht*. Stauffenburg, Tübingen.

Marion Weller, Fabienne Cap, Stefan Müller, Sabine Schulte im Walde, and Alexander Fraser. 2014. Distinguishing degrees of compositionality in compound splitting for statistical machine translation. In *Proceedings of the 1st Workshop on Computational Approaches to Compound Analysis*, pages 81–90.

Patrick Ziering and Lonneke van der Plas. 2016. Towards Unsupervised and Language-independent Compound Splitting using Inflectional Morphological Transformations. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–653.

Marco Zorzi, Chiara Barbiero, Andrea Facoetti, Isabella Lonciari, Marco Carrozzi, Marcella Montico, Laura Bravar, Florence George, Catherine Pech-Georgel, Johannes C Ziegler, and Michael Posner. 2012. Extra-large letter spacing improves reading in dyslexia. *Proceedings of the National Academy of Sciences*, 109(28):11455–11459.

A Supplemental Material

A.1 Screenshots of COAST

Figure 5: COAST – View for text analysis and enhancement. Users can insert or edit text in the lower text box. The preview of syllable enhancement is given in the upper box. On the left side, users can edit settings regarding syllable annotation: boldness, colors of stressed and unstressed syllables (either background or foreground), background color of words, font size, and spacing between syllables, words, and lines.

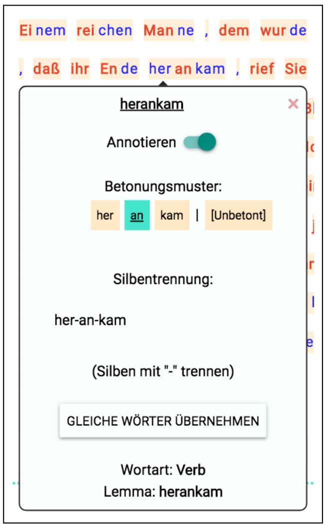


Figure 6: COAST – Word-Popup. Popup with additional information that is invoked when a known word is clicked in the text-view. At the moment, we offer information about syllabification, part of speech, and lemma. Users can additionally manually change the syllabification or stress assignment of the selected word and apply it to the preview.

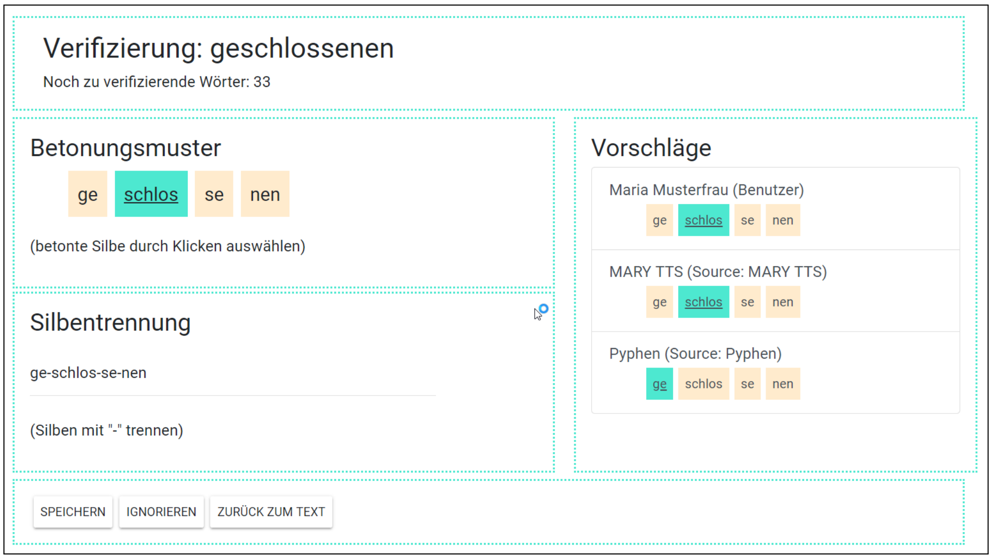


Figure 7: COAST – Verification-view of entries added by users unknown to the global database. Current word is *geschlossenen* (*ge*-***schlos***-*sen*-*en*, closed). Users can edit stress assignment and syllabification on the left side or agree to a user's judgment or to automatically generated suggestions on the right side.

Scenario 2: Text Analysis and Enhancement

1. Log in with your credentials.
2. Go to "Text Analysis".
3. Insert the given text into the text box.
4. Let the tool analyze the text.
5. Clarify all unknown words. Unknown words are shaded in red.
6. Play around with the annotation settings until the preview suits you.

Scenario 3: Annotation and Enhancement Template

1. Please try to rebuild the following annotation scheme by changing the annotation settings.

Ei = nem rei = chen Man = ne , dem wur = de sei = ne Frau krank , und als

2. Save your annotation settings as a template with the name „Template 1".
3. Now, please try to rebuild the following annotation scheme.

Einem reichen Manne , dem wurde seine Frau krank , und als

4. Save your annotation settings as a template with the name „Template 2".
5. Now, switch between "Template 1" and "Template 2" back and forth.

Scenario 5: Verification of User-Generated Entires

1. Switch to "Verification".
2. Approve or edit five entries.

Figure 8: User tests of scenario 2 (text analysis and enhancement), 3 (creation and use of annotation and enhancement templates), and 5 (verification of user-generated entries).

OPTIMIZING THE QUALITY OF SYNTHETICALLY GENERATED PSEUDOWORDS FOR THE TASK OF MINIMAL-PAIR DISTINCTION

Heiko Holz[*‡], *Maria Chinkina*[*§], *Laura Vetter*

LEAD Graduate School & Research Network
ICALL Research Group[†]
University of Tübingen

[‡] https://orcid.org/0000-0002-5211-2181, [§] https://orcid.org/0000-0002-6566-5332

ABSTRACT

Training the distinction of vowel lengths or learning to differentiate between voiced and voiceless plosive sounds in form of minimal pair differentiation is one of the treatments fostering phonological awareness for people with reading and/or writing disabilities. While text-to-speech systems can automatically generate minimal pairs (e.g., *bin* and *pin*), the quality of the pronunciation of pseudowords is not always optimal. We present a novel approach for using text-to-speech tools to artificially generate the pronunciation of German pseudowords, which is evaluated in a crowdsourcing task of the discrimination of minimal pairs. While the input for generating audio files for real words is provided as plaintext, the audio files for pseudowords are generated from the SAMPA transcription, a computer-readable phonetic alphabet, of their real-word counterparts. The task of selecting the correct word from a minimal pair of a pseudoword and its lexical counterpart was completed equally successfully when a pseudoword was generated by our method or pronounced by a human ($\chi^2(1) = 2.43, p = .119$).

Index Terms— text-to-speech, pseudoword pronunciation, minimal pair distinction, speech synthesis

1. INTRODUCTION

Phonological awareness, i.e., the ability to recognize and modify speech segments of spoken language, is regarded as a reliable predictor of literacy skills [1–4]. Therefore, trainings to improve phonological awareness can be seen as a first therapeutic intervention for reading and writing disabilities. Along with rhyme detection and syllable segmentation, training of phonological awareness includes treatments that emphasize the role of speech sounds (see [5]). One such treatment, minimal pair therapy, teaches the ability to distinguish speech sounds through the use of pairs of words that differ by a single phoneme, such as *pin* and *bin* [6]. In German, orthographic marking is required to correctly spell short and long vowels (e.g., Biene (bee, [i:]) vs. Hilfe (help, [ɪ])), thus being able to distinguish the sounds is crucial for a successful literacy development [7, 8]. For an auditory training of such minimal pairs, trained educators, such as speech and learning therapists, or audio recordings are required. This is, however, costly in terms of effort and time. For example, computer-based games for German dyslexic children, such as [9], use spoken minimal pairs for the training of vowel length distinction and differentiation. In order to reduce the effort and increase the size of the provided audio-dictionary, the audio files can be generated automatically with text-to-speech (TTS) tools.

The aim of a TTS system is to simulate the complex interplay of different elements of natural language artificially in order to produce a naturally sounding output [10]. TTS systems usually consist of two main components: a text-to-phoneme module and a phoneme-to-speech module [11, 12]. The purpose of the text-to-phoneme module is to generate a phonetic transcription of the text input [12]. The transcription is extended with prosody-related specifications that help the phoneme-to-speech module synthesize the spoken language.

Uses of TTS systems are many-fold. They play a major role in telecommunications, help people with physical limitations in everyday life. TTS systems can also be relevant for the communication between humans and machines, measurement and control systems, and between different languages. Furthermore, TTS can be used in language learning systems [11, 13] or in the therapy of phonological awareness, which has numerous benefits. For example, Tallal and Merzenich could show that the use of synthetically slowed down language material can be a predictor for a better reading and writing performance [14]. Ptok and Meisen postulate that the ability to recite minimal pairs also correlates with the reading and writing performance [15]. A form of a minimal-pair training to learn spelling patterns was realized by Berkling et al. in the language learning program Phontasia [16]. It is a

[*] These authors contributed equally to this work.
[†] http://icall-research.de

digital learning application that utilizes an Apple TTS system and works on the basis of the phonics method widely used in England. As research on Phontasia yielded positive results, the use of TTS systems for learning of spelling could be confirmed as a promising approach.

As for the evaluation of TTS tools, Handley argues that TTS systems are not sufficiently studied for use in language learning programs [12]. Indeed, a clear research focus of TTS systems to date has been placed on the intelligibility of the speech output, which is not sufficient in the context of computer-assisted language learning. Handley suggests that further investigations with regard to accuracy, naturalness, and expressiveness have to be conducted in order to be able to exploit the full potential of TTS in learning applications. In addition to that, the performance of TTS tools has only been systematically examined with real lexical words as input, and not with pseudowords.

With these limitations of the current research in mind, we conducted the current study to investigate whether a consistent modification of the input into TTS systems can optimize the speech output for lexical and pseudowords. The paper is structured as follows: First, we give an overview of the pronunciation performance of lexical words and their pseudoword counterparts of freely accessible TTS tools when using plaintext as input. We then derive the requirements of a TTS tool suitable for our experiment. Lastly, we describe our approach to artificially generate pseudoword counterparts for German lexical words and evaluate in a crowdsourcing study on discrimination of minimal pairs producing $1,228$ judgments. We show that by modifying the input of a TTS tool, the pronunciation of artificially generated pseudowords can reach a quality level comparable to their lexical counterparts.

2. OVERVIEW OF TTS TOOLS

We selected several freely available TTS tools for our analysis, without any claim to completeness, to investigate the quality of pronunciation of lexical and pseudowords. Our analyses included the following TTS tools: gspeech[1], ispeech[2], MARY[3], MWS reader[4], oddcast[5], Polly[6], and Watson TTS[7].

For the quantitative analysis, we used the aforementioned TTS tools to generate pseudoword counterparts for a subset of 233 lexical words of our database. The input was provided in plaintext. The database consisted of (non-compound)

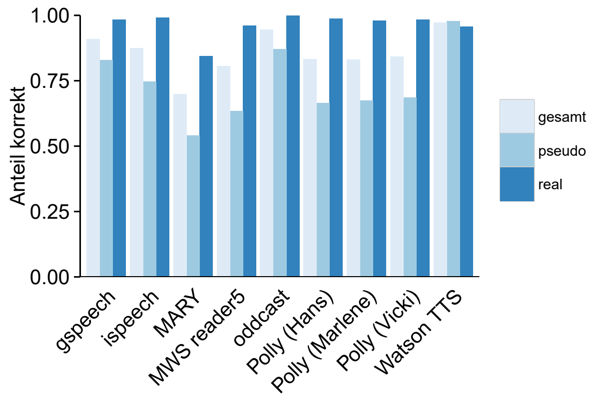


Fig. 1. Quantitative analysis of the subjectively perceived pronunciation quality of seven different freely accessible TTS tools with plaintext as input. The percentage of correctly pronounced lexical and pseudowords from the dictionary of 233 minimal pairs is plotted along the y-axis.

words that are part of the training in [9], composed of word lists for German primary-school children [17–19]. In total, the dictionary contained 90 words with a long and 106 with a short vowel in the stressed syllable and 60 words with either a voiced or a voiceless plosive as the initial sound. We had to remove words from the category 'vowel length' as no pseudoword counterparts can be generated with our algorithm described in 3.2 for words in which the vowel was succeeded by the phoneme *ch* ([X] – e.g., for the lexical word Kuchen (cake, ['Ku:-X@n]) with a long vowel in the stressed syllable, the pseudoword counterpart with a short vowel ['KU-X@n] does not have a plaintext representation). The final subset contained 79 words with a long vowel and 94 words with a short vowel in the stressed syllable and 60 words with either a voiced or a voiceless plosive as the initial sound.

We evaluated the pronunciation quality of the TTS tools with the previously described database of 233 lexical words and their pseudoword counterparts and summarized the results in Fig. 1. The results show that all freely accessible TTS tools pronounce lexical words with sufficient quality ($[84\%, 100\%]$). However, all tested TTS tools – except for Watson TTS (96%) – pronounce pseudowords that are entered in plaintext very poorly ($[54\%, 87\%]$). We assume that this is caused by the fact that TTS systems perform lookups for lexical words, and if the word is not found (as is the case for pseudowords), the synthesis for the pseudoword will be based on models trained with lexical words, thus, resulting in outputs pronounced very similar or equal to their lexical counterparts.

The evaluation reveals two drawbacks of using plaintext as input: First, Fig. 1 suggests that this results in subjectively unpredictable and insufficient pronunciation of the synthetically generated pseudoword (e.g., some TTS tools pro-

[1] `https://creative-solutions.net/wordpress/gspeech`
[2] `https://ispeech.org/text.to.speech`
[3] `http://marytts.phonetik.uni-muenchen.de:59125`
[4] `http://mwsreader.com/de`
[5] `http://oddcast.com/home/demos/tts/tts_example.php`
[6] `https://aws.amazon.com/polly`
[7] `https://ibm.com/watson/services/text-to-speech/`

Table 1. Requirement analysis of freely available TTS tools. The second and third column list the number of supported languages and voices, respectively. The other columns indicate whether the TTS tool supports a required feature.

TTS tool	# languages	# voices	API	SAMPA	SSML
gspeech	>50	?	✗	✗	✗
ispeech	29	2	✓	✗	✗
MARY	9	6	✓	✓	✓
MWS reader5	16	1	✓	✗	✗
oddcast	29	5	✓	✗	✗
Polly	17	3	✓	✓	✓
Watson TTS	10	2	✓	✗	✓

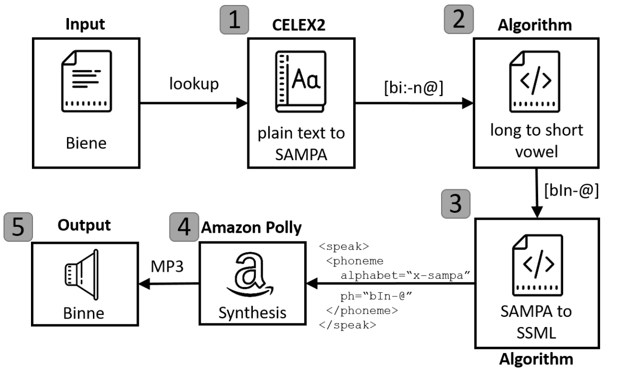


Fig. 2. Our proposed algorithm to synthetically generate pseudoword counterparts of a lexical word. As an example, for the word Biene (bee, ['bi:-n@]), a pseudoword counterpart with a short vowel is generated (Binne, ['bI-n@]).

nounced Wiehnter ['wi:-nt6], the pseudoword counterpart for Winter ['wIn-t6], too similar to its lexical counterpart). Secondly, as mentioned above, using plaintext excludes some German words as no pseudoword counterpart can be generated.

3. OUR APPROACH

3.1. Choice of a TTS Tool

To choose an appropriate TTS tool for our experiment, we derived the following features as requirements and queried the seven accessible TTS tools for their support (see Tab. 1): First, the TTS should come with various voices in order to exclude the risk of people recognizing a voice and labeling it as synthetic or human in the conducted experiment. Secondly, the TTS needed to provide a ready-to-use API to be later integrated into learning applications and systems. Thirdly, as using plaintext excludes some German words for which no pseudoword counterpart can be generated, we needed a TTS system with support of SAMPA transcription to generate properly pronounced pseudoword counterparts for all lexical words used in the experiment. Lastly, the support of SSML[8] (synthetic speech markup language) is required to adjust prosodic features, such as the speech rate of single syllables or sounds. With the help of SSML, the distinctions of short and long vowels can be emphasized more strongly.

Based on the aforementioned criteria, only Polly and MARY were considered as candidates for our experiment. Finally, we opted for Amazon Polly as its three voices consistently outperformed MARY in our quantitative analysis (see Fig. 1).

3.2. Algorithm

In this section, we describe our proposed algorithm to generate pseudoword counterparts of lexical words to form a minimal pair (see Fig. 2). We describe the algorithms for generating pseudowords from plaintext and SAMPA transcriptions.

As a first step, the SAMPA transcription of a lexical word is retrieved by querying the language corpus CELEX2 [20][9]. With the help of the SAMPA transcription, we can extract the category of the word: long vs. short vowel in the stressed syllable and/or voiced vs. voiceless plosive as initial sound. For each of this category, the SAMPA transcription is processed differently in step two.

In the second step, depending on the category extracted in step one, the corresponding counterpart is generated (short $\leftrightarrow$ long vowel; voiced [b,d,g] $\leftrightarrow$ voiceless [p,t,k] plosive). In case of plaintext as output of step two, the letters of lexical words are adjusted accordingly. For short vowels, consonant doublings are removed and letters of German orthography marking elongation are added (e.g., Antenne $\rightarrow$ Antehne). In case of a long vowel, letters marking elongation are removed and consonant doubling is added (e.g., Biene $\rightarrow$ Binne). Voiced plosives are exchanged by their unvoiced counterparts and vice versa (e.g., Pause $\rightarrow$ Bause; Biene $\rightarrow$ Piene). In case of SAMPA transcription as output of the second step, the SAMPA transcription generated in the first step is changed accordingly. For the categories 'vowel length', the symbols corresponding to the vowel in the stressed syllable are adjusted: short vowels are replaced by their long vowel counterpart and an elongation mark ([:]) is added (e.g., Antenne [an-'tE-n@] $\rightarrow$ Antehne [an-'te:-n@]). In the case of long vowels, the elongation mark is removed and the symbol for the long vowel is exchanged by its short-vowel-counterpart (e.g., Biene ['bi:-n@]$\rightarrow$ Binne ['bI-n@]). For the categories of plosives, initial voiced ([b,d,g]) and voiceless ([p,t,k]) plosives are exchanged by their counterpart. Examples generated by our algorithm can be seen in Tab. 2.

[8]https://www.w3.org/TR/speech-synthesis11/

[9]The CELEX2 contains more than 50.000 noninflected entries and is seen as the gold standard for our algorithm.

Table 2. Exemplary results of our proposed algorithm. The last three columns depict the methods used to generate the pseudoword counterpart. L2S stands for long-to-short-vowel, S2L for short-to-long-vowel and PL for voiced↔voiceless plosive.

plaintext lexical word	SAMPA lexical word	SAMPA pseudoword	L2S	S2L	PL
Ton	'to:n	'tOn	✓		
Vogel	'fo:-g@l	'fOg-@l	✓		
reservieren	re:-zEr-'vi:-r@n	re:-zEr-'vIr-@n	✓		
passieren	pas-'i:-r@n	pas-'Ir-@n	✓		
bekommen	b@-'kOm-@n	b@-'ko:-m@n		✓	
endlich	'Ent-lIx	'e:n-tlIx		✓	
Biene	'bi:-n@	'pi:-n@			✓
tanzen	'tan-ts@n	'dan-ts@n		✓	

In the third step, the SSML code based on the output of step two for the pseudoword is generated. To improve the output for the categories of 'vowel length', the speech rate was adjusted automatically for each individual (pseudo-) word. In case of a long vowel, the SSML tag 'x-slow' was used. No SSML tag was used for short vowels. The fourth step feeds the SSML code into the TTS engine Polly of the Amazon Web Services[10] (AWS). The audio generated in step four is exported and saved as a MP3-file in the fifth and final step. The quality of the minimum pairs synthesized using the Amazon Polly TTS system was tested in the experiment described below.

4. EVALUATION OF THE APPROACH

In order to be able to use synthesized output in speech therapy, it has to be as close as possible to the language produced by humans. For longer utterances, this usually means ensuring the optimal use of intonation and speech pauses in the output. For single words, naturalness and clarity are the two factors that determine the quality of the synthesized language.

To evaluate our TTS approach described in the previous section, we designed a crowdsourcing study[11]. Human- and synthetically produced audio files were presented to the participants in the form of a minimal-pair distinction task. The participants were asked to first select the correct pronunciation and then rate each audio file on the five-point scales of naturalness and clarity.

4.1. Materials

4.1.1. Audio Files

We used the algorithm described in Sec. 3.2 to artificially generate pseudoword counterparts based on the SAMPA transcription of lexical words to form a minimal pair. Amazon Polly offers three different voices for the audio output, one of which was randomly chosen for each minimal pair. The human production of both lexical words and pseudowords was recorded by three linguists. It is important to note that we did not include lexical words whose pseudoword counterpart in the minimal pair was identical to any other existing lexical word (e.g., we excluded lexical words such as schief (crooked) and Schiff (ship) which could serve as pseudoword counterpart for each other).

For the crowdsourcing experiment, the parameters *existence* (pseudoword or lexical word), *source* (HC = human-created and TTS = generated by Text-To-Speech), and *category* (S2L = short-to-long vowel, L2S = long-to-short vowel, plosive) were varied in each minimal pair. Table 3 shows the types of resulting items. In total, 282 minimal pairs were tested in the experiment.

4.1.2. Questionnaire

To find out whether the speech quality of artificial and human generated lexical words and their pseudoword counterparts are comparable, we asked one question about a minimal pair and three questions about each of the two audio files in the minimal pair.

To investigate if participants were able to identify the lexical word of a minimal pair, we asked which of the two audio files was pronounced correctly.

To investigate how natural and clear the audio files were pronounced, we designed five-point Likert-items for each audio file of the minimal pair, with options ranging from *very unnatural* to *very natural* and *very unclear* to *very clear*, respectively.

To investigate the perceived source attribution, we let the participants judge how the version of the word was generated, with *artificially generated, spoken by a human*, or *don't know* as options.

4.1.3. Items

An item consisted of the two audio files of a minimal pair and the corresponding set of questions. Items were answered in the following way: First, the participants were asked to listen to both audio files of the minimal pair and to answer the first question. Once they have selected the audio file they thought would refer to a lexical word, they could not change their answer and were forwarded to questions covering naturalness, clarity, and source attribution of each of the two audio files. We randomized the order of presentation of human-created or synthesized audio files and the categories 'plosive' or 'vowel length'.

We selected 42 items whose correct pronunciation could be easily recognized by native German speakers to serve as test items. Test items were used to ensure participants' suit-

[10]https://us-east-2.console.aws.amazon.com/polly/home/SynthesizeSpeech

[11]For this study, we used Figure Eight platform, formerly known as CrowdFlower: `https://www.figure-eight.com/`

Table 3. Total number of items used in the study grouped by *source* (HC = human-created, TTS = text-to-speech) and *category* (S2L = short-to-long vowel, L2S = long-to-short vowel, plosive).

	S2L	L2S	plosive	Total
HC	36	32	0	68
TTS	78	81	55	214
Total	114	113	55	282

Table 4. Percentages of correctly resolved minimal pairs. The results are grouped by *source* (HC = human-created, TTS = text-to-speech) and *category* (S2L = short-to-long vowel, L2S = long-to-short vowel, plosive).

	S2L	L2S	plosive	Total
HC	96%	89 %	0%	93 %
TTS	85%	93 %	82%	87 %
Total	88%	92%	82%	88 %

ability for the study, i.e. participants were expected to recognize the correct pronunciation of the word.

4.2. Participants

A total of $n = 37$ so-called "workers" from German-speaking countries, i.e. Germany, Austria, and Switzerland, participated in the experiment through the crowdsourcing platform Figure Eight. The setting to select contributors was set to "higher quality", described as a smaller group of more experienced, higher accuracy contributors. Demographic data, such as age or gender, is not accessible through Figure Eight. Each participant worked on 41.2 items on average.

4.3. Procedure

After agreement, the instructions were shown to the participants. First, we explicitly asked the crowd workers to only participate in the study if they were proficient speakers of German. The instructions, the task, and the test items were in German. Then, the participants were told that the task is to rate the speech quality of audio files. Therefore, they were instructed to use headphones, to work on the experiment in a quiet environment, and to check that their sound system works properly. We explicitly told the participants that we were not testing their hearing but our system. They were then instructed how to answer an item that consisted of two audio files and the set of questions as explained in Sect. 4.1.3.

Before proceeding to the main task, the participants had to take a so-called quiz consisting of five test items. If participants incorrectly answered more than one test item regarding the correct pronunciation of the word, they were excluded from the task.

The main task started after a successful completion of the quiz. We displayed five items per page. Participants needed to answer all five items in order to get paid and to proceed to the next page. Thirty-eight additional test questions randomly inserted to each page throughout the main task ensured the participants' suitability for the study. Participants were excluded from the main task if their recognition rate of test items dropped below 80% and were paid for the number of items already completed in the main task.

We collected five judgments from different participants for each item. The participants received five cents for each completed item.

4.4. Results

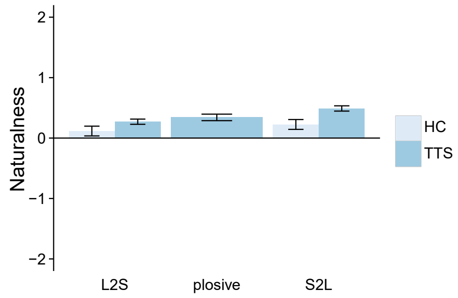
Which word was pronounced correctly?
The question of which version is the correctly pronounced one (and thus, which word is a real lexical one) was answered correctly in 88% of cases. In the subset of artificially synthesized minimal pairs, the detection performance of the correctly pronounced word was 87%, and it was 93% for human-produced audio files.

The difference in the detection of the correct version between the human- and synthetically produced items was non-significant ($\chi^2(1) = 2.43, p = .119$). The results broken down into different categories (S2L, L2S, plosive) are presented in Tab. 4. The percentages represent the number of times that the participants could tell a real lexical word from its pseudoword counterpart and recognized it as the correct pronunciation. There was a significant difference in the detection of the correct pronunciation between the three categories only for synthesized items ($\chi^2(2) = 17.98, p < .001$). The detection rate of the correct pronunciation generated by TTS for L2S was significantly higher than that for S2L ($\chi^2(1) = 9.08, p = .002$). The participants performed the lowest when detecting the correct pronunciation of synthesized items in the plosive category, although the difference with the next category was non-significant ($\chi^2(1) = 1.21, p = .271$).

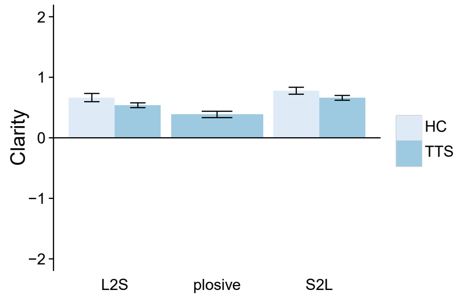
Was the word produced by a human or a computer?
In 4% of cases, the participants chose the option 'don't know' when answering this question, and in 66% of cases, they mistakenly marked artificially generated audio files as having been produced by humans. At the same time, in only 27% of cases, human-recorded items were mistakenly thought to be artificially generated. The assumption that an audio file has been spoken by a real human – although it was artificially generated – has been selected significantly more often than

vice versa ($\chi^2(1) = 9.37, p = .002$).[12]

How natural and clear did the word sound?
The synthetically generated words were judged to be significantly more natural than the words recorded by humans ($W = 680670, p = .013$). On the other hand, the latter were perceived as significantly clearer than the synthesized words ($W = 791030, p < .001$). Significant effects were also found for the two categories L2S and S2L: The short-to-long vowel transformation from a real word to a pseudoword resulted in more naturally-sounding items ($W = 602040, p < .001$). There also was a significant difference between S2L and L2S with respect to clarity ($W = 620440, p = .021$), S2L transformed words were reported to be pronounced more clearly. Finally, pseudo- and real words did not differ significantly in their perceived naturalness ($W = 1037300, p = .927$) or clarity ($W = 1029400, p = .780$). Figures 3(a) and 3(b) provide a visual overview of the naturalness and clarity ratings, respectively.



(a) Naturalness (SE)



(b) Clarity (SE)

Fig. 3. Naturalness and clarity ratings of human-created (HC) and synthesized (TTS) items grouped into the categories of long-to-short vowel (L2S), plosive, and short-to-long vowel (S2L). Bars represent standard error.

4.5. Summary of the Findings

We conclude this section with an overview of the results from our crowdsourcing experiment. The task of selecting the correct word from a minimal pair of a pseudoword and a real word was completed equally successfully when a pseudoword was generated by our method or pronounced by a human. In this task, it was the hardest for the participants to differentiate between voiced and voiceless plosives ([b, d, g] and [p, t, k], respectively) in synthesized words. The participants performed worse when presented with synthetically generated minimal pairs containing a real word with a short vowel and a pseudoword with a long (or rather, elongated) one. Interestingly, this was not the case for the words pronounced by humans.

When asked about the naturalness and clarity of recordings, participants perceived the words generated by TTS tools as significantly more natural and the human-produced words as significantly clearer. Finally, when asked to guess whether the presented word was produced by a human or a computer, participants tended to identify synthesized items as human recordings. This indicates that the quality of automatically generated files was perceived as comparable to that of human-recorded ones.

5. CONCLUSION

According to the findings of this work, replacing natural language with TTS tools in German language learning systems seems promising. This way, the dictionary of items can be extended as required, thus saving costs and time. The low naturalness ratings and the high clarity ratings for words produced by humans could be due to the fact that these were recorded for the target group while putting special emphasis on clarity, distinction, and articulation. However, the extent to which the naturalness and clarity of the pronunciation of words actually play a role in literacy learning still needs to be evaluated with the target group, namely dyslexic children.

As a next step, we aim at synthetically generating more distinct long and short vowels, e.g., by using more SSML features, not only on the word but also on the syllable level, or splitting up syllables to artificially elongate vowels. Finally, replicating the experiment with German (dyslexic) primary-school children could shed more light on the suitability of synthesized speech in the language learning context.

[12]The participants who chose the option 'don't know' (4%) were excluded from this analysis.

6. ACKNOWLEDGMENTS

This research was funded by the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments. Heiko Holz and Maria Chinkina are doctoral students at the LEAD Graduate School & Research Network.

7. REFERENCES

[1] Linnea C Ehri, Simone R Nunes, Dale M Willows, Barbara Valeska Schuster, Zohreh Yaghoub-Zadeh, and Timothy Shanahan, "Phonemic awareness instruction helps children learn to read: Evidence from the national reading panel's meta-analysis," *Reading research quarterly*, vol. 36, no. 3, pp. 250–287, 2001.

[2] Usha Goswami, Natasha Mead, Tim Fosker, Martina Huss, Lisa Barnes, and Victoria Leong, "Impaired perception of syllable stress in children with dyslexia: A longitudinal study," *Journal of Memory and Language*, vol. 69, no. 1, pp. 1–17, 2013.

[3] Gracia Jiménez-Fernández, Nicolás Gutiérrez-Palma, and Sylvia Defior, "Impaired stress awareness in Spanish children with developmental dyslexia," *Research in Developmental Disabilities*, vol. 37, pp. 152–161, feb 2015.

[4] Katharina Sauter, Jürgen Heller, and Karin Landerl, "Sprachrhythmus und Schriftspracherwerb," *Lernen und Lernstörungen*, vol. 1, no. 4, pp. 225–239, 2012.

[5] Katharina Galuschka and Gerd Schulte-Körne, "The Diagnosis and Treatment of Reading and/or Spelling Disorders in Children and Adolescents.," *Dtsch Arztebl Int*, vol. 113, no. 16, pp. 279–86, 2016.

[6] Jessica A Barlow and Judith A Gierut, "Minimal pair approaches to phonological remediation," in *Seminars in speech and language*. Copyright© 2002 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel.:+ 1 (212) 584-4662, 2002, vol. 23, pp. 057–068.

[7] K Landerl, "Categorization of vowel length in German poor spellers: An orthographically relevant phonological distinction," *Applied Psycholinguistics*, vol. 24, no. 4, pp. 523–538, 2003.

[8] Christian Klicpera, Barbara Gasteiger-Klicpera, and A Schabmann, "Lesen und Schreiben–Entwicklung und Schwierigkeiten," *Bern: Huber*, 1993.

[9] Heiko Holz, Katharina Brandelik, Jochen Brandelik, Alexandra Kirsch, Jürgen Heller, and Detmar Meurers, "Prosodiya – A Mobile Game for German Dyslexic Children," in *Games and Learning Alliance: 6th International Conference, GALA 2017, Lisbon, Portugal, December 5–7, 2017, Proceedings*, João Dias, Pedro A. Santos, and Remco C. Veltkamp, Eds., pp. 73–82. 2017.

[10] Thierry Dutoit, *An introduction to text-to-speech synthesis*, vol. 3, Springer Science & Business Media, 1997.

[11] Ruslan Mitkov, *The Oxford Handbook of Computational Linguistics*, Oxford University Press, 2005.

[12] Zöe Handley, "Is text-to-speech synthesis ready for use in computer-assisted language learning?," *Speech Communication*, vol. 51, no. 10, pp. 906–919, 2009.

[13] Kay Berkling, Nadine Pflaumer, and Rémi Lavalley, "German Phonics Game using Speech Synthesis - A Longitudinal Study about the Effect on Orthography Skills," in *Workshop on Spoken Language Technology for Education (SLaTE)*, 2015, pp. 167–172.

[14] Paula Tallal, Steve L Miller, Gail Bedi, Gary Byma, Xiaoqin Wang, Srikantan S Nagarajan, Christoph Schreiner, William M Jenkins, Michael M Merzenich, et al., "Language comprehension in language-learning impaired children improved with acoustically modified speech," *Science-AAAS-Weekly Paper Edition*, vol. 271, no. 5245, pp. 81–83, 1996.

[15] Meisen R. Ptok M., "Die Fähigkeit, Minimalpaare nachzusprechen, korreliert mit der Rechtschreibleistung," *HNO*, vol. 56, pp. 73–80, 2008.

[16] Kay Berkling, "Phontasia: A phonics game for German and its effect on orthographic skills–first corpus explorations," in *WOCCI 2017: 6th International Workshop on Child Computer Interaction*, ISCA, nov 2017, number November, pp. 7–16, ISCA.

[17] GUT1., "Grundwortschatz 500," .

[18] Gerd Schulte-Körne and Frank Mathwig, "Das Marburger Rechtschreibtraining," *Ein regelgeleitetes Förderprogramm für rechtschreibschwache Kinder. Bochum: Winkler*, 2001.

[19] Lisa Dummer-Smoch and Renate Hackethal, *Kieler Leseaufbau*, Veris Kiel, 2011.

[20] R Harald Baayen, Richard Piepenbrock, and Leon Gulikers, "The celex lexical database (release 2)," *Distributed by the Linguistic Data Consortium, University of Pennsylvania*, 1995.

Interaction Styles in Context: Comparing Drag-and-Drop, Point-and-Touch, and Touch in a Mobile Spelling Game

Heiko Holz[a,b,*], Detmar Meurers[a,b]

Abstract
Which interaction styles to use for children is controversial. In this paper, we argue for the need to choose the interaction style in context, considering a range of factors. We compare drag-and-drop, point-and-touch, and simple touch for selecting letters to form words in a spelling line as part of an educational spelling game. We evaluate the perceived workload, user experience, preference, and writing times of twenty-five children (8–11 years). We found that touch received better ratings and was ranked highest most often on all subscales compared to drag-and-drop and point-and-touch. Children needed less time using touch and 68% chose it as their favorite interaction style. We also found small advantages for drag-and-drop over point-and-touch, which runs counter to some recent recommendations. This becomes particularly clear when using ranking responses, which support a particularly fine-grained picture.

Keywords
child computer interaction; touch interaction styles; drag-and-drop; point-and-click

[a] *ICALL-Research.com Group, Department of Linguistics, University of Tübingen*
[b] *LEAD Graduate School & Research Network, University of Tübingen*
*__Corresponding author__: heiko.holz@uni-tuebingen.de

1. Introduction

When developing educational applications for children, one of the central design decisions is how actions are performed to complete a task. For example, in a simple educational spelling game in which children write single words by selecting letter objects, the question arises which interaction style should be implemented to perform the action of moving and arranging the letters. The term "interaction style" is commonly used to refer to the different ways users can interact with an application or a game (Soegaard, 2015). For the given example, the question is whether to implement a drag-and-drop, a point-and-click, or a click interaction style to arrange the letters in the spelling line. The design decision impacts performance (Donker & Reitsma, 2007a; Girard & Johnson, 2009; Hamza & Salivia, 2015; Inkpen, 2001; Joiner, Messer, Light, & Littleton, 1998; Ward, 2014), user experience (e.g., Barendregt & Bekker, 2011; Girard & Johnson, 2009; Inkpen, 2001), and educational effectiveness (Schwartz & Plass, 2014). For young learners, it is especially challenging to implement appropriate interaction styles as it not only depends on their physical and mental development (e.g., Hamza & Salivia, 2015; Vatavu, Cramariuc, & Schipor, 2015), but also on their previous experience with the use of technology (Barendregt, 2015).

When consulting current research and guidelines on interaction styles, the conflicting results and recommendations leave the designer puzzled (Hourcade, 2008). Often point-and-click is said to be more appropriate than drag-and-drop (Chiasson & Gutwin, 2005; Gelderblom & Kotzé, 2009; Girard & Johnson, 2009; Inkpen, 2001; Joiner et al., 1998; Roman, 2015; Soni, Aloba, Morga, Wisniewski, & Anthony, 2019; Ward, 2014) – other times the opposite (Barendregt, 2015; Barendregt & Bekker, 2011; Donker & Reitsma, 2007a; Hamza & Salivia, 2015). As we argue in the next section, the conclusions seem to be dependent on the input modality (mouse vs. touch), the type of task, the age of the children, and the typical interaction experience of children with technology at the time the study was conducted.

In this article, we focus on spelling games designed to enhance dyslexic children's experiences in spelling acquisition. The acquisition and use of orthographic regularities (Galuschka, Ise, Krick,

& Schulte-Körne, 2014) implemented in trainings based on spelling rules can effectively improve children's spelling performance (Galuschka & Schulte-Körne, 2016). This is also true for computer-based spelling trainings for German dyslexic children (e.g., Baschera, 2011; Berkling, 2017; Kargl, Purgstaller, Weiss, & Fink, 2008; Kast, Baschera, Gross, Jäncke, & Meyer, 2011; Klatte, Bergström, Steinbrink, Konerding, & Lachmann, 2018) – and spelling games for mobile touch devices are reportedly more engaging than paper-pencil exercises are (e.g., Rello, Bayarri, & Gòrriz, 2013; Rello, Bayarri, Otal, & Pielot, 2014). The purpose of this article is to determine the best interaction style for a mobile touch-based spelling game for children. We compare drag-and-drop, point-and-touch, and touch with regard to subjectively perceived workload, user experience, and writing times in a lab experiment with twenty-five German children aged 8–11. We first discuss the current state of research on interaction styles for children and mobile spelling games, then introduce the characteristics of our educational learning game, before presenting the results of our lab experiment. Concluding, we compare our results with the current state of research and suggest some future directions.

2. Related Work

2.1 Interaction Styles and Children

We start by summarizing the empirical findings on interaction styles for young children and the published design guidelines based on those findings. Joiner et al. (1998) compared point-and-click and drag-and-drop using a computer mouse with children aged 7–12. In their first study, 7-year-old children were found to be faster and more accurate with pointing compared to dragging. The second study compared the performance of children from three age groups (5–6, 8–9, and 11–12 years). Older children were found to be faster and made fewer errors than younger children, independent of the interaction style. The youngest children were slower and made more errors with dragging than with pointing, while no such performance differences arose for the other age groups. Therefore, they concluded that point-and-click is the most appropriate interaction style for young children.

Inkpen (2001) found in a study with 68 children aged 9–13 that point-and-click was faster, led to fewer errors, and was preferred to drag-and-drop. The examined task was a puzzle-solving game using a computer mouse. With point-and-click, objects were moved by clicking on them, then another click on the point to where they should move, and a final click to make the move. With drag-and-drop, the mouse had to be kept down on the object while dragging it to the desired position. One of the reasons why children preferred point-and-click was that their fingers became tired of holding the mouse button down. Again, point-and-click was concluded to be more appropriate. In line with these findings, in their design principles for children's technology, Chiasson and Gutwin (2005) recommend that drag-and-drop should be replaced with point-and-click.

Donker and Reitsma (2007b) found the opposite results when comparing drag-and-drop and point-and-click in a "moving objects" task with 107 Dutch children aged 6–7 years. Children were asked to move one letter that was falsely written in a word into a trash bin. Drag-and-drop was found to be faster and to cause fewer interaction errors than point-and-click. They conclude that drag-and-drop is the most appropriate for educational software. They discuss whether error-handling differences may have influenced previous results favoring point-and-click. E.g., when children made an error in drag-and-drop in the study by Inkpen, Booth, and Klawe (1996), they had to perform the entire trial again, whereas for point-and-click they only had to redo the incorrect click. Children thus may perform drag-and-drop more accurately and slowly to avoid time-consuming rectifications (Donker & Reitsma, 2007b).

Girard and Johnson (2009) compared drag-and-drop with point-and-click in multiplication tasks with children aged 7–9 using a computer mouse. Children computed the result by first selecting the column of the digit (units, tens, hundreds) and then selecting numbers by point-and-click or using sliders with drag-and-drop. They found that point-and-click was more effective in terms of achievement, interaction error, speed and accuracy of answer, and was preferred over drag-and-drop. Gelderblom and Kotzé (2009) concur in their lessons learned in the design of technology for children aged 6–8.

Barendregt and Bekker (2011) and Barendregt (2015) performed two studies to investigate children's spontaneous use of drag-and-drop and point-and-click interaction styles in two educational math games on a computer using a computer mouse and internal touch pads. In the first game, focusing on division, children had to draw lines to equally divide candy among four babies. In the second game, children had to position numbers on a number line. In the main experiment of the first study (Barendregt & Bekker, 2011), in which both games only supported point-and-click, they tested 12 Dutch children aged 7–12 years and found that many children tried to apply drag-and-drop as a first spontaneous reaction and struggled with the point-and-click interaction style. In the second study (Barendregt, 2015), the games supported both interaction styles. Twenty-six Dutch children aged 4–6 years were asked to play without previously being informed about interaction styles. When the children played a game without watching a demo showing an interaction style, children spontaneously used drag-and-drop instead of point-and-click for both games, independent of how often they had used the different interaction styles before. The spontaneous use of drag-and-drop was especially clear for actions with a natural mapping to keep the mouse button pressed, such as cutting or drawing lines, as in the cutting game. Point-and-click was more easily adopted in the moving game. The authors assume that the children developed habits from their experiences with drag-and-drop-like functionality in desktop and mobile phone interfaces. The results suggest that drag-and-drop may be appropriate for use in educational applications, even for very young children, and that it is important to not only consider performance measures but also take habits and spontaneous use of interaction styles into account.

Ward (2014) reported on the HCI requirements of a Computer Assisted Language Learning tool for 4–5 year old Irish primary school students. The tool used keyboard and mouse input and included an exercise in which jumbled letters had to be rearranged to spell the word correctly. Two interaction styles were considered: selecting a letter and dragging it to the spelling line, or placing any letter on the spelling line that is clicked on. Both options were tested. Drag-and-drop was reported as problematic, requiring more dexterity for the movement, so the click option was adopted in the final system.

Roman (2015) conducted a usability study with 4–6 year-olds for an online streaming service. Children either used a mouse or a laptop touchpad. She concludes that clicking is favored over drag-and-drop. The same year, Hamza and Salivia (2015) investigated the performance of, among others, point-and-touch and drag-and-drop of 4–5-year-olds with an iPad application. They report that the children tried to perform drag-and-drop in the point-and-touch tasks as it apparently was more familiar, easier, and less confusing to them. They also observed that the children had better motor control in the tasks using drag-and-drop. They showed an effect of age, with 5-year-olds outperforming the younger ones. In other research (Azah, Syuhada, Batmaz, Stone, & Wai, 2014; Soliman & Nathan-Roberts, 2018), toddlers and infants were shown to be capable of using various gestures with mobile touch devices, with the drag-and-drop gesture being acquired at around the age of 3. Interestingly, the touchscreen interaction design recommendations for children by Soni et al. (2019) still recommend avoiding drag-and-drop.

In sum, the varied findings summarized here support no simple conclusion as to the most appropriate interaction style for educational applications for children. Instead, we argue that it is necessary to consider the interaction modality (touch vs. computer mouse), the particular implementation of the interaction styles, the context and type of task, and the ages and interaction experiences of the participating children. While all of the discussed interaction styles are used in published spelling games, we did not find any research that systematically compares different interaction styles in spelling games for children in which single or multiple words are written.

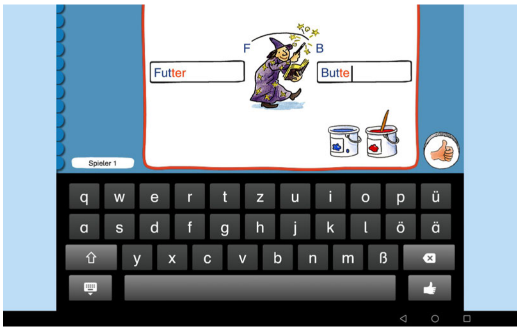
2.2 Spelling Exercises for Mobile Touch Devices

Spelling exercises available for mobile touch devices differ, among other features, how letters are arranged to write words and in the interaction styles used to select or move letters. As foundation for a concise overview, we downloaded exemplary apps for Android devices from the Google Play Store that contained spelling games for German children. Four exemplary apps are shown in Figure 1.

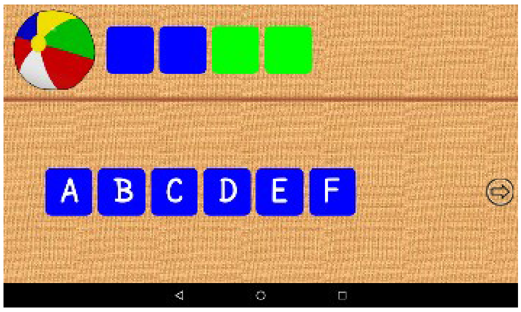
2.2.1 Arranging Letters to Write Words

Writing words using a complete keyboard: Words are written using the the device's default or a custom keyboard, see Figure 1a. Touching a letter inserts it into or appends it to the end of the word, depending on the cursor's position.

Writing words using a restricted keyboard:

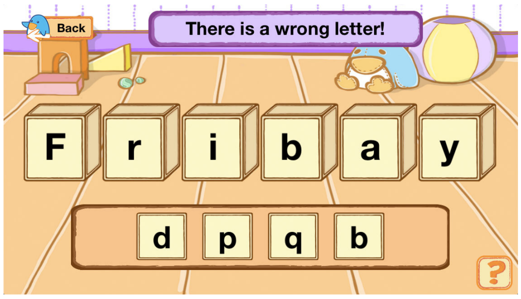


(a) *ABC der Tiere 1* (Animal Alphabet 1; Mildenberger Verlag GmbH, 2016). A complete keyboard is used for writing words. Target word in the right gap is *Butter* [butter].



(b) *ABC Deutsch Lernen Grundschule* (ABC – Learning German in Primary School; Sonnenwald Apps, 2013). Only parts of a keyboard are visible at once, other letters become available by swiping or tapping on arrows. Target word is *Ball* [ball].

(c) *Kinder lernen deutsche Wörter - lesen und schreiben* (Children Learn German Words – Reading and Spelling; Brainy Ape Studio, LLP, 2017). Only pre-defined letters are available to write words. Target word is *Mais* [corn].



(d) *Dyseggxia* (Rauschenberger et al., 2015; Rello et al., 2014). A misspelled word needs to be corrected using insertion, deletion, or replacing of letters. Target word is *Friday*.

Figure 1. Exemplary mobile spelling games.

Words are written using a partially displayed keyboard. The letter colors in Figure 1b refer to different keyboard parts. Other parts become visible by swiping left/right, or by pressing arrows keys.

Writing words using a predefined set of letters: Words are written using a set of letters, see Figure 1c. Optionally, distracting letters are included to increase task difficulty. Depending on the app, a letter can be used once or multiple times. In the former case, the number of available letters equals the word length (plus optionally distracting letters).

Correcting or writing words with limited interactions: An unfinished or misspelled word is presented and needs to be completed or corrected by inserting, deleting, or replacing letters, see Figure 1d. The focus is on major spelling challenges derived by typical dyslexic errors.

2.2.2 Interaction Styles in Mobile Touch-Based Spelling Games

Touch-based spelling games (apart from those with a complete keyboard) use three different interaction styles to select or move letters in order to write a word: drag-and-drop (DnD), point-and-touch (PaT), and touch (T). As we only refer to touch devices in this article, we renamed "click" to "touch".

Drag-and-drop works as follows to write words:

1. The child touches and holds the letter that he wants to insert into, remove from, or move within the word.
2. The child moves the letter by dragging his finger without releasing its touch.

a) *Writing letters*

3. The child moves the letter to the spelling line.
4. The child drops the letter by releasing its touch.
5. The letter is placed at the current position.

b) *Deleting letters*
   3. The child moves the letter out of the spelling line.
   4. The child drops the letter by releasing its touch.
   5. The letter is deleted.

Point-and-touch works as follows to write words:

a) *Writing letters*
   1. The child touches the letter that it wants to insert into or move within the word.
   2. The touched letter is highlighted as "currently selected".
   3. The child moves its finger to the position the letter should be placed.
   4. The child touches the desired position.
   5. The letter is placed at the current position.

b) *Deleting letters*
   1. To delete an already placed letter, the child touches the letter twice – either as a double tap or consecutively without restriction on the delay of the touches.

Touch works as follows to write words:

a) *Writing letters*
   1. The child touches the letter.
   2. The letter is immediately inserted into the word at the current position (either end of word or first free gap within the word).

b) *Deleting letters*
   1. The child touches the letter in the spelling line that it wants to delete.
   2. The touched letter is deleted.

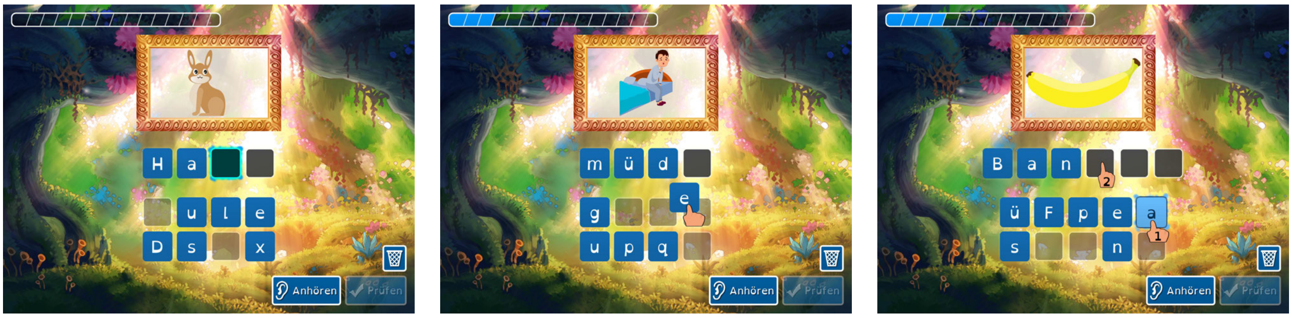
3. Contextualizing the Interaction Style

We developed a touch-based spelling game within the scope of an educational game for spelling acquisition for German primary school children aged 7–12 years (cf. Holz, Brandelik, Beuttler, Brandelik, & Ninaus, 2018; Holz, Ninaus, Meurers, & Kirsch, 2018). In the spelling game, children write single words by arranging letters from a limited set of letters in the so-called letter area in the correct order into the slots of the spelling line, see Figure 2. To be able to compare the three above-mentioned interaction styles, we implemented each interaction style for this spelling game.

The game works similar to cloze spelling tests. To avoid confusion and ambiguities, the target word is contextualized semantically and presented as follows: First, the word (e.g., *Hase* [bunny]) is spoken and simultaneously displayed visually as an image (e.g., an image of a bunny). Second, the word is spoken and contextualized in a sentence (e.g., "*der **Hase** hoppelt über die Wiese*" [the bunny hops across the meadow]). Lastly, the word is spoken again to focus the child's attention. After the word is spoken the second time, the letter area is displayed and the child may start writing the word. The letter area contains a unique key of each letter of the target word and, optionally, distracting letters. Children can listen to the word again by pressing the *listen* button. The spelling line can be reset by pressing the *reset* button, represented by a trash bin. Answers are logged in by pressing the *check*-button. If the answer is correct, positive feedback is given and the game continues with the next word. If not, the children are motivated to give it another try. After two failed attempts, the solution is presented and the game continues with the next word. Various parameters can be changed to adapt the difficulty of the levels, such as capitalization, usage and difficulty of distracting letters, and whether the number of letters of the target word is exposed.

In the drag-and-drop condition, letters are moved and placed as introduced in the previous section and shown in Figure 2b. Drag-and-drop interaction offers the most options: a letter can either be dropped on an empty slot in the spelling line, be inserted between two already placed letters, or replace an already placed letter. In the last case, if the moved letter originated from the spelling line, the replaced letter swaps the position with it. If the moved letter originated from the letter area, the replaced letter returns there. As recommended to reduce errors caused by drag-and-drop (cf. Donker & Reitsma, 2007a), we visually indicate where dragged letters can be dropped: For (re-)placement and insertion, the dragged letter and the letter/empty slots in the spelling line wiggle when their responsive zones intersect.

Point-and-touch as explained in the previous section is illustrated in Figure 2c. It offers the possibilities to place, delete, and swap letters. When a letter is first selected, the color is changed to light blue and the border begins to sparkle. Letters from the letter area can be placed on empty or occupied slots in the spelling line. In the latter case,



(a) Interaction style "Touch". Target word is *Hase* [bunny].

(b) Interaction style "Drag-and-drop". Target word is *müde* [tired]. The letter "e" is currently being dragged from the letter area to the spelling line.

(c) Interaction style "Point-and-touch". Target word is *Banane* [banana]. The letter "a" is currently selected in the letter area. While selected, touching a spelling line position will move the letter there.

Figure 2. Interaction styles of the proposed spelling game. Hand symbols are used in animated instructions of tutorials.

the already placed letter is removed to the letter area. Letters in the spelling line can be swapped by consecutively touching them. To delete letters from the spelling line, the respective letter needs to be touched twice, without temporal restriction. All interactions can be canceled by touching anywhere outside the spelling line. This results in the formerly selected letter to be deselected and its visual highlighting is reset.

The touch interaction style also introduced in the previous section is illustrated in Figure 2a. The first free slot of the spelling line has a luminous border and pulses lightly to indicate where the next touched letter from the letter area will be placed.

In interactive tutorials, children are step-wise introduced to the individual features of the interaction styles. To ensure that children understand everything, we used animated audio instead of textual instruction (cf. McKnight & Fitton, 2010), and the system waits upon successful execution of requested actions before the tutorial continues, i.e., placement, deletion, swapping, and insertion of letters, and using the buttons.

We followed guidelines regarding target size (cf. Hamza & Salivia, 2015) and recommendations to allow slightly out-of-bounds touches (cf. Anthony et al., 2014; Vatavu et al., 2015) to counteract inaccuracies in touches observed with drag-and-drop. The responsive zone of the letters and slots of the spelling line is slightly larger than their visible shapes and measures 16×16 mm on a 9.7" screen. Responsive zones additionally expand slightly in each direction not neighbored by another letter.

Table 1. Demographic data of participants by spelling proficiency and grade.

Grade	*Variables*	Typically Developing (TD) *M* (*SD*)	Developmental Dyslexics (DD) *M* (*SD*)	**All** **M (SD)**
3	*N*	9	3	**12**
	Boys/girls	4/5	3/0	**7/5**
	Age (years)	9.44 (0.52)	10.01 (0.79)	**9.58 (0.62)**
4	*N*	8	5	**13**
	Boys/girls	3/5	4/1	**7/6**
	Age (years)	10.1 (0.19)	10.34 (0.47)	**10.12 (0.33)**
All	*N*	17	8	**25**
	Boys/girls	7/10	7/1	**14/11**
	Age (years)	9.75 (0.52)	10.22 (0.58)	**9.9 (0.57)**

4. Experiment

4.1 Participants

Twenty-five German primary school children (14 boys and 11 girls) from third and fourth grade, aged 8–11 years ($M = 9.9$, $SD = 0.57$), participated in the study. Eight of the children were diagnosed with developmental dyslexia (seven boys and one girl). A detailed overview of the demographics is given in Table 1. The study was approved by the local ethics committee for psychological research (file number 2017_0823_76).

4.2 Material

Tablets. A Samsung Galaxy S2 (type SM-T813) and S3 tablet (type SM-T820), both with 9.7 inches screen size and running Android 7.0, were used for the study. The tablets were put in original Samsung covers (type EF-EF-BT810P and EF-BT820PB, respectively). The covers can be used as a stand and offer flat as well as steep

(a) Wooden ruler with movable peg used to collect children's workload ratings, developed by Laurie-Rose et al. (2014).

(b) Digital slider for touch devices to collect children's workload ratings, used in this article.

Figure 3. Sliders to collect children's workload ratings for use with adapted versions of the NASA-TLX, displayed for the mental demand subscale. The position of the arrow corresponds to values between 0 to 100 in steps of 5.

viewing angles.

Game. The spelling game explained in Section 3 was used for the study. In order to focus on the interaction style and not on the difficulty of spelling, we decided to set the capitalization automatically, to expose the number of letters of the target word, and to deploy an easy set of distracting letters, i.e., the distracting letters cannot cause the children to write homophones (i.e., words having the same pronunciation but different spelling and meaning). For example, the children were asked to write the word *Hase* [bunny] by arranging the respective letters from the letter area containing $\{H\ u\ l\ e\ D\ s\ a\ x\}$ in the four slots of the spelling line, see Figure 2a.

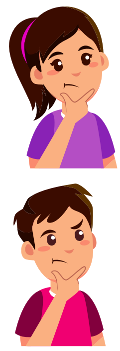
Word Lists. We created three equally difficult word lists of fifteen words each. To ensure equal difficulty among the word lists, we matched the words according to word length, orthographic challenges, and word frequency. The word difficulty increases linearly within each word list by increasing the word length and by adding orthographic challenges. For example, the word lists contain words you spell the way you hear them at the beginning (e.g., *Hase* [bunny] or *Hose* [pants]), then longer words (e.g., *Banane* [Banana]), then words with consonant clusters (e.g., *Blume* [flower]), and finally words with vowel length markers, such as words with *ie* (e.g., *besiegen* [to defeat sb.]) or with a consonant doubling (e.g., *brennen* [to burn]). Using word lists with increasing word difficulty, the children can first accustom with the new interaction style before writing orthographic challenging words. Since more difficult words may lead to misspellings (e.g., omitting one consonant of a double consonant), corrections may be required, depending on the individual spelling proficiency. This way, children may experience the writing of words with a particular interaction style with and without correcting a word by using its unique features to move, insert, replace, or delete letters.

4.3 Questionnaires

4.3.1 NASA-TLX

To measure perceived workload for each interaction style, we use an adapted version of the highly regarded, multidimensional workload scale NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). For this study, we decided to use the subscales *mental demand*, *physical demand*, *frustration*, and *effort* (leaving out the subscales *temporal demand* and *overall performance*). Laurie-Rose et al. (2014) adapted the NASA-TLX to measure perceived workload in children. The authors were able to show that, using two different tasks, task type elicited different task demands in children of first, second, fourth, and fifth grade. The authors suggest that an adapted NASA-TLX can be applied in educational settings to assess workload. To adapt the NASA-TLX for use with elementary school-aged children, the authors modified (and simplified) the NASA-TLX instructions to make the scale's wording accessible for the target group. Additionally, due to concerns about numerical literacy, they designed a wooden apparatus containing a numberless ruler with a movable peg. The children were asked to place the peg at a location that reflected their subjective experience. Finally, anime drawings for each subscale endpoints were attached to the wooden squares at either end of the numberless ruler to keep the children engaged in the task and to give visual anchors, which is advocated to use in studies were children are asked to use rating scales (cf. Tomlinson, von Baeyer, Stinson, & Sung, 2010). The wooden ruler

Table 2. Mental demand subscale of the adapted NASA-TLX with verbal descriptions for the touch interaction style. For drag-and-drop and point-and-touch, the phrase "touch the letters" is replaced by "dragging and dropping the letters" and "first selecting and then placing the letters", respectively.

Subscale Simple Title		**Adapted TLX Description** (verbal description translated from German)	
Mental Demand Nachdenken (Thinking)		How much did you have to think when you wrote the words by touching the letters? Did you have to do a lot of thinking and figuring out or not very much when you wrote the words by touching the letters? Look at the girl in the pictures. Over here (in the right picture), the girl looks like she is thinking very hard. On this side (in the left picture), it looks like the girl did not have to think a lot. How did you feel when you wrote the words by touching the letters? Show me on the slider how much thinking or figuring out you had to do.	

to collect children's workload ratings by Laurie-Rose et al. (2014) is displayed in Figure 3a.

For this study, we further adapted the kids version of the NASA-TLX described above. First, we developed a digital version of the slider for use with touch devices to smooth the movement of the peg. We do not expect any difficulties in use as the children will extensively work with the tablet in this study prior to answering the NASA-TLX. Second, we designed new drawings in comic style of each subscale endpoint for boys and girls individually. The sex of the drawings is adjusted automatically in the tablet application, according to the sex of the children. Lastly, we added a simplified title of each subscale above the slider. The digital kids version of the NASA-TLX is displayed in Figure 3b, the slider's value ranging from 0 to 100 with increments of 5. The adapted description for the mental demand subscale is listed in Table 2; physical demand, effort, and frustration are adapted accordingly. As the objective of this study was to compare the different interaction styles, we focused in the verbal descriptions explicitly on the interaction style and not on writing in general.

4.3.2 Smileyometer

In addition to the workload subscales of the NASA-TLX, we used self-designed questions to assess the subscales *ease of use*, *speed*, and *fun*. For this, we used 5-point Smileyometer (cf. Read, 2008) combined with a 5-point word scale, see Figure 4. The answer options for the subscales ease of use, speed, and fun ranged from *very hard* to *very easy*, *very slow* to *very fast*, and *not fun at all* to *very fun*, respectively.

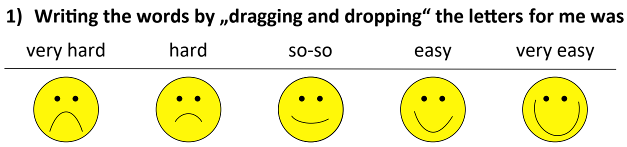


Figure 4. Smileyometer used for the subscales ease of use (shown), speed, and fun.

4.3.3 Fun-Sorter

Children do not necessarily respond honestly to rating scales (Hall, Hume, & Tazzyman, 2016). This may be caused, inter alia, by the desirably bias, i.e., children may not accurately respond regarding socially desirable characteristics in order to appear more appealing to the researcher (Evans et al., 2007), or by the acquiescence bias, i.e., the tendency to agree or respond positively to not tell the researcher that their product is not great (Danner, Aichholzer, & Rammstedt, 2015). One thus may fail to find a difference between the interaction styles even though the children subjectively perceive the interaction styles differently. To take this into account, we implemented the Fun-Sorter (cf. Read, 2008; Read, Macfarlane, & Casey, 2002) at the end of the experiment to explicitly compare the three interaction styles for the subscales *like*, *ease of use*, *fun*, *speed*, *effort*, and *mental demand*. The Fun-Sorter requires the children to rank the three interaction styles in order of preference of the respective subscale. We printed a representative picture of each interaction style on a magnetic card that can be placed in a magnetic

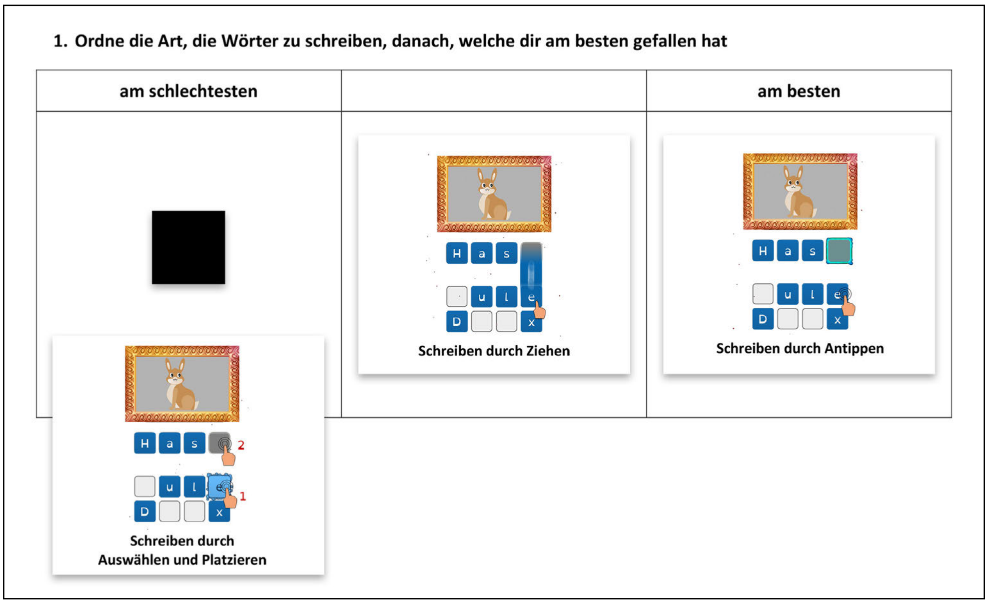


Figure 5. Fun-Sorter for the subscales like (shown), ease of use, fun, effort, speed, and mental demand. Interaction styles are represented by magnetic cards which are placed in individual cells. Order of the cells from left to right: worst, mediocre, best.

cell of the Fun-Sorter table, see Figure 5. We used a separate sheet for each subscales to prevent children from trying to balance their responses. Using magnetic cards, we aimed at maximizing the haptic and physical experience of the ranking of the interaction styles.

4.4 Procedure

We used a within-subject design with order of interaction styles and word lists as independent variables. To balance order of interaction styles and word lists, we computed all 36 unique combinations in advance and assigned them randomly to the 25 participants.

The experiment was carried out either in a specifically prepared room at the University or in a classroom of a local learning therapy facility.

After the parents and their children gave their informed consent, the parents were sent to a waiting room for the duration of the experiment. After the parents left, the children were seated in front of the tablet and told to make themselves feel comfortable and to adjust the viewing angle of the tablets according to their needs. We told them that they always can change the viewing angle of the tablet or take the tablet on their lap, whatever is the most comfortable.

The experiment started with the tutorial of the first interaction style. Upon completion, any remaining questions were clarified and the children continued to write the fifteen words of the corresponding word list with the corresponding interaction style. Upon completion, the children answered the subscales of the adapted NASA-TLX on the tablet. For the first interaction style, the NASA-TLX was introduced with the exemplary subscale "tiredness", for which the children were asked to rate how tired they felt this morning on a scale from "very tired" to "totally awake". We told them to drag the slider between the very sleepy looking boy and the very awake looking boy as to reflect their feeling this morning. After making sure that the children understood the rating scale and the control of the digital slider, they continued with the workload subscales.

Afterwards, the Smileyometer questionnaire was answered on paper. We explained in detail the Smileyometer rating scale. Based on other's experience of children's unfamiliarity with such rating scales (Holz, Beuttler, & Ninaus, 2018), we read aloud each question and its response options individually and clarified posed questions. The children were then asked to circle or cross the according smiley. Lastly, we asked the children to tell us what they specifically liked or disliked about the interaction style or how we could improve it. In total, one interaction style took about fifteen minutes. A short break of five minutes succeeded the questionnaires in which the children were invited to play board games with the experimenter.

The second and third interaction styles were carried in the same way. After finishing with the third interaction style, the children answered the final questionnaire including the subscales of the Fun-Sorter and the question regarding which interaction style they would choose to practice writing words if they only can choose one. Additionally, we gave them the opportunity for open comments about the interaction styles and the spelling game in general. The experiment lasted about fifty to sixty minutes. Upon completion, children were rewarded with a gift voucher from a local book store.

4.5 Analysis

We computed linear mixed regression models fitted by REML using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2014), in order to contrast the influence of interaction style (touch, drag-and-drop, point-and-touch), spelling proficiency (dyslexic [DD], typically developing [TD]), and grade (third, fourth) on the rating of the workload subscales of the adapted NASA-TLX, on the subscales of the Smileyometer questionnaire, and on writing times. For workload and user experience, we entered interaction style, grade, and spelling proficiency as fixed effect and rating as response. The models included random intercepts for children.[1] For writing times, we entered interaction style, grade, and spelling proficiency as fixed effect and writing time as response. The model included random intercepts for children and words.[2]

Degrees of freedom and p-values were calculated with the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017) using Satterthwaite's approximation for the denominator degrees of freedom. Pairwise comparisons were obtained from the table of differences of least square means provided by the *lmerTest* package.

To investigate the interaction effects between interaction style and spelling proficiency, and between interaction style and grade, we enhanced the aforementioned models by additionally entering the interaction effect in question as fixed effect and report F-tests of the interaction effect in question.[3]

[1]Model specification for workload and user experience:
`Rating~InteractionStyle+Grade+Spelling+(1|Subject)`

[2]Model specification for writing time:
```
Time~InteractionStyle+Grade+Spelling+(1|Subject)
                                    +(1|Item)
```

[3]Alternatively, likelihood ratio tests (LRT) of the full model with the effect in question against a reduced model without the effect in questions can be used to obtain p-values. F-tests and LRTs yielded the same test decisions for all interaction effects in question.

The criterion of statistical significance was set at $\alpha = .05$. Visual inspection of residual plots and Fligner-Killeen median centering χ^2 tests (cf. Conover, Guerrero-Serrano, & Tercero-Gómez, 2018) did not reveal significant deviation from homoskedasticiy for any model.

5. Results

Descriptive statistics are listed in Table 3. Model fits of all subscales are listed in Table A1 and Table A2. The data from one child (female typically developing third grader) was excluded from the analyses of the workload subscales of the NASA-TLX and from the analyses of the writing times due to tablet crashes during the experiment.

5.1 NASA-TLX

Descriptive statistics are given in Figure 6 and Table 3, Table A1 shows the respective model fits for the workload subscales. The children's ratings ranged between 0 and 100 with increments of five, see. Section 4.3.1.

Mental Demand. Touch was perceived as significantly less mentally demanding than point-and-touch, $\beta = -19.37$, $SE = 4.12$, $t(46) = -4.7$, $p < .001$. Further, touch tended to be perceived as less mentally demanding than drag-and-drop, although the difference was not significant, $\beta = -7.08$, $SE = 4.12$, $t(46) = -1.72$, $p = .093$. However, this difference is primarily driven by typically developing children while dyslexic children did not seem to report touch as being less mentally demanding than drag-and-drop, see Figure 6.

Drag-and-drop was perceived as significantly less mentally demanding than point-and-touch, $\beta = -12.29$, $SE = 4.12$, $t(46) = -2.98$, $p = .005$. While the main effect of spelling proficiency was not significant, $\beta = 9.73$, $SE = 5.16$, $t(21) = 1.89$, $p = .073$, dyslexic children tended to report higher mental demand than typically developing children. The main effect of grade was not significant, $\beta = 3.69$, $SE = 4.88$, $t(21) = 0.76$, $p = .458$. We neither found a significant interaction between interaction style and spelling proficiency, $F_{2,\,44} = 2.38$, $p = .104$, nor between interaction style and grade, $F_{2,\,44} = 1.62$, $p = .209$.

Physical Demand. Touch was perceived as significantly less physically demanding than point-and-touch, $\beta = -19.79$, $SE = 4.84$, $t(46) = -4.09$, $p < .001$. Further, touch was perceived as significantly less physically demanding than drag-and-drop, $\beta = -19.17$, $SE = 4.84$, $t(46) = -3.96$,

Table 3. Mean (SD) writing times and workload and user experience ratings by subscale, interaction style (i.e., touch, drag-and-drop [DnD], and point-and-touch [PnT]), and spelling proficiency.

				Interaction Style		
Children	Scale	Subscale	N	Touch	DnD	PnT
All Children	Workload (NASA-TLX)	Mental Demand	24	14.2 (14.2)	21.3 (14.6)	33.5 (21.1)
		Phys. Demand	24	12.9 (22.8)	32.1 (29.4)	32.7 (26.0)
		Frustration	24	5.0 (13.5)	14.2 (19.2)	16.5 (22.9)
		Effort	24	8.8 (12.4)	23.1 (22.9)	24.2 (23.4)
	Smileyometer	Ease of Use	25	4.6 (0.6)	4.2 (0.8)	4.3 (0.7)
		Fun	25	4.5 (0.7)	4.0 (0.9)	4.1 (0.9)
		Speed	25	4.2 (0.8)	3.8 (0.6)	3.7 (0.8)
		Writing Time	24	9.2 s (6.5)	14.4 s (9.2)	14.0 s (8.7)
Typically Developing	Workload (NASA-TLX)	Mental Dem.	16	10.3 (9.4)	21.6 (15.8)	27.8 (21.4)
		Phys. Dem.	16	6.3 (7.4)	30.9 (25.7)	28.1 (21.1)
		Frustration	16	3.8 (8.1)	15.3 (20.3)	15.9 (23.8)
		Effort	16	5.0 (5.8)	25.3 (25.7)	22.5 (23.6)
	Smileyometer	Ease of Use	17	4.7 (0.5)	4.1 (0.9)	4.3 (0.6)
		Fun	17	4.5 (0.7)	3.9 (1.0)	4.1 (0.9)
		Speed	17	4.2 (0.8)	3.7 (0.5)	3.7 (0.7)
		Writing Time	16	8.1 s (5.0)	13.9 s (8.9)	12.6 s (6.1)
Developmental Dyslexics	Workload (NASA-TLX)	Mental Dem.	8	21.9 (19.3)	20.6 (12.9)	45.0 (16.3)
		Phys. Dem.	8	26.3 (35.9)	34.4 (37.7)	41.9 (33.7)
		Frustration	8	7.5 (21.2)	11.9 (17.7)	17.5 (22.5)
		Effort	8	16.3 (18.3)	18.8 (16.6)	27.5 (25.4)
	Smileyometer	Ease of Use	9	4.5 (0.9)	4.4 (0.7)	4.3 (0.9)
		Fun	9	4.5 (0.8)	4.3 (0.7)	4.1 (0.6)
		Speed	9	4.4 (0.7)	4.1 (0.6)	3.8 (1.0)
		Writing Time	8	11.6 s (58.6)	15.4 s (9.7)	16.8 s (12.0)

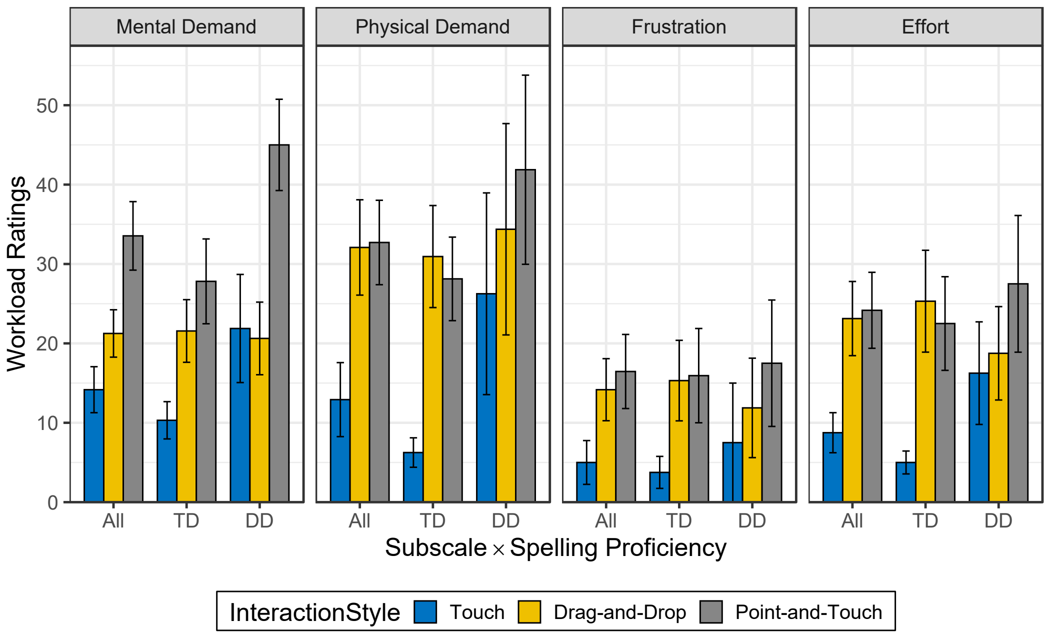


Figure 6. Children's workload ratings of the adapted NASA-TLX by subscale (facets), interaction style (color), and spelling proficiency (i.e., all, typically developing [TD], and dyslexic children [DD]). Bars represent standard errors of the mean.

$p < .001$. The difference between drag-and-drop and point-and-touch was not significant, $\beta = -0.63$, $SE = 4.84$, $t(46) = -0.13$, $p = .898$. The main effect of spelling proficiency was not significant, $\beta = 14.28$, $SE = 9.21$, $t(21) = 1.55$, $p = .136$. Although third graders tended to report higher physical demand than fourth graders, the main effect of grade was not significant, $\beta = 15.04$, $SE = 8.72$, $t(21) = 1.72$, $p = .099$. We neither found a significant interaction between interaction style and spelling proficiency, $F_{2,\,44} = 1.35$, $p = .270$, nor between interaction style and grade, $F_{2,\,44} = 2.14$, $p = .130$.

Frustration. Touch was perceived as significantly less frustrating than point-and-touch, $\beta = -11.46$, $SE = 4.65$, $t(46) = -2.46$, $p = .018$. Further, touch tended to be perceived as less frustrating than point-and-touch, although the difference was not significant, $\beta = -9.17$, $SE = 4.65$, $t(46) = -1.97$, $p = .055$. The difference between drag-and-drop and point-and-touch was not significant, $\beta = -2.29$, $SE = 4.65$, $t(46) = -0.49$, $p = .625$. We neither found a significant main effect of spelling proficiency, $\beta = 1.15$, $SE = 6.12$, $t(21) = 0.19$, $p = .852$, nor of grade, $\beta = 4.22$, $SE = 5.79$, $t(21) = -0.73$, $p = .47$. We neither found a significant interaction between interaction style and spelling proficiency, $F_{2,\,44} = 0.27$, $p = .765$, nor between interaction style and grade, $F_{2,\,44} = 1.01$, $p = .371$.

Effort. Children reported significantly lower effort in writing words with touch compared to with point-and-touch, $\beta = -15.42$, $SE = 5.18$, $t(46) = -2.94$, $p = .005$. Further, touch was perceived to require significantly less effort than drag-and-drop, $\beta = -14.37$, $SE = 5.18$, $t(46) = -2.77$, $p = .008$. The difference between drag-and-drop and point-and-touch was not significant, $\beta = -1.04$, $SE = 5.18$, $t(46) = -0.20$, $p = .842$). We neither found a significant main effect of spelling proficiency, $\beta = 3.76$, $SE = 6.24$, $t(21) = 0.60$, $p = .55$, nor of grade, $\beta = 4.26$, $SE = 5.90$, $t(21) = -0.72$, $p = .48$. We neither found a significant interaction between interaction style and spelling proficiency, $F_{2,\,44} = 1.37$, $p = .264$, nor between interaction style and grade, $F_{2,\,44} = 2.41$, $p = 0.102$.

5.2 Ease of Use, Fun, and Speed

Descriptive statistics are given in Figure 7 and Table 3, Table A2 shows respective model fits. The ratings of the children range from 1 to 5 in increments of 1, cf. Section 4.3.2.

Ease of Use. Children reported significantly higher ease of use for touch than for point-and-touch, $\beta = 0.36$, $SE = 0.17$, $t(48) = 2.17$, $p = .035$. Further, touch was perceived as significantly easier to use than drag-and-drop, $\beta = 0.44$, $SE = 0.17$, $t(48) = 2.65$, $p = .011$. The difference between drag-and-drop and point-and-touch was not significant, $\beta = -0.08$, $SE = 0.17$, $t(48) = -0.48$, $p = .633$. We neither found a significant main effect of spelling proficiency, $\beta = -0.02$, $SE = 0.24$, $t(22) = -0.08$, $p = .940$, nor of grade, $\beta = -0.13$, $SE = 0.22$, $t(22) = -0.60$, $p = .556$. We did not find significant interactions between interaction style and spelling proficiency, $F_{2,\,46} = 0.86$, $p = .428$, nor between interaction style and grade, $F_{2,\,46} = 0.93$, $p = .403$.

Fun. Touch tended to be perceived as more fun than point-and-touch, however the difference was not significant, $\beta = 0.40$, $SE = 0.22$, $t(48) = 1.81$, $p = .077$. Further, touch was perceived as significantly more fun than drag-and-drop, $\beta = 0.48$, $SE = 0.22$, $t(48) = 2.17$, $p = .035$. The difference between drag-and-drop and point-and-touch was not significant, $\beta = -0.08$, $SE = 0.22$, $t(48) = -.36$, $p = .719$. We neither found a significant main effect of spelling proficiency, $\beta = 0.04$, $SE = 0.25$, $t(22) = 0.15$, $p = .882$, nor of grade, $\beta = -0.38$, $SE = 0.23$, $t(22) = -1.63$, $p = .118$. We did not find significant interactions between interaction style and spelling proficiency, $F_{2,\,46} = 0.30$, $p = 0.744$, nor between interaction style and grade, $F_{2,\,46} = 2.55$, $p = .089$.

Speed. Children reported to perceive touch as significantly faster than point-and-touch, $\beta = 0.52$, $SE = 0.18$, $t(48) = 2.86$, $p = .006$. Further, touch was perceived as significantly faster than drag-and-drop, $\beta = 0.40$, $SE = 0.18$, $t(48) = 2.12$, $p = .032$. The difference between drag-and-drop and point-and-touch was not significant, $\beta = 0.12$, $SE = 0.18$, $t(48) = 0.66$, $p = .510$. We did not find a significant main effect of spelling proficiency, $\beta = 0.19$, $SE = 0.21$, $t(22) = -0.88$, $p = .386$, nor of grade, $\beta = -0.22$, $SE = 0.20$, $t(22) = -1.12$, $p = .275$. We did not find significant interactions between interaction style and spelling proficiency, $F_{2,\,46} = 0.46$, $p = .633$, nor between interaction style and grade, $F_{2,\,46} = 0.06$, $p = .939$.

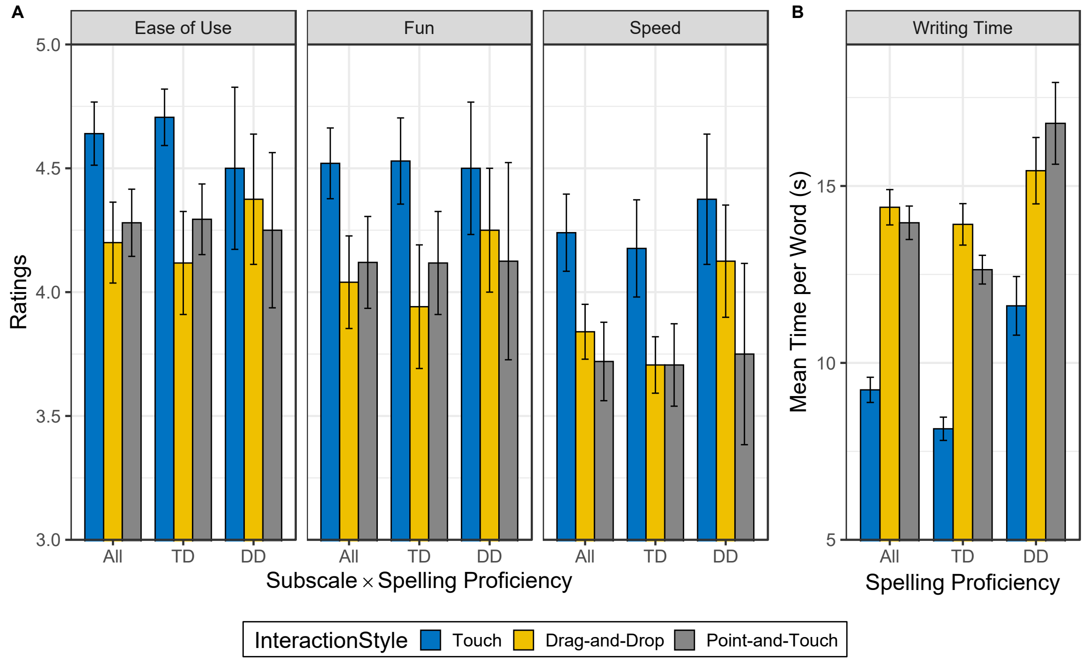


Figure 7. Children's subjective user experience ratings and writing time by subscale (facets), interaction style (color), and spelling proficiency (i.e., all, typically developing [TD], and dyslexic children [DD]). Bars represent standard errors of the mean. **A**: Children's user experience ratings; **B**: Children's average writing times.

5.3 Writing Times

We computed the writing time by subtracting the children's first interaction with a letter from the completion time of a task (i.e., the time when the correct answer was logged in). The writing time is reported in seconds. As our interest was in how fast children write words with a specific interaction style, we excluded all trials in which the word was not written correctly at the first attempt (41 trials excluded). Additionally, we removed trials per interaction style and child whose writing time deviated by more than 2.5 standard deviations (28 trials excluded). In total, 1011 trials were used in the analyses. Descriptive statistics of the writing times by interaction style and spelling proficiency are given in Figure 7 and Table 3. The model fit is listed in Table A2.

Children were significantly faster in writing words with touch than with point-and-touch, $\beta = 4.81$ s, $SE = 0.47$ s, $t(942.1) = 10.14$, $p < .001$. Further, children wrote words significantly faster with touch than with drag-and-drop, $\beta = 5.26$ s, $SE = 0.47$ s, $t(944.5) = 11.00$, $p < .001$. The difference between drag-and-drop and point-and-touch was not significant, $\beta = 0.45$ s, $SE = .47$ s, $t(942) = 0.95$, $p = .341$.

We found a significant main effect of spelling proficiency on writing time, $\beta = 3.54$ s, $SE = 1.15$ s, $t(21.1) = 2.91$, $p = .008$. Dyslexic children needed significantly more time to write words than typical developing children. We also found a significant main effect of grade, $\beta = 3.22$ s, $SE = 1.15$ s, $t(20.9) = 2.81$, $p = .011$. Third graders needed significantly more time than fourth graders.

5.4 Fun-Sorter

The children's rankings of the interaction styles by subscale and spelling proficiency are listed in Figure 8. We used Pearson's Chi-squared test to test the differences in ranking for significance.

Touch was favored by most children over drag-and-drop and point-and-touch in all subscales. The difference of ranking is significant for all subscales except for fun (*like*: $\chi^2(4) = 26.88$, $p < .001$; *ease of use*: $\chi^2(4) = 39.36$, $p < .001$; *fun*: $\chi^2(4) = 6.72$, $p = .15$; *effort*: $\chi^2(4) = 52.08$, $p < .001$, *speed*: $\chi^2(4) = 20.64$, $p < .001$; and *mental demand*: $\chi^2(4) = 50.16$, $p < .001$). Drag-and-drop was preferred to point-and-touch in all subscales except for like, where drag-and-drop seemed to polarize the children: five children liked drag-and-drop best while eleven children liked it least. Point-and-touch was ranked worst in all categories except like. We did not observe divergent preferences between dyslexic and typically developing children.

Seventeen children reported that touch would be their preferred interaction style if they only can choose one to write words with (Figure 8, subscale

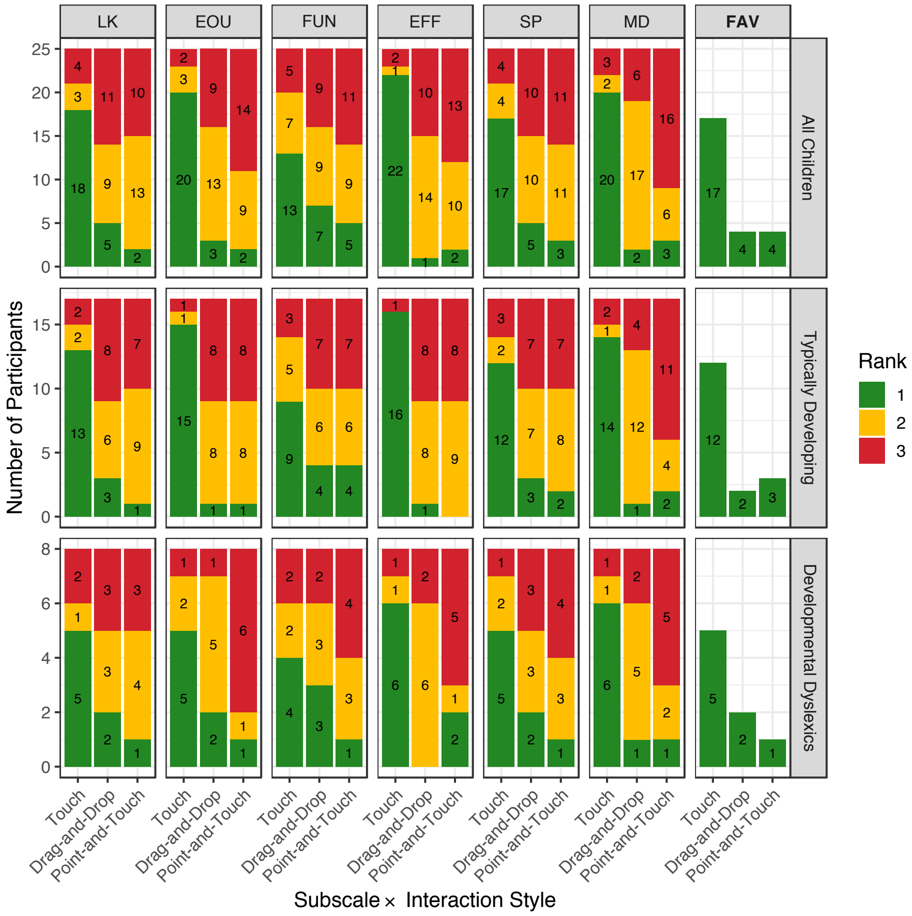


Figure 8. Children's Fun-Sorter rankings of the interaction styles by subscale (facets) and spelling proficiency (all = all children; TD = typically developing; DD = dyslexic). LK = like; EOU = ease of use; FUN = fun; EFF = effort; SP = speed; MD = mental demand; FAV = favorite.

FAV). Drag-and-drop and point-and-touch were chosen by four children each.

The choice did not differ significantly between dyslexic and typically developing children, $\chi^2(4) = 6$, $p = .20$.

6. Discussion

Our results suggest that simple touch interaction is the most appropriate interaction style for children in the proposed mobile spelling game when compared to point-and-touch and drag-and-drop interaction. We found that children were faster in writing words with touch than with drag-and-drop or point-and-touch. This is confirmed by the subjective ratings of perceived speed, writing times, and free comments. Looking at the repeated subjective ratings of workload and user experience subscales, touch was reported to be less physically demanding, less effortful, easier to use, more fun, and – compared to point-and-touch mainly – less frustrating and less mentally demanding. While typically developing children also reported touch as being less mentally demanding and less frustrating than drag-and-drop, this was not observable as such for dyslexic children (see Figure 6), possibly due to the low sample size of only eight dyslexic children. But the appropriateness of touch becomes systematically visible when looking at the results of direct rankings: touch was ranked better than drag-and-drop and point-and-touch by typically developing and dyslexic children in all subscales, i.e., like, ease of use, fun, effort, speed, and mental demand. Finally, touch was selected most often by 17 out of 25 children as their interaction style of choice.

The findings indicating that touch should be favored over drag-and-drop and point-and-touch in a mobile spelling exercise may be caused by the fact that touching requires less interactions and hand/finger movements if words are written without correction, that children are getting increasingly familiar with mobile touch devices, or that click or simple touch is used in (on-screen) keyboards. This is supported by comments stating that touch felt like writing on a computer keyboard or on children's (or their parent's) mobile devices.

Interestingly, our findings contradict earlier results and design recommendations concluding that point-and-click or point-and-touch is more appropriate than drag-and-drop (Chiasson & Gutwin, 2005; Gelderblom & Kotzé, 2009; Girard & Johnson, 2009; Inkpen, 2001; Joiner et al., 1998; Roman, 2015; Soni et al., 2019; Ward, 2014). We found small advantages of drag-and-drop over point-and-touch in reported mental demand; subjective ratings of other workload and user experience subscales and writing times did not differ. Further advantages become more visible in ranking responses providing a more fine-grained picture showing that drag-and-drop was generally favored over point-and-click. Thus, our results concur more with reports on drag-and-drop to be unproblematic or even more appropriate than point-and-click (Barendregt, 2015; Barendregt & Bekker, 2011; Donker & Reitsma, 2007a), specifically for touch-based devices (Hamza & Salivia, 2015). We assume contradictory results arise from the type of task children performed, the input modality (mouse vs. touch), children's ages, and the previous experience of children interacting with technology at the time the study was conducted. In the current era of mobile touch devices, swiping and dragging becomes more and more part of children's everyday lives. Children may expect the same interaction styles they have been using in other applications. Also, whether the performed actions have a more natural mapping to either of the interaction styles is an important factor to consider

Point-and-Touch vs. Drag-and-Drop. Considering point-and-touch and drag-and-drop, we did not observe any difficulties in children using these interaction styles. Furthermore, children seemed to be already experienced in dragging and dropping, concurring recent findings (Barendregt, 2015; Hamza & Salivia, 2015). We observed that children successfully used all features of point-and-touch and drag-and-drop, i.e., deletion, swap, and insertion of letters. Two children exclaimed that they specifically liked that drag-and-drop – compared to the other interaction styles – offers the possibility of directly inserting letters between two already written letters. This was confirmed when we asked the children how we could improve the spelling game: some of them responded that they really liked swapping and inserting letters and suggested to implement these features also for the touch interaction style.

Comparing drag-and-drop and point-and-touch, considering repeated subjective workload and user experience measures, the results suggest that drag-and-drop was less mentally demanding. The difference in subjectively perceived mental demand may result from drag-and-drop interaction making it possible to free cognitive resources through offloading (Antle, 2013), i.e., by dragging letters in a playful way before dropping them. Although drag-and-drop was only rated to be more physically demanding and taking more effort than touch, but not than point-and-touch, a few children commented that drag-and-drop is generally "more exhausting for the fingers and the arm". A more distinct differentiation becomes visible when children are forced to directly compare the two interaction styles: drag-and-drop was ranked better than point-and-touch regarding ease of use, fun, effort, speed, and mental demand. The subscale like yields no clear difference between drag-and-drop and point-and-touch and seemed to polarize the children. While drag-and-drop was liked best more often, it was also liked least more often and point-and-touch was mostly rated as the second best interaction style. The implications of the comparison of drag-and-drop and point-and-touch are two-fold. First, it seems that dyslexic and typically developing children perceived and handled these interaction styles generally somewhat alike, with a slight tendency towards the advantage of drag-and-drop. This is also reflected in that drag-and-drop and point-and-touch were selected by four children each as their interaction style of choice. Second, it highlights the importance of direct rankings in user studies to get better insights.

Relevance of Implementation Specifics. Respective point-and-touch, we observed five children trying to swap letters between the letter area and the spelling line by first selecting the letter in the spelling line and then touching a letter in the letter area. The study version only supported to swap letters the other way round. We suggest to implement swap both ways when opting for point-and-touch.

In pilot tests, we observed that children using the drag-and-drop interaction styles tried to drag letters very accurately to destined locations before dropping them, taking more time and effort than necessary. We thus enhanced the tutorial for drag-and-drop by clarifying the sufficiency to drag the letter just close to the destined location. We assume this additional explanation prevented ratings and writing times that would erroneously have led to the disadvantage of drag-and-drop. Thus, besides sufficiently big responsive zones and allowance of out-of-bound touches, it is of utmost importance to inform the children *precisely* how the interaction style works to make inferences on performance and other metrics.

Impact of Child Differences. Children's spelling proficiencies were expectantly reflected in writing times and reported mental demand: dyslexic children needed more time and tended to report higher mental demand than typically developing children.

The influence of children's physical and mental development was reflected in that third graders tended to report higher physical demand and were slower in writing words than fourth graders, concurring with recent studies (e.g., Hamza & Salivia, 2015; Vatavu et al., 2015).

Relevance of Different Measurement Methods. In this experiment, apart from writing times, we used subjective reports of workload and user experience, observations from the experimenter, and ranking responses to compare the interaction styles. During the experiment, it was sometimes hard to tell for the experimenter whether the children actually rated the interaction styles differently on the Smileyometer or on the adapted kids version of the NASA-TLX. However, our results suggest that (repeated) subjective reports and specifically direct rankings are valuable in comparing multiple systems or versions of a system. Future research could include objective biophysical markers, such as facial expression, pupillometry, or heart rate, to avoid relying solely on subjective reports and observations.

7. Conclusion

The current state of research is puzzling when it comes to determining the most appropriate (touch) interaction style for children (Barendregt, 2015; Donker & Reitsma, 2007a; Hourcade, 2008). To determine the most appropriate interaction style in a tabled-based spelling game, and to disentangle currently reported contradictions, we compared drag-and-drop, point-and-touch, and touch in a lab experiment. We asked twenty-five German children aged 8–11 years, eight of whom were dyslexic, to arrange letters in a spelling line to write single words.

We were able to demonstrate that children aged 8–11 years can use drag-and-drop, point-and-touch, and touch without problems in the presented spelling game. Further, we observed that children are aware of and use unique features that constitute each interaction style, e.g., swapping letters using point-and-touch or inserting letters using drag-and-drop.

Our results suggest that touch is the least mentally and physically demanding, the least effortful and frustrating, the easiest to use, the most fun, and the fastest interaction style among the three. Additionally, touch was favored to drag-and-drop and point-and-touch in direct rankings with regard to liking, ease of use, fun, effort, speed, and mental demand. Finally, touch was selected as the interaction style of choice by 17 out of 25 children, whereas four children each chose drag-and-drop or point-and-touch as their favorite. Based on our results, touch seems to be the most appropriate interaction style in an educational touch-based spelling game – independent of spelling proficiency and grade.

Possibly, a hybrid interaction style combining features of various interaction styles would be even more appropriate. This was also proposed by the children who reportedly liked the possibility to swap and insert letters and wanted to have the same features in the touch interaction style. E.g., the interaction style starts in the touch mode and long pressing or movements of a fingertip length (McKnight & Fitton, 2010) switches into drag-and-drop or point-and-touch, enabling swapping of letters or inserting between already written letters in the spelling line without prior deletion.

Regarding the different groups of spelling proficiency, i.e., dyslexic and typically developing children, we found no significant interaction effects in the analyses of the three interaction styles. But dyslexic children were systematically slower, they tended to report higher mental demand when spelling the words, and their ratings did not differ as strongly between touch and drag-and-drop on some subscales. Thus, the conclusions drawn above for the different interaction styles in general hold for both, typically developing and dyslexic third and fourth graders.

Further research is still required to compare touch-based interaction styles in contexts that specifically address not only performance metrics, but also take into account workload, user experience, and children's habits and preferences. This is specifically the case with the emerging use of educational touch-based applications. We advice researchers and designers to select the interaction style carefully with regard to age-group, input modality, and context – and to not rely on results that do not address the requirements of their applications.

Acknowledgments

This research was funded by the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments. Heiko Holz is a doctoral student at the LEAD Graduate School & Research Network.

References

Anthony, L., Brown, Q., Tate, B., Nias, J., Brewer, R., & Irwin, G. (2014). Designing smarter touch-based interfaces for educational contexts. *Personal and Ubiquitous Computing*, *18*(6), 1471–1483. doi:10.1007/s00779-013-0749-9

Antle, A. N. (2013). Research opportunities: Embodied child–computer interaction. *International Journal of Child-Computer Interaction*, *1*(1), 30–36. doi:10.1016/j.ijcci.2012.08.001

Azah, N., Syuhada, N., Batmaz, F., Stone, R., & Wai, P. (2014). Selection of touch gestures for children's applications: Repeated experiment to increase reliability. *International Journal of Advanced Computer Science and Applications*, *5*(4), 97–102. doi:10.14569/IJACSA.2014.050415

Barendregt, W. (2015). Determining the most appropriate interaction style for children: Taking spontaneous use into account. doi:10.13140/RG.2.1.2463.2163

Barendregt, W., & Bekker, M. M. (2011). Children may expect drag-and-drop instead of point-and-click. In *CHI '11 extended abstracts on human factors in computing systems* (p. 1297–1302). doi:10.1145/1979742.1979764

Baschera, G.-M. (2011). *Modeling and Evaluation of Computer-Assisted Spelling Learning in Dyslexic Children* (Doctoral dissertation, ETH Zurich). doi:10.3929/ethz-a-007086954

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Berkling, K. (2017). Phontasia: A phonics game for German and its effect on orthographic skills–first corpus explorations. In *Proc. WOCCI 2017: 6th international workshop on child computer interaction* (pp. 7–16). doi:10.21437/WOCCI.2017-2

Brainy Ape Studio, LLP. (2017). Kinder lernen deutsche Wörter - lesen und schreiben [Children Learn German Words – Reading and Spelling] [Mobile application software]. Retrieved from https://play.google.com/store/apps/details?id=com.puzzlingapps.spelling.de.free

Chiasson, S., & Gutwin, C. (2005). Design principles for children's technology. *interfaces*, *7*(28), 1–9. Retrieved from http://www.hci.usask.ca/publications/2005/HCI_TR_2005_02_Design.pdf

Conover, W. J., Guerrero-Serrano, A. J., & Tercero-Gómez, V. G. (2018). An update on 'a comparative study of tests for homogeneity of variance'. *Journal of Statistical Computation and Simulation*, *88*(8), 1454–1469. doi:10.1080/00949655.2018.1438437

Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality*, *57*, 119–130. doi:10.1016/j.jrp.2015.05.004

Donker, A., & Reitsma, P. (2007a). Drag-and-drop errors in young children's use of the mouse. *Interacting with Computers*, *19*(2), 257–266. doi:10.1016/j.intcom.2006.05.008

Donker, A., & Reitsma, P. (2007b). Young children's ability to use a computer mouse. *Computers & Education*, *48*(4), 602–617. doi:10.1016/j.compedu.2005.05.001

Evans, G. W., Brauchle, G., Haq, A., Stecker, R., Wong, K., & Shapiro, E. (2007). Young children's environmental attitudes and behaviors. *Environment and Behavior*, *39*(5), 635–658. doi:10.1177/0013916506294252

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS ONE*, *9*(2), e89900. doi:10.1371/journal.pone.0089900

Galuschka, K., & Schulte-Körne, G. (2016). Clinical practice guideline: The diagnosis and treatment of reading and/or spelling disorders in children and adolescents. *Dtsch Arztebl Int.*, *113*(16), 279–86. doi:10.3238/arztebl.2016.0279

Gelderblom, H., & Kotzé, P. (2009). Ten design lessons from the literature on child development and children's use of technology. In *Proceedings of the 8th international conference on interaction design and children - IDC '09* (p. 52). doi:10.1145/1551788.1551798

Girard, S., & Johnson, H. (2009). Educational software for children: analysis of interaction techniques for direct manipulation. In (pp. 259–262). doi:10.1145/1629826.1629868

Hall, L., Hume, C., & Tazzyman, S. (2016). Five degrees of happiness: Effective smiley face likert scales for evaluating with children. In *Proceedings of the the 15th international conference on interaction design and children - IDC '16* (pp. 311–321). doi:10.1145/2930674.2930719

Hamza, Z., & Salivia, G. (2015). Study of touch gestures performance in touch devices by young children. *International Journal on Recent and Innovation Trends in Computing and Communication*, *3*(3), 1395–1400. doi:10.17762/ijritcc2321-8169.1503107

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (Vol. 52, pp. 139–183). North-Holland. doi:10.1016/S0166-4115(08)62386-9

Holz, H., Beuttler, B., & Ninaus, M. (2018). Design rationales of a mobile game-based intervention for german dyslexic children. In *Proceedings of the 2018 annual symposium on computer-human interaction in play companion extended abstracts* (pp. 205–219). doi:10.1145/3270316.3272053

Holz, H., Brandelik, K., Beuttler, B., Brandelik, J., & Ninaus, M. (2018). How to train your syllable stress awareness – A digital game-based intervention for German dyslexic children. *International Journal of Serious Games*, *5*(3), 37–59. doi:10.17083/ijsg.v5i3.242

Holz, H., Ninaus, M., Meurers, D., & Kirsch, A. (2018). Validity and player experience of a mobile game for German dyslexic children. In *Proceedings of the 2018 annual symposium on computer-human interaction in play companion extended abstracts* (pp. 469–478). doi:10.1145/3270316.3271523

Hourcade, J. P. (2008). Interaction design and children. *Human–Computer Interaction*, *1*(4), 277–392. doi:10.1561/1100000006

Inkpen, K. (2001). Drag-and-drop versus point-and-click mouse interaction styles for children. *ACM Transactions on Computer-Human Interaction*, *8*(1), 1–33. doi:10.1145/371127.371146

Inkpen, K., Booth, K. S., & Klawe, M. (1996). *Drag-and-drop vs. point-and-click mouse interaction for children* (Technical Report No. 96-17). Vancouver, BC, Canada: Department of Computer Science, University of British Columbia.

Joiner, R., Messer, D., Light, P., & Littleton, K. (1998). It is best to point for young children: a comparison of children's pointing and dragging. *Computers in Human Behavior*, *14*(3), 513–529. doi:10.1016/S0747-5632(98)00021-1

Kargl, R., Purgstaller, C., Weiss, S., & Fink, A. (2008). Effektivitätsüberprüfung eines morphemorientierten Grundwortschatz-Segementierungstrainings (MORPHEUS) bei Kindern und Jugendlichen [Effectiveness of a morpheme-oriented basic vocabulary segmenting training (MORPHEUS) for children and adolescents]. *Heilpädagogische Forschung*, *34*(3), 147–156.

Kast, M., Baschera, G.-M., Gross, M., Jäncke, L., & Meyer, M. (2011). Computer-based learning of spelling skills in children with and without dyslexia. *Annals of Dyslexia*, *61*(2), 177–200. doi:10.1007/s11881-011-0052-2

Klatte, M., Bergström, K., Steinbrink, C., Konerding, M., & Lachmann, T. (2018). Effects of the computer-based training program Lautarium on phonological awareness and reading and spelling abilities in German second-graders. In T. Lachmann & T. Weis (Eds.), *Reading and dyslexia: From basic functions to higher order cognition* (pp. 317–333). doi:10.1007/978-3-319-90805-2_15

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). doi:10.18637/jss.v082.i13

Laurie-Rose, C., Frey, M., Ennis, A., & Zamary, A. (2014). Measuring perceived mental workload in children. *The American Journal of Psychology*, *127*(1), 107. doi:10.5406/amerjpsyc.127.1.0107

McKnight, L., & Fitton, D. (2010). *Touch-screen technology for children: Giving the right instructions and getting the right responses* (No. January 2010). doi:10.1145/1810543.1810580

Mildenberger Verlag GmbH. (2016). ABC der Tiere 1 [Animal Alphabet 1] [Mobile application software]. Retrieved from https://play.google.com/store/apps/details?id=air.de.mildenbergerverlag.abc1android

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.r-project.org/

Rauschenberger, M., Fuechsel, S., Rello, L., Bayarri, C., & Gòrriz, A. (2015). A game to target the spelling of german children with dyslexia. In *Proceedings of the 17th international acm sigaccess conference on computers & accessibility* (pp. 445–446). doi:10.1145/2700648.2811345

Read, J. C. (2008). Validating the fun toolkit: An instrument for measuring children's opinions of technology. *Cognition, Technology & Work*, *10*(2), 119–128. doi:10.1007/s10111-007-0069-9

Read, J. C., Macfarlane, S., & Casey, C. (2002). Endurability, engagement and expectations: Measuring children's fun. In *Interaction design and children* (Vol. 2, pp. 1–23).

Rello, L., Bayarri, C., & Gòrriz, A. (2013). Dyslexia exercises on my tablet are more fun. *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility - W4A '13*, 1. doi:10.1145/2461121.2461148

Rello, L., Bayarri, C., Otal, Y., & Pielot, M. (2014). A computer-based method to improve the spelling of children with dyslexia. In *Proceedings of the 16th international acm sigaccess conference on computers & accessibility - ASSETS '14* (pp. 153–160). doi:10.1145/2661334.2661373

Roman, K. (2015). *Children's ability to navigate through a streaming services such as TV4 Play* (Master's thesis, KTH, School of Computer Science and Communication (CSC)). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-169931

Schwartz, R. N., & Plass, J. L. (2014). Click versus drag: User-performed tasks and the enactment effect in an interactive multimedia environment. *Computers in Human Behavior*, *33*, 242–255. doi:10.1016/j.chb.2014.01.012

Soegaard, M. (2015). Interaction Styles. In *The Glossary of Human Computer Interaction.* The Interaction Design Foundation. Retrieved from https://www.interaction-design.org/literature/book/the-glossary-of-human-computer-interaction/interaction-styles

Soliman, S., & Nathan-Roberts, D. (2018). Evaluating children's interaction with touchscreens from 0 to 8 years old. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 260–264. doi:10.1177/1541931218621060

Soni, N., Aloba, A., Morga, K. S., Wisniewski, P. J., & Anthony, L. (2019). A framework of touchscreen interaction design recommendations for children (TIDRC). In *Proceedings of the 18th acm international conference on interaction design and children* (pp. 419–431). doi:10.1145/3311927.3323149

Sonnenwald Apps. (2013). ABC Deutsch lernen Grundschule [ABC – Learning German in Primary School] [Mobile application software]. Sonnenwald Apps. Retrieved from https://play.google.com/store/apps/details?id=kellerm.org.school.writer

Tomlinson, D., von Baeyer, C. L., Stinson, J. N., & Sung, L. (2010). A systematic review of faces scales for the self-report of pain intensity in children. *PEDIATRICS*, *126*(5), e1168–e1198. doi:10.1542/peds.2010-1609

Vatavu, R.-D., Cramariuc, G., & Schipor, d. M. (2015). Touch interaction for children aged 3 to 6 years: Experimental findings and relationship to motor skills. *International Journal of Human-Computer Studies*, *74*, 54–76. doi:10.1016/j.ijhcs.2014.10.007

Ward, M. (2014). HCI requirements for young primary school CALL learners. In P. Zaphiris & A. Ioannou (Eds.), *Learning and collaboration technologies. technology-rich environments for learning and collaboration* (Vol. 8524, pp. 339–348). doi:10.1007/978-3-319-07485-6_34

A. Fitted Linear Mixed-Effects Models

The fitted linear mixed-effects models are listed on the next page. Contrasts were obtained from the table of differences of least square means provided by the *lmerTest* package (Kuznetsova et al., 2017).

Table A1. Estimates for the children's ratings of the workload subscales of the adapted kids version of the NASA-TLX. Contrasts were obtained from the table of differences of least square means provided by the *lmerTest* package (Kuznetsova et al., 2017).

	Mental Demand				**Physical Demand**				**Frustration**				**Effort**			
Predictors	β	*SE*	*t*	*p*	β	*SE*	*t*	*p*	β	*SE*	*t*	*p*	β	*SE*	*t*	*p*
(Intercept)	22.65	5.70	3.98	**0.001**	30.58	9.66	3.17	**0.004**	8.05	6.70	1.20	0.239	13.56	6.94	1.96	0.060
IS. Touch — Point-and-Touch	-19.38	4.12	-4.70	**<0.001**	-19.79	4.84	-4.09	**<0.001**	-11.46	4.65	-2.46	**0.018**	-15.42	5.18	-2.98	**0.005**
IS. Touch — Drag-and-Drop	-7.08	4.12	-1.72	0.093	-19.17	4.84	-3.96	< 0.001	-9.17	4.65	-1.97	0.055	-14.37	5.18	-2.77	**0.008**
IS. Drag-and-Drop — Point-and-Touch	-12.29	4.12	-2.98	**0.005**	-0.63	4.84	-0.13	0.898	-2.29	4.65	-0.49	0.624	-1.04	5.18	-0.20	0.842
Spelling. DD — TD	9.73	5.16	1.89	0.073	14.28	9.21	1.55	0.136	1.15	6.12	0.19	0.852	3.76	6.24	0.60	0.553
Grade. 3rd — 4th	3.69	4.88	0.76	0.458	15.04	8.72	1.72	0.099	4.22	5.79	0.73	0.474	4.26	5.90	0.72	0.479
Random Effects																
σ^2	204.06				280.96				259.47				332.18.4			
τ_{00}	71.95_{Subject}				352.75_{Subject}				110.37_{Subject}				97.06_{Subject}			
ICC	0.26				0.56				0.30				0.23			
N	24_{Subject}				24_{Subject}				24_{Subject}				24_{Subject}			
Observations	72				72				72				72			
Marginal R^2 / Conditional ^{2}	0.241 / 0.439				0.281 / 0.653				0.074 / 0.350				0.120 / 0.334			

Table A2. Estimates for writing times and children's ratings of ease of use, fun, and speed. Contrasts were obtained from the table of differences of least square means provided by the *lmerTest* package (Kuznetsova et al., 2017).

	Ease of Use				**Fun**				**Speed**				**Writing Time**			
Predictors	β	*SE*	*t*	*p*	β	*SE*	*t*	*p*	β	*SE*	*t*	*p*	β	*SE*	*t*	*p*
(Intercept)	4.56	0.26	17.72	**<0.001**	4.35	0.28	15.24	**<0.001**	4.25	0.24	18.02	**<0.001**	13.55	1.42	9.53	**<0.001**
IS. Touch — Point-and-Touch	0.36	0.17	2.17	**0.035**	0.40	0.22	1.81	0.077	0.52	0.18	2.86	**0.006**	-4.81	0.48	-10.14	**<0.001**
IS. Touch — Drag-and-Drop	0.44	0.17	2.65	**0.011**	0.48	0.22	2.17	**0.035**	0.40	0.18	2.12	**0.032**	-5.26	0.48	-11.00	**<0.001**
IS. Drag-and-Drop — Point-and-Touch	-0.08	0.17	-0.48	0.633	-0.08	0.21	-0.36	0.719	0.12	0.18	0.66	0.510	0.45	0.48	0.95	0.341
Spelling. DD — TD	-0.02	0.24	-0.08	0.940	0.04	0.25	0.15	0.882	0.19	0.22	0.88	0.386	3.54	1.22	2.91	**0.008**
Grade. 3rd — 4th	-0.13	0.22	-0.60	0.556	-0.38	0.23	-1.63	0.118	-0.22	0.20	-1.12	0.275	3.22	1.15	2.81	**0.011**
Random Effects																
σ^2	0.35				0.61				0.41				37.57			
τ_{00}	0.19_{Subject}				0.13_{Subject}				0.10_{Subject}				6.85_{Subject} 20.38_{Item}			
ICC	0.35				0.17				0.20				0.42			
N	25_{Subject}				25_{Subject}				25_{Subject}				24_{Subject} 45_{Item}			
Observations	75				75				75				1011			
Marginal R^2 / Conditional R^2	0.072 / 0.397				0.100 / 0.253				0.125 / 0.298				0.138 / 0.500			

A Digital Game-Based Training Improves Spelling in German Primary School Children – A Randomized Controlled Field Trial

Heiko Holz[a,b,*], Manuel Ninaus[b,c], Benedikt Beuttler[d], Katharina Brandelik[e], Detmar Meurers[a,b]

Abstract

Digital tools have shown great promise to support reading and spelling development of children, specifically of those suffering from learning disorders such as dyslexia. However, more research is needed on the evaluation of digital game-based trainings carried out in the home environment. In the present study, we investigated the feasibility, effectiveness, and validity of a novel digital game-based spelling training. The training is designed to be used unassisted at home and differs from similar approaches in that it systematically teaches orthographic knowledge in combination with the awareness of syllable stress. 116 German second- to fourth-grade children with mainly poor spelling skills participated in a randomized controlled field trial with a two-period, wait-list controlled crossover treatment design in which children from the immediate treatment group ($N = 58$) received the training during the first training period and the delayed treatment group ($N = 58$) during the second, while the training groups served as control in the opposite training periods. In the active training condition, children practiced at home over a short period of nine to ten weeks. Results showed significant training effects on syllable stress awareness and spelling abilities in trained and untrained domains. The training was also found to be easy to use, motivating, and provided high game experience, proving its feasibility for the use in the home environment. Lastly, we confirmed the validity of our novel pedagogical approach in correlation analyses investigating the relationship between syllable stress awareness, reading, spelling, and training performances. Thus, the training may expand the traditional pool of training methods.

Keywords

game-based learning, elementary education, dyslexia, spelling, special needs education

[a] *ICALL-Research.com Group, Department of Linguistics, University of Tübingen*
[b] *LEAD Graduate School & Research Network, University of Tübingen*
[c] *Department of Psychology, University of Insbruck, Austria*
[d] *Department of Psychology, University of Tübingen, Germany*
[e] *Tübinger Institut für Lerntherapie GmbH, Germany*
*****Corresponding author**: heiko.holz@uni-tuebingen.de

1. Introduction

Reading and writing belong to the most important skills acquired by young learners. Unfortunately, approximately 4–10% of German children do not master these skills adequately and suffer from developmental dyslexia (Katusic, Colligan, Barbaresi, Schaid, & Jacobsen, 2001; Moll, Kunze, Neuhoff, Bruder, & Schulte-Körne, 2014; Moll & Landerl, 2009), which constitutes one of the most frequent learning disorders. Compared to their classmates, dyslexic children acquire reading and writing skills in a much slower pace and not as proficient (Schulte-Körne & Remschmidt, 2003) and suffer massively from their impaired literacy acquisition. If reading and spelling disorders are not diagnosed and treated adequately, they negatively affect children's academic (Daniel et al., 2006), personal (Schulte-Körne, 2010), and social development (Beddington et al., 2008) in the short and long run. Thus, appropriate interventions are indispensable to support reading and spelling development of affected children as early as possible in order to counteract the negative consequences, and to improve their future prospects (Galuschka & Schulte-Körne, 2016). The effectiveness of traditional teaching methods to improve literacy skills applied in standard classroom or individual learning therapy is widely proven and much is known on effective treatment components of spelling disorders (cf. Galuschka & Schulte-Körne, 2016). In addition to traditional learning therapy, digital reading and spelling trainings, that can be used during or outside of class, have shown great promise to support children's literacy acquisition (cf. Holz, Brandelik, Beuttler, Brandelik, & Ninaus, 2018). However, more empirical research is needed in order to evaluate the benefits of digital (game-based) spelling trainings for German (dyslexic) primary school children in the home environment (cf. Holz, Brandelik, et al., 2018).

In order to extend the current state of research on digital spelling trainings, we present and evaluate an innovative mobile game-based spelling training for German primary school children. The training program differs from similar approaches in that it focuses on teaching orthographic regularities of German orthography in combination with the awareness of syllable stress, and combines the educational approach with foundations of digital game-based learning.

1.1 The Benefits of Digital Game-Based Interventions

In the following, we elaborate on the benefits and disadvantages of therapeutic, computer-based, and digital game-based interventions. We refer the reader to (Holz, Brandelik, et al., 2018) for a more detailed overview of the advantages and disadvantages of the different forms of intervention.

Therapeutic Interventions. Commonly, reading and spelling disorders are treated in therapeutic interventions administered by trained practitioners, such as teachers or learning therapists, in weekly individual or group sessions over several months. Therapeutic interventions are recommended treatments for dyslexic children (Galuschka & Schulte-Körne, 2016) and can reliably improve reading and spelling (e.g., Groth, Hasko, Bruder, Kunze, & Schulte-Körne, 2013; Ise & Schulte-Körne, 2010; Klicpera, Weiss, & Gasteiger-Klicpera, 2013; Reuter-Liehr, 1993; Tacke, 2005) when administered by experts (Galuschka, Ise, Krick, & Schulte-Körne, 2014). However, therapeutic interventions are cost-intensive, time- and location-dependent, and might not be available timely or long enough due to tedious application processes for financing or reimbursement, disadvantaging families who cannot afford to pay for learning therapy privately.

Computer-Based Interventions. In addressing the disadvantages of therapeutic interventions and offering new ways to engage young learners, computer-based interventions have been shown in recent years to successfully complement traditional teaching and learning therapy in improving reading and spelling in German dyslexic children (e.g., Kargl, Purgstaller, Weiss, & Fink, 2008; Kast, Baschera, Gross, Jäncke, & Meyer, 2011; Klatte, Bergström, Steinbrink, Konerding, & Lachmann, 2018). More generally, computer-based intervention have been shown to facilitate literacy acquisition in dyslexic children (e.g., Cidrim & Madeiro, 2017; Drigas & Batziaka, 2016). Moreover, computer-based interventions are independent of time and place and can automatically adapt the learning content to the specific needs of individual children. This is necessary for dyslexics who have heterogeneous difficulties in different levels of literacy acquisition (Rose, 2009). While computer instructions may be equally effective as human tutors (e.g., in handwriting and spelling cf. Berninger, Nagy, Tanimoto, Thompson, & Abbott, 2015), children have shown to concentrate better while engaged with computer-based interventions than in traditional school tasks (Ronimus, Kujala, Tolvanen, & Lyytinen, 2014). Additionally, interactive experiences motivate young learners and help to attenuate their daily struggles in literacy acquisition (Cidrim & Madeiro, 2017). Furthermore, gamification, i.e., *"the use of game design elements in non-game contexts"* (Deterding, Dixon, Khaled, & Nacke, 2011), mostly positively affects learning and increases motivation, engagement in, and enjoyment of learning tasks (Hamari, Koivisto, & Sarsa, 2014).

Digital Game-Based Interventions. Digital game-based trainings, which are also referred to as serious or educational games, take it to the next level. While gamified computer-based interventions merely incorporate elements of games to existing tasks that may be unengaging, tedious, or boring (Plass, Homer, & Kinzer, 2015), game-based interventions are designed as full-fledged games for educational purposes (Deterding et al., 2011) that focus on designing activities as playful tasks (Plass et al., 2015). Research on digital game-based learning has become more popular in recent years (for an overview see Boyle et al., 2016; Hainey, Connolly, Boyle, Wilson, & Razak, 2016) and it has been shown to be effective or even outperform conventional instruction methods, especially for language learning (Wouters & van Oostendorp, 2013). Specifically for learning disorders, educational games have proven to support children with dyslexia or dyscalculia (e.g., Abrami, Borohkovski, & Lysenko, 2015; Ninaus, Kiili, McMullen, & Moeller, 2016), and, most importantly for this article, the acquisition of reading and spelling in German dyslexics (e.g., Berkling, 2017; Görgen, Huemer, Schulte-Körne, & Moll, 2020; Lenhard & Lenhard, 2016). Game elements embedded in digital game-based interventions, such as feedback, reward, or narratives, influence learning positively (Wouters & van Oostendorp, 2013) and play a crucial role to achieve learning goals (Boyle et al., 2016). They address negative feelings, such as frustration, demotivation, or boredom (Deterding et al., 2011), promote engagement and learning for children with special needs (Ke & Abras, 2013), and may even reengage learners who disengaged from learning, i.e., learners who lost interest, motivation, and engagement in learning and cannot be engaged with other methods (Griffiths, 2002; Squire, 2008).

1.1.1 State of Research on the Effectiveness of Treatment Approaches for Spelling Disorders

Research on the spelling remediation in German dyslexics was mainly done with weekly therapeutic interventions administered by experts (e.g., Ise & Schulte-Körne, 2010; Reuter-Liehr, 1993; Schulte-Körne & Mathwig, 2013; Tacke, 2005) or instructed parents (e.g., Schulte-Körne, Deimel, & Remschmidt, 1998; Schulte-Körne, Schäfer, Deimel, & Remschmidt, 1997); or with digital (game-based) interventions in daily to weekly supervised training sessions during school lessons (Kargl et al., 2008; Klatte et al., 2018) or after school (e.g., Berkling, 2017), sometimes with additional training at home (Kargl et al., 2008). However, randomized controlled field trials (RCFT) on the effectiveness of computer-based treatment approaches, i.e., when the training is carried out in the home environment without adult help – under "real-world" conditions – are missing in clinical practical guidelines (cf. Galuschka & Schulte-Körne, 2016) and meta reviews (cf. Galuschka et al., 2014; Ise, Engel, & Schulte-Körne, 2012; McArthur et al., 2012). While Görgen et al. (2020) could recently show in a RCFT that their digital game-based reading training carried out in the home environment can significantly improve reading abilities for trained word material in German children with reading disorders, we are not aware of such RCFTs on digital spelling trainings.

1.2 The Role of Syllable Stress in Literacy Acquisition

According to current research, dyslexia is not caused by a single factor, but rather is influenced by myriad factors, including genetic disposition, socioeconomic factors, cognitive functions, and the perception and processing of visual and acoustic information (Schulte-Körne & Remschmidt, 2003). In this regard, the phonological deficit theory is the most well-developed and evidence-based theory that sees a causal role of phonological skills in children's development of reading and spelling (cf. Ramus, 2003; Snowling, 2001) – children with good phonological skills become good readers and good spellers, while children with poor phonological skills progress more poorly (cf. Goswami, 1999). As such, a deficient phonological awareness – the ability to deal with the sound system of a language and to detect, distinguish, and manipulate segments of a language (Klicpera, Schabmann, & Gasteiger-Klicpera, 2013) – is known as one major cause of dyslexia (Bradley & Bryant, 1983; Snowling, 1995).

Phonological awareness also includes the perception and processing of prosodic features. A shortcoming in the perception of prosodic features is a strong predictor for dyslexia (Goswami et al., 2013; Leong, Hämäläinen, Soltész, & Goswami, 2011; Sauter, Heller, & Landerl, 2012). One of these features is syllable stress, an important characteristic of German speech rhythm. In stress-timed languages, such as German (Kohler, 1986), English, Russian, or other Germanic languages, speech rhythm is generated by the regular appearance of stressed syllables, whereby the intervals between stressed syllables tend to have a constant duration of approximately 500 milliseconds (Arvaniti, 2009; Pompino-Marschall, 2009). Stressed syllables are on average louder, longer (Jessen, Marasek, Schneider, & Claßen, 1995), and oftentimes higher in pitch than unstressed syllables and the rise time (the time required to reach peak signal intensity) is shorter (Thomson & Jarmulowicz, 2016) – the vowel sound of the stressed syllable gets loud faster (Pompino-Marschall, 2009). In contrast, unstressed syllables are compressed and reduced to fit the rhythm.

Recent empirical findings have shown that the perception of stress is impaired in dyslexic children (Goswami et al., 2013; Jiménez-Fernández, Gutiérrez-Palma, & Defior, 2015; Leong et al., 2011), and that syllable stress awareness is highly correlated with reading and spelling skills (Sauter et al., 2012).

For German dyslexics, one explanation is thought to be found in the association between stress and German orthographic markers. Orthographic markers, i.e., graphemes marking long and short vowels, generally occur in stressed syllables (markers for long vowels, such as the bigram *ie* in *BIE-ne* [bee]) or in conjunction with stressed syllables (markers for short vowels, such as the ambisyllabic consonant doubling *tt* in *Ge-WIT-ter* [thunderstorm]) (Staffeldt, 2010; Vennemann, 2011).

Mastering the complex orthographic rules to mark long and short vowels is a major difficulty for German children (Klicpera & Gasteiger Klicpera, 2000; Landerl, 2003). The phonological origin of orthographic markers lies in the basic form of the German trochee – the German disyllabic standard word in which the first syllable is stressed and the second syllable is unstressed (e.g., *FAL-len* [to fall], *REN-nen* [to run], *FEL-sen* [rock], *SE-geln* [to sail]).

Therefore, processing verbal stress adequately may help children to acquire the complex spelling rules that underlie vowel length spelling in German orthography. Further, rhythmic trainings that contain exercises to match the correct syllabic stress pattern to words have been shown to be beneficial for the development of literacy and phonological awareness of English poor readers (e.g., Bhide, Power, & Goswami, 2013; Thomson, Leong, & Goswami, 2013). Yet, syllable stress awareness has not been included comprehensively in digital spelling trainings for German.

To conclude, besides morphological skills, lexical knowledge and knowledge of spelling rules (Galuschka & Schulte-Körne, 2016; Ise & Schulte-Körne, 2010; Schulte-Körne & Mathwig, 2013), syllable stress awareness may play a role in the orthographic stage of spelling acquisition, particularly in the spelling of long and short vowels (Sauter et al., 2012).

1.3 Aims of the Present Study

This study aims to fill the research gaps in (i) digital game-based spelling trainings carried out at home and in (ii) digital training approaches that combine syllable stress awareness with spelling. For this, we present and evaluate a novel digital game-based spelling training for German primary school children for the use at home. Importantly, the training teaches orthographic knowledge and spelling rules in combination with the awareness of syllable stress. It systematically trains children's awareness and their analytical skills on the syllable level. In doing so, it is in line with the clinical practical guideline on the treatment of reading and/or spelling disorders (Galuschka & Schulte-Körne, 2016), which concludes that spelling can most effectively be improved by using systematic instructions of sound-letter correspondences, exercises analyzing sounds, syllables, and morphemes, as well as trainings enabling the acquisition and generalization of orthographic regularities.

The main purpose of the current study is to assess the feasibility, efficacy, and validity of the training program. For this, we carried out a randomized controlled field trial with a two-period, wait-list controlled crossover treatment design in which 116 German primary school played the game at home during a period of 9–10 weeks. The evaluation in the present study addresses in total three hypotheses explained in the following.

1.3.1 Hypotheses

Feasibility. Feasibility is a major design principle of digital interventions that aim at supporting children with special educational needs in the home environment without the need of adult help. For this, the interventions have to ensure that children are able to complete the training on their own and that they are motivated and engaged over longer time to maximize learning. We therefore embedded the proposed training in a digital game-based learning environment and aimed to ensure the game's feasibility with the use of different components, such as interactive instructions, immediate feedback, or rewards. As a result, we expect that the training program can successfully be used in the home environment as a supplementary tool to support literacy acquisition in (dyslexic) primary school children, i.e., that it can be used by the children without adult help and that it engages and motivates young learners over several weeks (*Hypothesis 1*).

Efficacy. Based on the empirical evidence and linguistic background of syllable stress awareness in literacy acquisition, we expect that the training program has a positive effect on literacy skills. Particularly, due to its focus on syllable stress and spelling rules, we expect that the training improves children's syllable stress awareness (*Hypothesis 2a*) and spelling (*Hypothesis 2b*). Further, we investigate the training's impact on untrained reading skills that are related to phonological awareness (*Hypothesis 2c*).

Validity. The validity of digital interventions does not only concern a theoretically sound pedagogical approach, but also that the educational content is

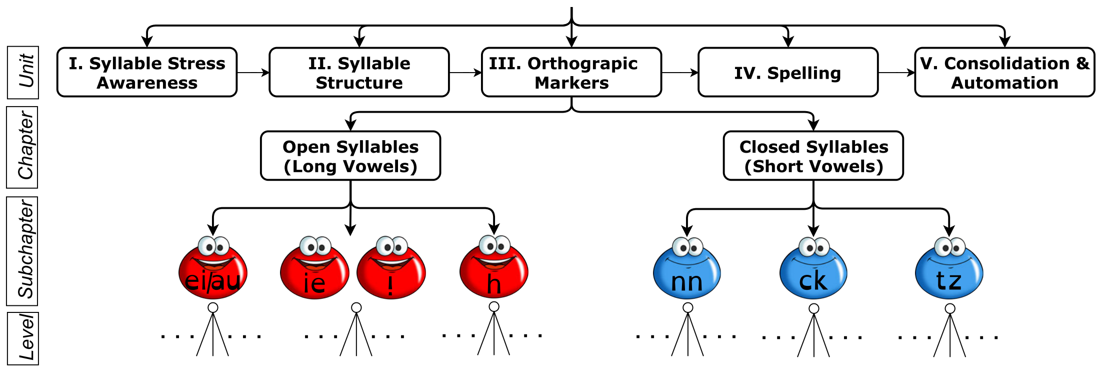


Figure 1. Overview of the pedagogical structure of the present version of Prosodiya. The game increases in complexity and difficulty on four levels at individual rates: units, chapters, subchapters, and levels.

implemented effectively. As to the validity of our approach, we expect that syllable stress awareness is correlated with reading and spelling skills in the present study (*Hypothesis 3a*) and we expect to find associations between the different exercises implemented in the training and real-life literacy skills (*Hypothesis 3b*), using learner analytics obtained from game logs.

2. Proposed Mobile Game-Based Spelling Training

In the following, we briefly explain our game-based spelling training. More detailed information on the linguistic background of the spelling training and its game design and pedagogical content is provided in Appendix 5. Videos demonstrating the training program and highlighting different aspects can be accessed at `https://prosodiya.com`.

"Prosodiya" is as an adaptive digital game-based spelling training for mobile touch devices that primarily aims at improving syllable stress awareness, the awareness of linguistic features related to syllable stress, and ultimately spelling abilities in German primary school children. The training is based on recent empirical findings and is to some extent similar to evidence-based rule-based spelling interventions (e.g., Ise & Schulte-Körne, 2010; Reuter-Liehr, 1993). It differs from similar empirically evaluated approaches in that it focuses on syllable stress awareness and on linking the linguistic features related to syllable stress to orthographic regularities of German orthography. These abilities play a special role in literacy acquisition and are specifically impaired in dyslexic children (cf. Section 1.2). This is where the training comes in. The training shifts the children's attention to relevant areas of words to clarify the association between syllable stress and orthographic marking of long and short vowels, and teaches the children how such syllables are spelled. In doing so, it ultimately leads to a rule-based orthographic spelling training.

Educational Content. The training is divided in five curriculum units that focus on syllable stress awareness, syllable segmentation, vowel length distinction, orthographic vowel length marking, and spelling. Exemplary games of the training are displayed in Figure 2 for the word *rennen* (/ˈʁɛnən/, to run), whose short vowel phoneme /ɛ/ is marked orthographically with the ambisyllabic consonant doubling *nn*.

In the first unit, children learn to identify the stress pattern of words and to segment words into syllables, see Figure 2a. In the second unit, children learn to distinguish vowel lengths and to identify open (ends with a vowel) and closed (ends with a consonant) syllables, see Figure 2b. In the third unit, children learn how open and closed syllables are spelled by teaching them the rules of orthographic marking of long and short vowels, see Figure 2c. In the fourth unit, children consolidate their previously acquired knowledge by spelling the words, see Figure 2d. The fifth unit aims at consolidating children's linguistic knowledge by practicing with all games in medium or hard difficulty to automate reading and spelling processes. Each part of the game starts with easy exercises and continuously increases in difficulty.

Narrative. The game's overall narrative revolves around little inhabitants called "*Kugellichter*" ["spherical lights"], which seek the children's help: A mysterious fog is haunting their homeland which causes the inhabitants to live in worries and sorrow, see Figure 4b. As the inhabitants are too weakened to dispel the fog on their own, the chil-

(a) Game 1: "stress pattern". Children identify stress pattern by placing the *Kugellichter* on respective platforms. The big green blob is used for stressed syllables, the small yellow blob for unstressed syllables.

(b) Game 2: "open and closed syllables" – or "vowel length distinction". Similar to the first game, children rebuild stress patterns of words but additionally need to distinguish whether the vowel of the stressed syllable is long (red blob with open mouth) or short (blue blob with closed mouth).

(c) Game 3: "orthographic markers". Children select the correct orthographic marker for the vowel of the stressed syllable.

(d) Game 4: "spelling". Children arrange letters from a predefined set in the spelling line to write words.

Figure 2. Games teaching the orthographic marking of long and short vowels and spelling of words.

dren are their last hope. Only they, accompanied by the Kugellichter through the world of syllables and orthography, can free the land from its dreadful destiny by mastering linguistic challenges. For this, they need to understand and use the "*power of the stressed syllable*" in order to obtain the "*wisdom of words*". Progressing through the course of the game, parts of land are saved and new regions await the children with challenges to be mastered.

Feasibility. To ensure that the training is feasible for the unassisted use at home, we implemented interactive tutorials and automatic feedback. The highly interactive tutorials teach game mechanics and linguistic knowledge of each featured game and linguistic characteristic. In addition, we also implemented short and spot-on task explanations, so-called tooltips, that appear at the start of each level and that can be accessed manually during play. Exemplary tutorials and tooltips are listed in Appendix A.3.2.

3. Methods

3.1 Design

To evaluate the feasibility, efficacy, and validity of the training, a two-period, wait-list controlled crossover treatment design was used, with participants randomized to the immediate treatment group (ITG) or to the delayed treatment group (DTG), see Figure 3. Pretests were conducted in February 2018 (T1) after which participants from the ITG performed 9–10 weeks of training. Midtests were conducted in May 2018 (T2) after which the training from the ITG was discontinued and participants from the DTG were crossed to the active training and performed 9–10 weeks of training. Posttests were conducted in July 2018 (T3). Test sessions were administered in classrooms of participating partner schools and learning institutions or in facilities of the university. A test session was as follows: first, classroom tests of spelling and reading fluency were administered

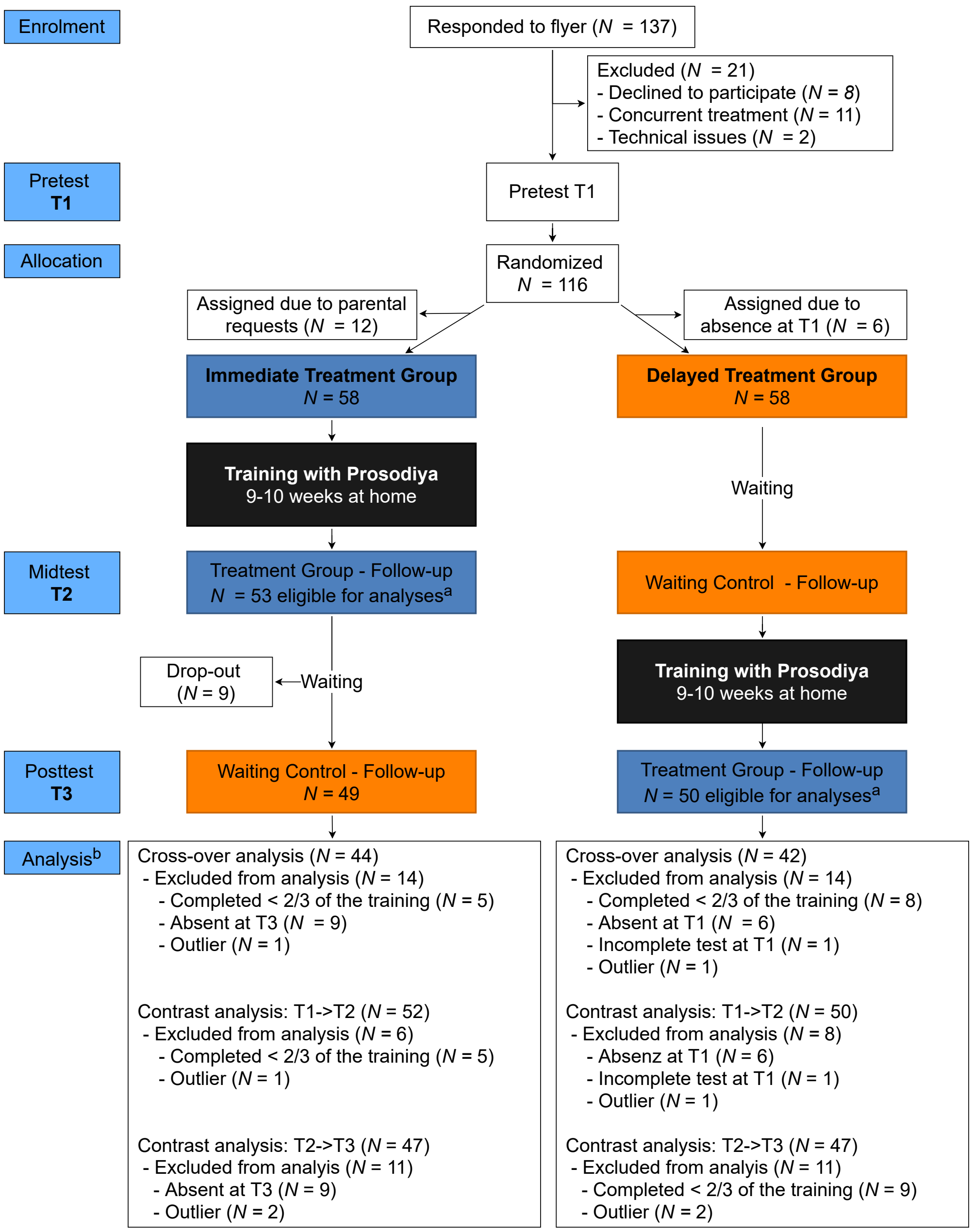


Figure 3. Flow diagram of the two-period, wait-list controlled crossover treatment design for the randomized controlled field trial of the training program.
Note. [a] Only children from the active training group who completed at least two-thirds of the training program were included in the analyses of respective training periods.
[b] Analysis of the standardized spelling scores assessed with the DRT.

in groups, followed by individually administered assessments of syllable stress awareness and word reading. At T2 and T3, training experience questionnaires were answered after the spelling tests by children from the active training group.

3.2 Participants

We recruited primary school children from second to fourth grade at the age of 7–11 years via learning institutions, the youth welfare office, newspaper advertisement, and eight public primary schools in the area of Tübingen, Germany. Flyer were sent to the institutions and we asked learning therapists, teachers, and employees of the youth welfare office to pass these to parents of poor spellers. In total, 137 families responded to the flyer of which eight dropped out before the study had started. Of the remaining 129 participants, we excluded thirteen children in the efficacy analyses, eleven children who received concurrent reading or spelling remediation and two children due to technical issues during training. Children not meeting the inclusion criteria were excluded

Table 1. Descriptive data of the treatment groups (ITG = immediate treatment; DTG = delayed treatment).

Variables	ITG ($N = 58$)		DTG ($N = 58$)			
	M (SD)	*Range*	M (SD)	*Range*	t	p
Age in years	8.9 (0.9)	7.5–10.6	8.8 (1.0)	7.3–11.0	0.75	.45
Spelling[a]	37.7 (8.3)	23.5–57.9	41.6 (8.1)[e]	25.8–61.8	−2.24	.03*
Reading fluency[b]	81.9 (14.0)[f]	62–113	86.5 (13.8)[g]	64–127	−1.22	.23
Word reading[c]	18.7 (22.7)[h]	1–82	22.0 (23.7)[i]	2–90	−0.69	.49
Syll. stress awareness[d]	7.4 (3.1)	1–13	7.9 (3.2)	1–14	−0.52	.60
	Frequencies				χ^2	p
Boys/girls	35/23		30/28		0.56	.45
Diagnosed dyslexics[j]	25/33		8/50		10.84	$< .001$*
Grade 2/grade 3/grade 4	23/24/11		27/20/11		0.68	.71

[a] Spelling (DRT): T-scores, $M = 50$, $SD = 10$.
[b] Reading fluency (SLS 2–9): LQ-scores, $M = 100$, $SD = 15$.
[c] Word reading (SLRT-II): percentile ranks.
[d] Syllable stress awareness (self-designed test): raw scores (max = 15).
[e] Based on $n = 51$ children from the DTG present at T1.
[f] Based on $n = 43$ children from the ITG present at T1 whose test met the inclusion criteria.
[g] Based on $n = 46$ children from the DTG present at T1 whose test met the inclusion criteria.
[h] Based on $n = 48$ children from the ITG present at T1 whose test met the inclusion criteria.
[i] Based on $n = 46$ children from the DTG present at T1 whose test met the inclusion criteria.
[j] External diagnosis.
[*] Significant difference between the two treatment groups.

from the analyses but were still allowed to complete the training.

The final sample for the efficacy analyses is listed in Table 1 and includes 116 children (65 boys and 51 girls), aged between 7–11 years ($M = 8.85$, $SD = 0.93$). Of the eligible 116 children, 58 children were assigned to the ITG and 58 to DTG. The assignment was mainly done randomly based on spelling and reading abilities assessed at T1. A full randomization of the participants was not possible due to ethical reasons and real-life circumstances of a field trial. Twelve parents of dyslexic children were not willing to participate in the study if their child would be assigned to the DTG and thus were assigned to the ITG. Three children, whose parents had contacted us just before the start of the first training period, and four children who were sick at T1 were allocated to the DTG. Nine children from the ITG did not participate at T3 because they continued with a spelling remediation after T2 or were sick at T3. The flow diagram of the present study including participant selection is depicted in Figure 3.

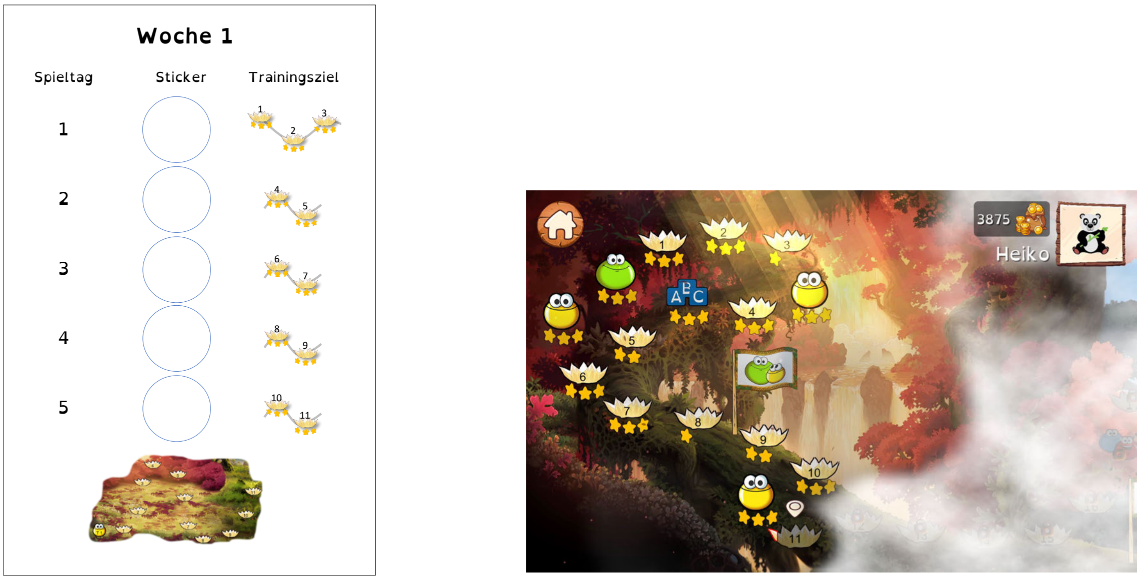
As listed in Table 1, the reading and spelling abilities of the participants were significantly below average and ranged between very poor and below average, with very few exceptions of average performance.

For efficacy analyses, we only included children from the active training group who completed at least two-thirds of the training program (ITG during the first training period and DTG during the second) or who served as the control (vice versa). The first two-thirds of the training cover the acquisition phase. Children acquire new skills and learn to use their new knowledge. The last third covers a training and automation phase. Analyses including only participants that completed the whole training yield the same test decisions with the drawback of smaller sample sizes and less ecological validity.

3.3 Materials

3.3.1 Game and Training Plan

The mobile game described in Section 2 (for more details see Appendix 5) was used. For the present study, we excluded the subchapter on the “silent h” since words that feature a silent h are exceptions in terms of orthographic marking. They do not follow explicit rules and must be memorized and learned by heart with memos such as “*das stumme h, das ist nicht schwer, steht meist vor l, m, n, und r*” [the silent h precedes mainly but not necessarily the letters l, m, n, and r after a long vowel phoneme]. Due to the brevity of the present study (training period: 8–10 weeks),



(a) One week of the training plan in the sticker book.

(b) Corresponding in-game map of the training. Glass blossoms are used as level symbols.

Figure 4. Training plan depicting what should be trained when to keep children on track and to engage them to complete their training.

we focused on the more consistent orthographic marking of long and short vowels. Further, the study version did not include capitalization rules. In spelling games, the available letters were displayed in lower- and uppercase, depending whether a noun was practiced or not, and the case could not be changed. For example, the available letters to spell the word *rennen* [to run] were all lowercase, whereas the available letters for the word *Biene* [bee] contained both lowercase and uppercase letters, e.g., a possible set of letters, including distractors, would be $\{B,n,i,n,e,P,h,ä\}$.

Schedule. During respective training periods, families were given Android tablets and children were asked to train at home five days per week twenty minutes each, following a training plan of eight weeks, see Figure 4. The training plan was given in the form of a sticker book with a set of 40 stickers to keep the children on track and to engage them to complete their training. The sticker book depicts for each training day and week the levels to be practiced, see Figure 4a. Each page contains one training week and corresponds to the map used in the game, see Figure 4b.

Due to school holidays during training, more levels than included in the sticker book were deployed in the game. In total, 80 levels were deployed. The training was officially completed at level 66, labeling the remaining levels as bonus. In each level, ten words were practiced. Depending on the levels' configuration and children's performances, the same levels may have to be practiced more than once. To avoid binge-playing and loss of training effect, content of a new training week was unlocked on Monday mornings.

3.4 Measures

3.4.1 Feasibility

To evaluate the feasibility of the training in the home environment, we examined the training behavior of children obtained from in-game data and logs as well as feedback from children and parents collected with questionnaires. The questionnaires were answered from the active training group after their respective training period. As the detailed evaluation of the questionnaires on game experience, usability, self-efficacy, and individual game elements are beyond the scope of this article, we refer the reader to Holz, Beuttler, and Ninaus (2018) and to Holz, Ninaus, Meurers, and Kirsch (2018) for detailed description and evaluation of these measures.

3.4.2 Efficacy

To evaluate the efficacy of the training, we examined the effect on trained literacy skills, i.e., syllable stress awareness and spelling. For spelling, we analyzed the general spelling ability as well as specific orthographic learning categories. Additionally, we examined transfer effects on untrained reading skills. In the following paragraph, we describe each measure in detail.

Syllable Stress Awareness. Syllable stress awareness was assessed using an individually administered paper version of the game “stress pattern” (Figure 2a), in which children had to identify the stress pattern of 15 trained words using printed versions of the *Kugellichter* for stressed and unstressed syllables. Each word was individually read out and displayed as a picture in a PowerPoint presentation. Scoring is based on the number of correctly identified stress patterns. Parallel test versions were used alternately in the test sessions, i.e., test version A at T1, test version B at T2, and test version C at T3.

Spelling. Spelling ability was assessed with a standardized classroom cloze spelling tests (*DRT 2/3/4, Diagnostischer Rechtschreibtest für 2./3./4. Klassen* [Diagnostic Spelling Test for 2nd/3rd/4th Grade]; Grund, Leonhart, & Nauman, 2017; Müller, 2003a, 2003b) in which children had to fill 32/44/42 dictated words (for grade 2/3/4, respectively) into sentence frames. The experimenter first read aloud the word to be spelled, then the full sentence and finally repeated the word to be spelled. Scoring is based on the number of correctly spelled words. Norm-referenced scores are standard scores (T-scores) with a mean of 50 and standard deviation of 10. For the analyses, we used the standard score for the entire test as well as the raw score of the number of misspelled words whose spelling mistakes violated the rules of the orthographic marking of short and long vowels (error category D in Müller, 2003a; *spelling mistakes in vowel length marking*).[1] Raw scores for the latter error category were used since standardized scores of this error category are not available in the DRT 4. Parallel test versions were used alternately to avoid testing-induced effects, i.e., test version A was used at T1 and T3 and test version B at T2.

We additionally administered a self-designed cloze spelling test at T2 and T3 in which children had to fill 30 dictated words into sentence frames that was administered similarly to the standardized spelling test. The items were the same for all grades, allowing to further investigate transfer of learning as the DRT contains grade-specific items that are not shared across all grades, making it hard to derive transfer effects independent of grade. The spelling test specifically addressed training-specific orthographic regularities and covered three explicit learning categories: (i) nine uninflected words that are part of the training (no transfer of learning), such as *fliegen* [to fly], for which no transfer of learning is required; (ii) ten uninflected words that are not part of the training but that have similar orthographic syllable structures (near transfer of learning), such as the word *stinken* [to stink], which is orthographically very similar to the training word *blinken* [to flash]; and (iii) eleven inflected words whose basic form is exposed in the training (far transfer learning), e.g., *rennt* [he runs], whose basic form *rennen* was included in the training. We consider the second category as near transfer learning since children must apply learned rules to unseen uninflected words. The third category is considered as far transfer of learning as it requires the children to apply the orthographic rules to the base form *rennen* of the word and not to the inflected form *rennt*. This morphological (word building) skill was not trained in the intervention. Scoring of the test is based on the number of correctly spelled words. Mistakes in upper- and lowercase were not counted as the primary goal of this test was to investigate the effect of the training on the spelling categories included in the training. For the analyses, we used in total five raw scores: the (i) raw score of the entire test, (ii) the raw score of misspelled words whose spelling mistakes violated the rules of the orthographic marking of long and short vowels, as well as the raw score of the three learning categories, i.e., (iii) uninflected training words, (iv) uninflected untrained words, and (v) inflected training words. Parallel test versions were used at T2 and T3.

Reading Fluency and Word Reading. Reading fluency was assessed with a standardized classroom reading test (*SLS 2–9, Salzburger Lese-Screening für die Schulstufen 2–9* [Salzburg Reading Screening for Grades 2–9]; Mayringer & Wimmer, 2014) in which children read as many sentences as possible in three minutes and mark them as either true or false (e.g., “you can drink water” is true while “strawberries can speak” is false). Scoring is based on the number of correctly marked sentences. Norm-referenced scores are standard reading scores (reading quotient, LQ-scores) with a mean of 100 and standard deviation of 15. The norm table of the handbook is limited to LQ-scores in the range between 62 and 138. For the analyses, we used the standard reading score. Parallel test versions were used alternately in the test sessions, i.e.,test version A at T1 and T3, test

[1] Transforming the raw scores into grade-specific z-scores lead to the same statistical test decisions.

version B at T3.

Word reading was assessed in a standardized one-minute reading speed test (*SLRT-II: Lese- und Rechtschreibtest* [SLRT-II: Reading and Spelling Test]; Moll & Landerl, 2010) in which children read aloud words as fast as possible without making errors from a reading list. The test contains a word and a pseudoword reading list with increasing word length and complexity. Scoring is based on the number of correctly read words. Norm-reference scores are percentile ranks. For the analyses, we calculated and used z-scores based on the norm sample. Parallel test versions were used alternately in the test sessions, i.e., test version A at T1 and T3, test version B at T3.

3.4.3 Validity

To examine the validity of the pedagogical approach and its implementation, we investigated the relationship between the aforementioned literacy skills and the relationship between literacy skills and training performances. For literacy skills, we used syllable stress awareness, spelling (standardized spelling score, spelling scores of our self-designed spelling test, spelling mistakes in vowel length marking), and reading fluency assessed at posttest T3, after all children received the training. Training performances are average scores and times obtained from in-game data, for which we computed the overall average score and completion time of a level as well as average scores and times per individual game type for each participant.

3.5 Analysis

All analyses were performed using the statistic software R (R Core Team, 2014). Type III sum of squares were used. The criterion of statistical significance was set at $\alpha = .05$.

3.5.1 Efficacy

The training's efficacy (*Hypothesis 2*) was analyzed in a two-step process. First, crossover analyses were performed to investigate if children's learning gains induced by the training is significantly higher than that obtained during waiting periods without extra training. For this, we compared the within-subject differences between the two training periods from the immediate and the delayed treatment group with regard to the outcome variables, following the analysis for two-group two-period crossover trials proposed by Hills and Armitage (1979). For this, we calculated changes in the outcome variables for both training periods (T2 − T1 and T3 − T2, respectively) by group and analyzed the within-subject period differences ([T2 − T1] − [T3 − T2]) in our outcome measures between the ITG and the DTG with two-sample t-tests. This analysis is recommended as the standard approach to investigate treatment effects for two-group two-period crossover trials when controlling for possible time effects (Senn, 2002; Wellek & Blettner, 2012). The crossover analyses included only those 89 children who completed at least two-thirds of the training program and who participated at each of the three test sessions, i.e., 45 children from the ITG and 44 children from the DTG. In case of significant treatment effects, Cohen's d effect sizes based on the pooled standard deviations were calculated. According to Hattie (2008), effect sizes can be considered small if $d = 0.2$, medium if $d = 0.4$, and large if $d = 0.6$ when evaluating educational outcomes.

In the second step, we examined whether a potential training effect is found during the first and/or second training period. For this, we applied planned contrasts to analyze separately changes in the outcome measures from pre- (T1) to mid- (T2) and from mid- (T2) to posttest (T3). Potential group differences in learning gains between T1 and T2 and between T2 and T3 were analyzed by means of one-way ANCOVAs, comparing group effects on gain scores of the outcome variables at T2 and at T3 with the pretest scores of respective tests of the respective training period (T1 for the first training period, T2 for the second), diagnosis of dyslexia, sex, and grade treated as covariates.[2] In case of significant group effects, we estimated between-group effect sizes $\hat{d}$ separately for the learning gain using the adjusted mean difference between the active intervention group and the control group divided by the estimated pooled standard deviation obtained from the square root of the mean squared error of the ANCOVA models, i.e., $\hat{d} = \frac{\overline{X'}_{training} - \overline{X'}_{control}}{\sqrt{MSE'}}$ (Grissom & Kim, 2012, p. 349). Estimated marginal means of ANCOVAs were extracted with the *effects* package (Fox & Weisberg, 2019). In case of unequal regression slopes, t-tests on gain scores were performed instead of ANVOCAs. In case of non-normally distributed gain scores, Wilcoxon rank-sum tests were used instead of t-tests.

[2] ANCOVAs on the respective posttest scores instead of the gain scores yield the exact same results (Jamieson, 2004; Zientek, Nimon, & Hammack-Brown, 2016). We opted for the gain scores as responses for illustrative purposes.

The assumptions for the applicability of ANCOVAs (Rausch, Maxwell, & Kelley, 2003) and t-tests were tested statistically. We used Levene's test (median-centered) from the *car* package (Fox & Weisberg, 2019) to test for homogeneity of variances, the Shapiro-Wilk test to test for normality, and testing of the interaction effect of the group assignment and the respective covariate to examine homogeneity of regression slopes.

3.5.2 Validity

We computed partial correlations using the *psych* package (Revelle, 2018) to determine the relationship between the assessed literacy skills (*Hypothesis 3a)* and between literacy skills and training performances (*Hypothesis 3b*) while controlling for sex and grade. We included the data of children who participated at T3 and completed at least two-thirds of the training program. We opted for Spearman's rank correlation due to non-normal distribution of the in-game data.

3.5.3 Exclusion of Participants

We excluded participants from respective analyses due to different reasons. For the sake of readability, we briefly describe the exclusion criteria. Resulting sample sizes for the contrast analyses of each outcome measure are listed in Table 2. We excluded children from respective analyses that were absent at respective testing sessions, children that did not participate in respective tests, and children who did not complete a respective test. Additionally, some children were excluded based on outlier analyses. In the crossover analyses, zero to two participants whose period differences deviated more than 2.5 standard deviations from the mean of the respective training group were excluded as outliers. In the ANCOVA models, zero to three participants whose residuals deviated more than 2.5 standard deviations from the mean of the residuals were excluded from respective analyses (cf. Baayen, 2008, Chapter 7). In t- and Wilcoxon rank-sum tests, one to three participants whose gain scores deviated more than 2.5 standard deviations from the mean were excluded.

Regarding reading fluency, we additionally excluded children whose raw scores were not listed in the norm table of the handbook (see Section 3.4.2), tests in which children continued working on the practice page during the three minutes of the actual test, tests that exceeded the time limit of three minutes due to flawed test administration, and tests of children who conducted more than four mistakes or skipped more than four sentences as we cannot reliably tell whether the lowered raw score reflects low reading fluency or results from poor concentration or lack of motivation. Regarding word and pseudoword reading, we excluded tests for which audio files were missing.

4. Results and Discussion

4.1 Feasibility

Hypothesis 1: Children practiced for about 18.5 minutes ($SD = 7.3$) over 27.9 days ($SD = 10.9$), reached, on average, level 69 ($SD = 18.7$), and practiced in total an average of 161.7 levels ($SD = 48.8$). It took them an average of 3.0 minutes ($SD = 0.8$) to complete a level that featured 10 words and they scored an average of 138.6 ($SD = 5.7$) out of 150 possible points per level, solving an average of 8.2 ($SD = 0.9$) tasks on the first go. The training behavior did not differ significantly between the ITG and the DTG.

Out of the 116 children eligible for the evaluation, 103 children (89%) completed at least two-thirds of the training and 88 children (76%) fulfilled the complete training plan, reaching level 66 or higher. Moreover, the number of children who successfully completed the training is comparable to that obtained in controlled intervention studies in which the training is carried out supervised in controlled settings at schools or learning facilities.

In addition, the training was perceived very positively by children, their parents, and teachers. The children reported that the game was easy to use and that they perceived high self-efficacy after training and a positive influence of the training on their spelling-related abilities. Many families responded that they would likely continue the training or recommend it to others. Furthermore, the children were engaged with the training, considered it more as a game, and liked in particular the pedagogical agents who have accompanied them throughout the training and taught them the linguistic knowledge. We refer the reader to (Holz, Ninaus, et al., 2018) for detailed analysis of training experience and usability, and to (Holz, Beuttler, & Ninaus, 2018) for the detailed evaluation of individual game elements, such as graphics, narrative, and pedagogical agents.

Consistent with *Hypothesis 1*, we may infer that the game-based spelling training is feasible as an intervention at home and that the results of the presented study may reflect real-life effectiveness whose indications go beyond controlled settings.

Table 2. Descriptive and inferential statistics of the estimated marginal means of learning gains during the first and second training period for both experimental groups (ITG = immediate treatment; DTG = delayed treatment). The ITG received the training during the first training period, while the DTG received the training during the second training period.

	First training period (T1→T2)				Second training period (T2→T3)				
	n	M (SE)	F- (p-$value$)	$\hat{d}$	n	M (SE)	F- (p-$value$)	$\hat{d}$	$N^{\dagger}$
Syllable stress awareness[a]									
ITG	51	4.36 (0.39)	46.9 (< .001)*	1.49	45	1.69 (0.28)	26.2 (< .001)*	1.25	43
DTG	52	0.44 (0.38)			50	3.78 (0.26)			44
Standardized spelling test (DRT)									
Spelling (T-scores)									
ITG	52	3.99 (0.66)	7.1 (.009)*	0.59	47	1.95 (0.73)	5.9 (.018)*	0.51	44
DTG	50	1.35 (0.67)			50	4.46 (0.71)			42
Vowel length marking[b]									
ITG	51	3.15 (0.36)	12.0 (< .001)*	0.76	49	0.06 (0.40)	8.8 (.004)*	0.63	45
DTG	51	1.32 (0.36)			50	1.76 (0.39)			43
Self-designed spelling test									
Total score (max=30)[c]									
ITG					46	−0.01 (0.62)	14.8 (< .001)*	0.86	
DTG					46	3.47 (0.62)			
Vowel length marking[b]									
ITG					46	−0.27 (0.57)	16.4 (< .001)*	0.92	
DTG					46	3.16 (0.57)			
Training words (max=9)[c]									
ITG					45	0.60 (0.25)	3.6 (.061)	0.43	
DTG					46	1.28 (0.25)			
Untrained words (max=10)[c]									
ITG					46	−0.03 (0.26)	20.4 (< .001)*	1.02	
DTG					45	1.72 (0.27)			
Inflected training words (max=11)[c]									
ITG					46	−0.32 (0.27)	5.2 (.026)*	0.51	
DTG					45	0.57 (0.27)			
Reading									
Reading fluency (SLS 2–9, LQ-scores)									
ITG	39	6.16 (0.76)	3.2 (.076)	0.42	37	2.87 (0.82)	1.1 (.293)	0.24	33
DTG	45	4.22 (0.71)			45	4.06 (0.74)			36
Word reading (SLRT-II, z-scores)									
ITG	45	0.15 (0.04)[d]	−1.5 (.130)[e]		41	0.16 (0.05)	0.1 (.701)		34
DTG	46	0.24 (0.04)[d]			41	0.13 (0.05)			33
Pseudoword reading (SLRT-II, z-scores)									
ITG	47	0.29 (0.06)[d]	1098 (.695)[f]		41	−0.04 (0.06)	0.1 (.701)		34
DTG	49	0.37 (0.07)[d]			43	0.01 (0.06)			35

[a] Number of correctly identified stress patterns (max=15).
[b] Number of words with mistakes in vowel length marking. Learning gains are inverted to reflect the improvement in vowel length marking.
[c] Number of correctly spelled words.
[d] Mean and standard errors of the raw learning gain (not marginal means of fitted models due to assumption violations).
[e] t-test results due to heterogeneity of regression slopes.
[f] Wilcox rank-sum test results due to non-normally distributed gain scores.
[*] Significant group differences on $\alpha = .05$.
[†] Number of participants included in crossover analyses.

4.2 Efficacy

In the following, we investigated the efficacy of the training on syllable stress awareness (*Hypothesis 2a*), spelling (*Hypothesis 2b*), and reading (*Hypothesis 2c*). Descriptive and inferential statistics of the learning gains during the first and second training period are listed in Table 2. Estimated marginal means of ANCOVAs are shown graphically in Figure 5, Figure 6, Figure 7, and Figure 8.

4.2.1 Effects on Syllable Stress Awareness

The crossover analysis revealed a large significant training effect on syllable stress awareness, $t(84.83) = 7.32$, $p < .001$, $d = 1.57$. The training-induced learning gain in identifying correct stress patterns was significantly higher than the change induced by waiting periods without extra training ($M_{\text{diff}} = 4.0$, 95% CI_{diff} [2.92, 5.10]).

To investigate whether the overall training effect found in the crossover analysis is present during individual training periods, planned contrast analyses were carried out for each training period separately. We found a large significant group effect on syllable stress awareness during the first training period, $F(1, 96) = 46.86$, $p < .001$, $\hat{d} = 1.49$, as well as during the second training period, $F(1, 88) = 26.22$, $p < .001$, $\hat{d} = 1.25$. Figure 5 indicates that during both training periods, children from the active training group improved at a significantly higher rate in syllable stress awareness than children from the control group not receiving extra training.

Taken together, the analyses revealed that children's abilities to correctly identify stress patterns improved at a significantly higher rate when they received the training, which is confirmed by significant effects in favor of the intervention found during both training periods. That is, the training had a strong positive impact on children's syllable stress awareness, providing first evidence of its pedagogical approach to support literacy acquisition.

4.2.2 Effects on Spelling

Next, we investigated whether the intervention goes beyond improving syllable stress awareness alone and fulfills its ultimate goal of positively affecting spelling abilities (*Hypothesis 2b*). Consequently, we first analyzed the data of the standardized spelling test followed by the analyses of our self-designed spelling test.

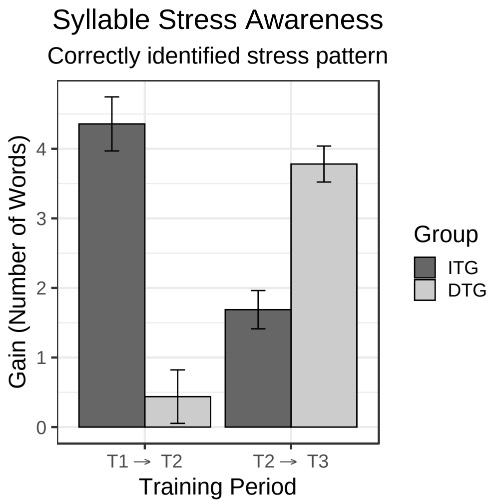


Figure 5. Estimated marginal means of learning gains in syllable stress awareness by group (ITG = immediate treatment; DTG = delayed treatment) and training period. Bars represent the standard errors of the mean.

Standardized Spelling Test. The crossover analysis revealed a large significant training effect on standardized spelling scores, $t(80.82) = 2.79$, $p = .007$, $d = 0.60$. The training-induced learning gain was significantly higher than the learning gain obtained during waiting periods ($M_{\text{diff}} = 2.45$ T-scores, 95% CI_{diff} [0.71, 4.20]). We also found a large significant training effect on the spelling mistakes in vowel length marking, $t(84.30) = 3.28$, $p = .001$, $d = 0.70$. The training-induced improvement in the orthographic vowel length marking was significantly higher than the learning gain during waiting periods without extra training ($M_{\text{diff}} = 1.59$, 95% CI_{diff} [0.63, 2.56]).

The treatment effect was confirmed in the planned contrast analyses, see Table 2 and Figure 6. The ANCOVA revealed a large significant group effect on the standardized spelling scores during the first training period, $F(1, 95) = 7.13$, $p = .009$, $\hat{d} = 0.59$, and a medium to large significant group effect during the second training period, $F(1, 90) = 5.85$, $p = .018$, $\hat{d} = 0.51$. The contrast analyses on the raw score of spelling mistakes in vowel length marking yielded similar results. A large significant group effect was found during the first training period, $F(1, 95) = 12.02$, $p < .001$, $\hat{d} = 0.76$, as well as during the second training period, $F(1, 92) = 8.77$, $p = .004$, $\hat{d} = 0.63$. Figure 6 indicates that children from the active training group improved significantly more in spelling than children from the control group.

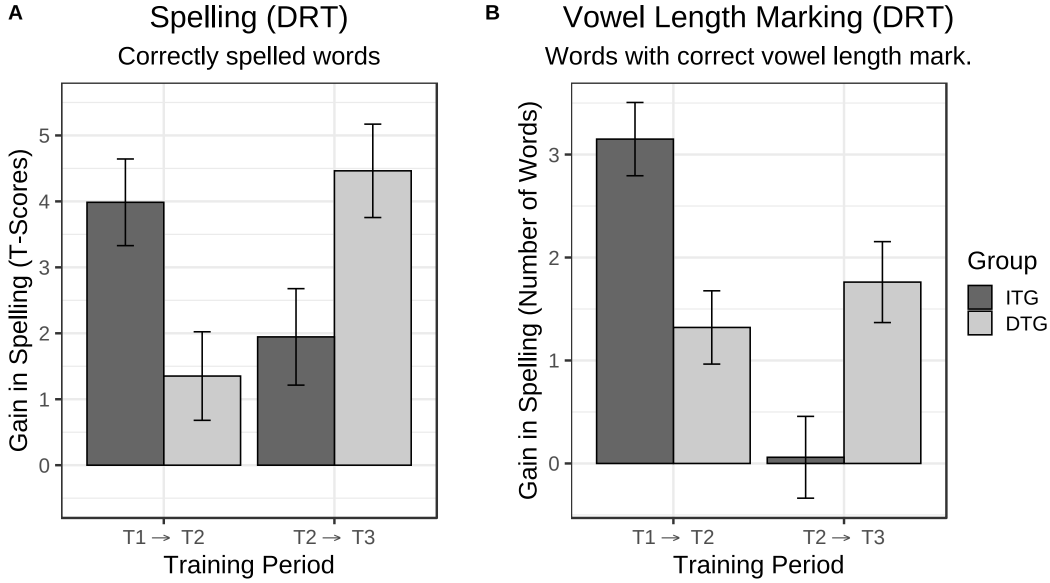


Figure 6. Estimated marginal means of spelling gains on the standardized spelling test (DRT) by group (ITG = immediate treatment; DTG = delayed treatment) and training period. Bars represent the standard errors of the mean. **A**: Gain in the standardized total score; **B**: Gain in vowel length marking (.i.e., number of words with *correct* vowel length marking).

In summary, the analyses of the standardized spelling test revealed that the children's spelling abilities improved at a significantly higher rate during the training as compared to waiting periods, demonstrating the efficacy of the training. We found significant training effects on general spelling ability as well as on the explicitly practiced orthographic marking of long and short vowels. Moreover, the ITG did not decline in spelling during the second training period, i.e., they could maintain their performance level at T3 without further training, indicating a long-term effect of the training.

Self-Designed Spelling Test. To further investigate the effect of the training on specific orthographic learning categories that were not available across grades in the standardized spelling test, we examined the results of our self-designed spelling test that was administered at T2 and T3. During the second training period, the DTG received the training and the ITG did not.

We found similar results for the self-designed spelling test as for the standardized spelling tests, see Table 2 and Figure 7. That is, we found a large significant group effect on the total number of correctly spelled words, $F(1,\ 85) = 14.80$, $p < .001$, $\hat{d} = 0.86$, as well as on the spelling mistakes in vowel length marking, $F(1,\ 85) = 16.40$, $p < .001$, $\hat{d} = 0.92$. Additionally, we found a large significant group effect on the spelling of uninflected untrained words (near transfer learning), $F(1,\ 84) = 20.40$, $p < .001$, $\hat{d} = 1.02$, and a medium sized significant group effect on the spelling of inflected training words (far transfer learning), $F(1,\ 84) = 5.17$, $p = .026$, $\hat{d} = 0.51$. As indicated in Figure 7, children from the DTG, who received the training, improved their spelling at a considerable higher rate during the second training period than children from the ITG. For uninflected training words (no transfer learning), we found a marginal yet not significant group effect, $F(1,\ 84) = 3.61$, $p = .061$, $\hat{d} = 0.43$. As indicated in Figure 7, the group difference in uninflected training words is not significant due to a noteworthy learning gain in the ITG, which may result from consolidation effects.

The results of the self-designed spelling tests confirm the findings of the standardized spelling test. We found a significantly higher spelling improvement in the active training group compared to the control group in the general spelling ability, in the orthographic marking of long and short vowels, as well as in the categories of near and far transfer of learning. Importantly, children did not only improve in spelling of training words, but were also able to apply the acquired knowledge on the trained spelling rules to uninflected words that were not part of the training as well as to inflected training words that were inflected in the spelling test.

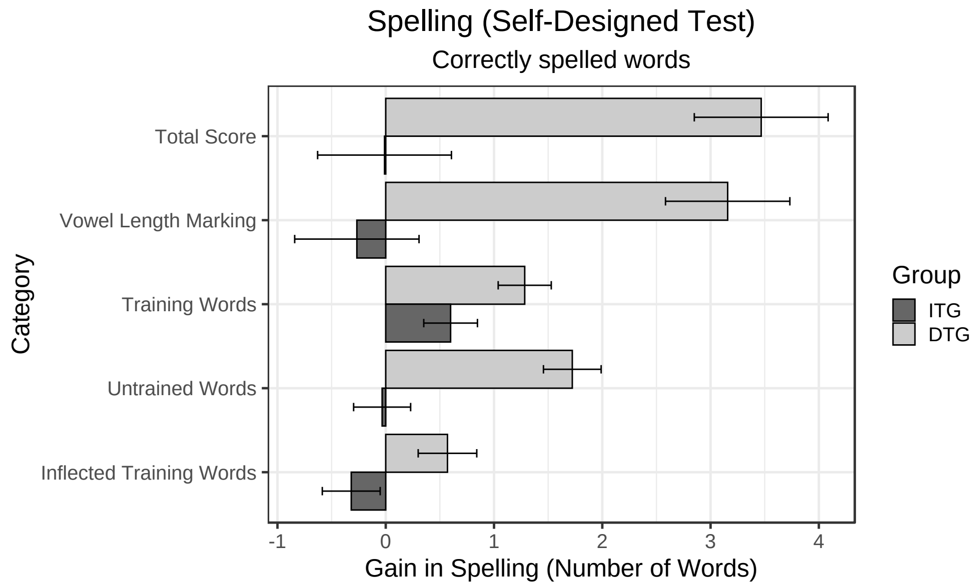


Figure 7. Estimated marginal means of spelling gains on our self-designed spelling test by group during the second training period (T2→T3), in which the DTG received the training. Categories from top to bottom: total score in the test; reduction of spelling mistakes in vowel length marking; training words (no transfer learning); untrained words in basic form (near transfer learning); and inflected training words (far transfer learning). Bars represent the standard errors of the mean.

4.2.3 Effects on Reading

After demonstrating that the training contributes to its primary goals of improving syllable stress awareness and spelling abilities, we further examined the training's effect on reading (*Hypothesis 2c*). Reading was not explicitly trained but might have been positively affected by the training. Accordingly, we analyzed children's reading fluency and word reading.

Reading Fluency. The crossover analysis revealed no significant training effect on reading fluency, $t(66.97) = 0.74$, $p = .465$.

As for the planned contrasts, the ANCOVA of the first training period revealed a marginal yet not significant group effect, $F(1,\ 77) = 3.24$, $p = .076$, while the group effect during the second training period was not significant, $F(1,\ 75) = 1.12$, $p = .293$. Figure 8 indicates that the improvement in reading fluency was more pronounced yet not significantly higher in the active training group than in the control group.

Word Reading. The crossover analysis revealed no significant training effect on word reading, $t(63.31) = 0.11$, $p = .909$, nor on pseudoword reading, $t(66.43) = 0.37$, $p = .710$.

As for the planned contrasts, the group effects on word and on pseudoword reading were not significant during the first training period, $t(88.95) = -1.53$, $p = .130$, and $W = 1097.5$, $p = .695$, nor during the second training period, $F(1,\ 75) = 0.15$, $p = .700$, and $F(1,\ 75) = 0.36$, $p = .551$.

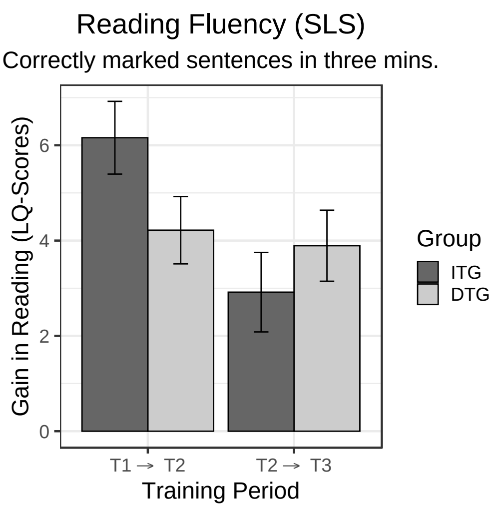


Figure 8. Estimated marginal means of learning gains in reading fluency by group (ITG = immediate treatment; DTG = delayed treatment) and training period. Bars represent the standard errors of the mean.

In sum, we did not find significant treatment effects on untrained reading skills (*Hypothesis 2c*). Yet, we found primary indications that the training meliorates reading fluency of some children, which should be thoroughly investigated in future studies.

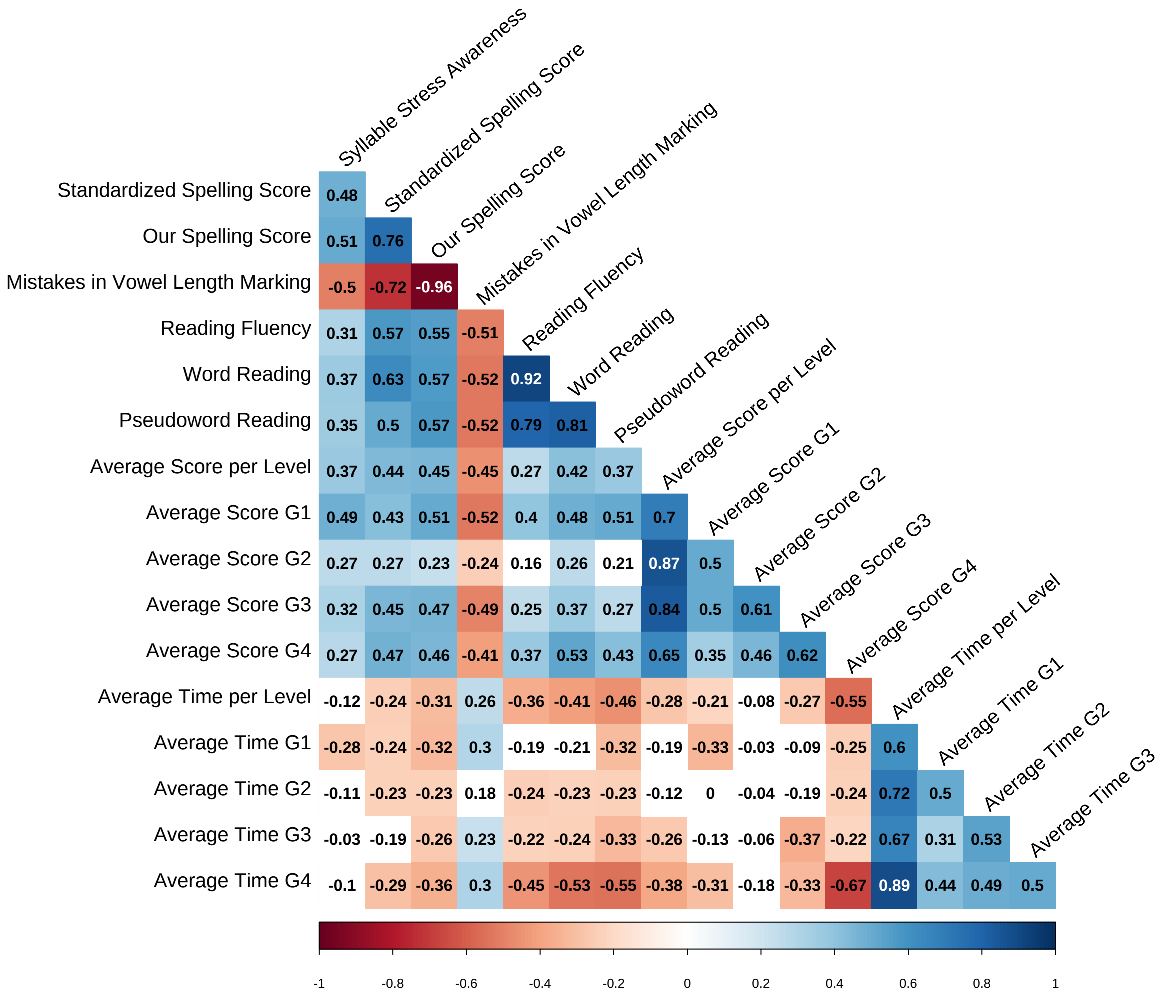


Figure 9. Correlations between literacy skills assessed at T3 (syllable stress awareness, standardized spelling score, our spelling score, words with incorrect vowel length marking, reading fluency, and (pseudo-) word reading) and average in-game scores and times per game type (G1 = "stress pattern", G2 = "open and closed syllables", G3 = "orthographic marker", G4 = "spelling"). Correlations significant on $\alpha = .05$ are colored.

4.3 Validity

Finally we investigated the validity of the pedagogical motivation of the training by investigating the relationship between literacy skills (*Hypothesis 3a*) as well the extent to which the exercises implemented in the training relate to real-life challenges of children with poor spelling and reading skills (*Hypothesis 3b*). The partial correlations are listed in Figure 9.

4.3.1 Relationship Between Syllable Stress Awareness and Reading and Spelling Skills

We found significant positive correlations between syllable stress awareness and reading and spelling skills. Particularly, we found that syllable stress awareness significantly correlated with reading fluency, $r_s = .31$, $p < .001$, word reading, $r_s = .37$, $p < .001$, and pseudoword reading, $r_s = .35$, $p < .001$. Moreover, syllable stress awareness was significantly correlated with the spelling score of the standardized spelling test, $r_s = .48$, $p < .001$, with the more specific spelling score of our self-designed spelling test, $r_s = .51$, $p < .001$, as well as with spelling mistakes in vowel length marking, $r_s = -.50$, $p < .001$. These correlations of moderate effect sizes are in line with current research findings that syllable stress awareness is impaired in children with poor reading and/or spelling skills (Goswami, Gerson, & Astruc, 2010; Goswami et al., 2013; Jiménez-Fernández et al., 2015; Leong et al., 2011; Sauter et al., 2012; Weber, Hahne, Friedrich, & Friederici, 2004). Accordingly, the current results further validate our approach of improving literacy skills by focusing on syllable stress awareness and linking the linguistic features of the stressed syllable to orthographic regularities, in particular to vowel length marking.

4.3.2 Relationship Between Assessed Literacy Skills and In-Game Performances

Moreover, we found that spelling, reading, and syllable stress awareness were significantly correlated with the overall average score achieved in-game as well as the average score achieved in levels of individual game types, see Figure 9. The standardized spelling score significantly correlated positively with all in-game scores, particularly a moderate positive correlation with the average score achieved and its in-game counterpart "spelling" (G4) was found, $r_s = .47$, $p < .001$. The spelling score of our self-designed spelling test that addresses the educational content of the training correlated even more strongly with the average score achieved per level, $r_s = .45$, $p < .001$, as well as with average score of the game "stress pattern" (G1), $r_s = .51$, $p < .001$, and with the average score of the game "orthographic markers" (G3), $r_s = .47$, $p < .001$. The correlations of the spelling mistakes in vowel length marking are inverted but strikingly similar to the total score of our spelling test. Syllable stress awareness and all in-game scores were also significantly correlated. In particular, a moderate positive correlation between syllable stress awareness and the average score of its in-game counterpart "stress pattern" (G1) was found, $r_s = .49$, $p < .001$. Reading fluency correlated significantly with the overall average in-game score as well as with the average score of all game types except for the game "open and closed syllables" (G2), whereas reading fluency most strongly correlated with the average score of the game "stress pattern" (G1), $r_s = .40$, $p < .001$, and with the average score of the game "spelling" (G4), $r_s = 0.37$, $p < .001$. Word reading also correlated significantly with all in-game scores, particularly with the average score of the game "stress pattern" (G1), $r_s = .48$, $p < .001$, with the average score of the game "orthographic markers" (G3), $r_s = .37$, $p < .001$, and with the average score of the game "spelling" (G4), $r_s = .53$, $p < .001$.

The indications of the correlations between literacy skills and in-game performances are twofold. First, they provide support for the validity of the implementation of the game's pedagogical approach (*Hypothesis 3b*). Specifically, the results indicate that the game addresses the difficulties of children with poor literacy skills. This applies to syllable stress awareness, to the general reading and spelling abilities assessed by standardized tests, as well as to the more specific spelling categories included in our self-designed spelling test, particularly the orthographic marking of long and short vowels. Furthermore, the relationship between the literacy skills can also be found in the correlations between the scores of individual game-based exercises. Second, the results are in line with previous research providing evidence that in-game measures such as times (e.g., Sense, Behrens, Meijer, & van Rijn, 2016) and scoring (e.g., Ninaus, Kiili, Mcmullen, & Moeller, 2017) may allow for valid assessment of skills and knowledge.

4.4 Additional Analysis

After the efficacy analyses revealed a significant training effect on syllable stress awareness and spelling abilities, we investigated potential factors that may have influenced the success of the training.

We calculated the total change in the T-scores of the standardized spelling test that can be attributed to the training. That is, we subtracted the waiting-induced improvement from the training-induced improvement for each child included in the crossover analysis. This absolute improvement was subjected to a stepwise linear regression analysis with pre-treatment score (T1 for the ITG and T2 for the DTG), diagnosis of dyslexia, grade, sex, and group assignment as possible predictors. As the full model with all predictors was insignificant, $F(5,\ 580) = 2.19$, $p = .063, R^2 = 0.07$, we performed a bidirectional stepwise regression analysis based on Akaike's information criterion (AIC, Akaike, 1998) to find the most appropriate model. The final model with the lowest AIC was significant, $F(2,\ 83) = 4.17$, $p = .019$, $R^2 = 0.07$, and included the pre-treatment spelling score and sex as predictors, discarding group assignment, diagnosis of dyslexia, and grade.[3] We found that the pre-treatment spelling score was a significant predictor, $\beta = -0.2$, $SE = 0.1$, $t(83) = -2.09$, $p = .039$, indicating that the training success increased with a decreasing spelling ability before treatment. Sex also predicted the improvement significantly, $\beta = 3.7$, $SE = 1.7$, $t(83) = 2.16$, $p = .034$, indicating that the training success was more pronounced in girls than in boys. Interestingly, upon further investigation, we found a marginal significant interaction between pre-treatment score and sex on the spelling improvement attributed to the

[3] Group assignment, diagnosis of dyslexia, and grade were not significant predictors in the full model either.

training. While the spelling improvement in girls only increased little with decreasing spelling ability, boys tend to improve in spelling more strongly with decreasing initial spelling ability. Possibly, the attitude towards the training, i.e., the awareness of the child that it needs the training and the willingness to practice conscientiously, might be differently pronounced in boys and girls with different spelling abilities.

5. Summary and Conclusion

In the present study, we introduced and evaluated a mobile game-based spelling training for German primary school children to improve their syllable stress awareness and spelling skills. The current intervention is the first digital training program that focuses on training syllable stress awareness and linking the linguistic features of the stressed syllable to orthographic regularities of German orthography (i.e., primarily the marking of long and short vowels). The evaluation was carried out with 116 German primary school children from second to fourth grades (aged 7–11 years) in a randomized controlled field trial with a two-period, wait-list controlled crossover treatment design. During respective training periods of 9–10 weeks, children from the active training group were asked to train at home on Android tablets. The evaluation was guided by three hypotheses on the feasibility of the training, i.e., the appropriateness of the digital training program in the home environment (*Hypothesis 1*), the training effect on literacy skills (*Hypothesis 2*), and the validity of its pedagogical approach (*Hypothesis 3*).

Feasibility. To evaluate whether the training can be used at home by primary school children to support their literacy acquisition without extra help (*Hypothesis 1*), we examined the training behavior and collected feedback from children and parents. Investigating the applicability of the training in the home environment is important to determine whether the effects found in the present study may transfer to real-life context outside of scientific studies in controlled environments. Confirming *Hypothesis 1*, the game was found to be easy to use and children spent an average 10 hours with the game. 76% of the children completed the training. This completion rate is comparable to studies conducted in controlled environments. Moreover, as reported in (Holz, Beuttler, & Ninaus, 2018; Holz, Ninaus, et al., 2018), children reported positive training experiences and enjoyed the individual game elements. Overall, the training was received very positively by parents as well as teachers and many families reported that they would continue the training or recommend it to others (Holz, Ninaus, et al., 2018). The training behavior and overall positive feedback indicates the feasibility of the training program. Importantly, the game can be used quite easily by children without additional instructions from parents or teachers and kept children engaged in the training over several weeks.

Efficacy. The main outcome of the current study concerns the efficacy of the training. In particular, the effects of the training on syllable stress awareness (*Hypothesis 2a*), spelling (*Hypothesis 2b*), and reading (*Hypothesis 2c*). We demonstrated that children improved their syllable stress awareness and spelling skills at a significantly higher rate when they actively trained with the program at home, compared to waiting periods in which they did not receive extra training. We found medium to large effects of the training in crossover analyses evaluating within-subject period differences as well as in planned contrasts analyzing the individual training periods separately by means of analyses of covariance. As for spelling, we found significant training effects on the general spelling ability as well as on the orthographic marking of long and short vowels. Moreover, the ITG maintained their spelling improvement during the second training period, in which they did not receive the training, indicating long-term effects of the training. Additionally, we found evidence of near and far transfer of learning in the DTG. The results of the self-designed spelling test showed that children improved in spelling of untrained uninflected words as well as inflected training words at a significantly higher rate than their peers without training. Our results are in line with the consistent finding that improving orthographic knowledge improves the spelling ability in German primary school (dyslexic) children (cf. Galuschka & Schulte-Körne, 2016; Ise et al., 2012).

The training did not have a significant impact on untrained literacy skills, i.e., reading fluency and (pseudo-) word reading. This is not too surprising considering that reading-related (precursor) skills were not explicitly trained. Yet, we found first indications that the training meliorates the reading fluency of some children, which should be further investigated in future studies. For instance, as the different stages of the acquisition of each literacy skill require specific treatment approaches (cf., Galuschka et al., 2014; Galuschka & Schulte-Körne, 2016), the training could extend its current spelling-specific focus by adding modules that specifically target reading (precursor) skills.

In the present study, the average training-induced improvement in spelling, obtained from the estimated marginal means of ANCOVAs of a standardized spelling test, was +4.0 T-scores in the ITG and +4.5 T-scores in the DTG. These learning gains are comparable to other empirically evaluated interventions to improve spelling in German primary school children. Particularly, the learning gains are comparable to other computer-based interventions (training during schools lessons: e.g., Klatte et al., 2018; supervised training sessions and training at home: e.g., Kargl et al., 2008) and to paper-based interventions (training in weekly sessions with trained personnel: e.g., Ise & Schulte-Körne, 2010; Reuter-Liehr, 1993; Schulte-Koerne, Deimel, Huelsmann, Seidler, & Remschmidt, 2001). Of the referenced interventions, our approach is most similar to the *Marburger Rechtschreibtraining* (*Marburg Spelling Training*; Schulte-Körne & Mathwig, 2013), which has been shown to improve the spelling in dyslexic children from grade 2–4 by around +3.2 T-scores (twelve weekly training sessions with trained personnel of 45 minutes each, Schulte-Koerne et al., 2001) and the spelling in dyslexic children from grade 5–6 by between +3.5 and +5.3 T-scores (twelve to fifteen weekly training sessions with trained personnel of 60 minutes each, Ise & Schulte-Körne, 2010). Considering the treatment duration and absolute training time in the present study, our results show that digital game-based interventions can significantly improve spelling in primary school children with comparable learning gains that may even outperform individually administered training sessions. Moreover, the current training can be used by children independently without permanent supervision of trained personnel. Consequently, the training can take place anytime and anywhere – as long as they have access to a tablet or smartphone. Further, the current data demonstrated that our innovative approach yields results comparable to traditional training methods. The approach to systematically teach orthographic knowledge in combination with the awareness of syllable stress seems to be equally beneficial. It might therefore expand the traditional pool of training methods. In the future, we aim to further develop the training to include morphological skills.

Validity. Consistent with *Hypothesis 3a*, we found moderate positive correlations between syllable stress awareness and reading and spelling skills. This is in line with recent empirical findings (Goswami et al., 2010, 2013; Jiménez-Fernández et al., 2015; Leong et al., 2011; Sauter et al., 2012; Weber et al., 2004) and supports our pedagogical approach to improve literacy skills by training syllable stress awareness and shifting the attention to the stressed syllable to teach related spelling rules. Thus, it seems to be a reasonable approach to include stress awareness in the training of reading and spelling skills.

Moreover, our correlation analysis revealed significant associations between literacy skills (syllable stress awareness and reading and spelling skills) and training performances obtained from in-game data (*Hypothesis 3b*). Most interestingly, we found moderate correlations between syllable stress awareness and its in-game counterpart "stress pattern" (G1), $r_s = .49$, between reading skills and the average score of the game "stress pattern" (G1), $r_s = [.40,\ .51]$, and between spelling skills and the average score of the games "stress pattern" (G1), "orthographic marker" (G3), and "spelling" (G4), whose correlation coefficients ranged between $r_s = [.41,\ .52]$. Importantly, we found the correlations between spelling and in-game performances for the standardized spelling ability (assessed with a standardized spelling test) as well as for the more specific spelling score and for the orthographic marking of long and short vowels (assessed in our self-designed spelling test). Based on these findings, we may conclude that the pedagogical content implemented in the training deals with real challenges of children with poor literacy skills and is tailored to the improvement of spelling abilities of poor spellers.

However, the present study also has some limitations. First, due to the scope and complexity of the training, the learning gains in spelling cannot explicitly be attributed. It is not clear whether they result from specific training components (e.g., syllable stress awareness), the combination of specific components (e.g., syllable stress awareness and orthographic marking), or the integration of all components in the holistic intervention and to what extent the playful implementation meliorated the learning gains. Yet, it seems reasonable that a holistic approach in the orthographic stage of spelling acquisition is effective when it includes, besides morphological skills, lexical knowledge and knowledge of spelling rules (Galuschka & Schulte-Körne, 2016; Ise & Schulte-Körne, 2010; Schulte-Körne & Mathwig, 2013), also syllable stress awareness, particularly in the spelling of long and short vowels (Sauter et al., 2012). Second, we observed significant differences in learning gains among children. While the majority could profit from the training, each training group also included some non-responders, i.e., children whose spelling scores did not change or even declined over time. In the future, predictors of children's responsiveness could be addressed, e.g., by enhancing the adaptive learner model, to ensure effective training for each child.

To summarize, we could empirically demonstrate that Prosodiya is an effective, engaging, and easy to use digital game-based spelling training for primary school children. Importantly, the training can be used unassisted without the need of external instructors and evidentially supports improving syllable stress awareness and spelling abilities. Thus, the training program can be particularly useful for children who don't have access to or are waiting for special spelling support. Further, the training can also be used in addition to learning therapy to increase frequency of support.

Acknowledgments

This research was funded by the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments. Heiko Holz is a doctoral student at the LEAD Graduate School & Research Network.

References

Abrami, P., Borohkovski, E., & Lysenko, L. (2015). The effects of ABRACADABRA on reading outcomes: A meta-analysis of applied field research. *Journal of Interactive Learning Research*, *26*(4), 337–367.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of hirotugu akaike* (pp. 199–213). doi:10.1007/978-1-4612-1694-0_15

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1-2), 46–63. doi:10.1159/000208930

Baayen, H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.*

Beddington, J., Cooper, C. L., Field, J., Goswami, U., Huppert, F. A., Jenkins, R., … Thomas, S. M. (2008). The mental wealth of nations. *Nature*, *455*(7216), 1057–1060. doi:10.1038/4551057a

Berkling, K. (2017). Phontasia: A phonics game for German and its effect on orthographic skills–first corpus explorations. In *Proc. WOCCI 2017: 6th international workshop on child computer interaction* (pp. 7–16). doi:10.21437/WOCCI.2017-2

Berninger, V. W., Nagy, W., Tanimoto, S., Thompson, R., & Abbott, R. D. (2015). Computer instruction in handwriting, spelling, and composing for students with specific learning disabilities in grades 4–9. *Computers & Education*, *81*, 154–168. doi:10.1016/j.compedu.2014.10.005

Bhide, A., Power, A., & Goswami, U. (2013). A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind, Brain, and Education*, *7*(2), 113–123. doi:10.1111/mbe.12016

Boyer, N., & Ehri, L. C. (2011). Contribution of phonemic segmentation instruction with letters and articulation pictures to word reading and spelling in beginners. *Scientific Studies of Reading*, *15*(5), 440–470. doi:10.1080/10888438.2010.520778

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., … Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, *94*(Supplement C), 178–192. doi:10.1016/j.compedu.2015.11.003

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read – a causal connection. *Nature*, *301*(3), 419–421. doi:10.1038/301419a0

Bußmann, H. (2008). *Lexikon der Sprachwissenschaft* [Encyclopedia of linguistics]. Stuttgart: Kröner.

Cidrim, L., & Madeiro, F. (2017). Information and Communication Technology (ICT) applied to dyslexia: literature review. *Revista CEFAC*, *19*(1), 99–108. doi:10.1590/1982-021620171917916

Corbalan, G., Kester, L., & J.G. van Merriënboer, J. (2009). Dynamic task selection: Effects of feedback and learner control on efficiency and motivation. *Learning and Instruction*, *19*(6), 455–465. doi:10.1016/j.learninstruc.2008.07.002

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, *88*(4), 715–730. doi:10.1037/0022-0663.88.4.715

Daniel, S. S., Walsh, A. K., Goldston, D. B., Arnold, E. M., Reboussin, B. A., & Wood, F. B. (2006). Suicidality, school dropout, and reading problems among adolescents. *Journal of Learning Disabilities*, *39*(6), 507–514. doi:10.1177/00222194060390060301

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th international academic mindtrek conference: Envisioning future media environments* (pp. 9–15). doi:10.1145/2181037.2181040

Drigas, A., & Batziaka, E. (2016). Dyslexia and ICTs, assessment and early intervention in Kindergarten. *International Journal of Emerging Technologies in Learning (iJET)*, *11*(02), 53. doi:10.3991/ijet.v11i02.5193

Dummer-Smoch, L., & Hackethal, R. (2011). *Kieler Leseaufbau* [Kiel Reading Course] (Vol. 8). Veris.

Eisenberg, P. (1998). Der Laut und die Lautstruktur des Wortes [The sound and the sound structure of the word]. In G. Drosdowski, P. Eisenberg, & G. Bibliographisches Institut (Mannheim (Eds.), *Duden, grammatik der deutschen gegenwartssprache* (4th ed.). Mannheim: Dudenverlag.

Eisenberg, P. (2013). *Grundriss der deutschen Grammatik. Band 1: Das Wort* [Outline of German grammar. Volume 1: The word] (4th ed.; N. Fuhrhop, Ed.). Stuttgart: J.B. Metzler. doi:10.1007/978-3-476-00757-5

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, *67*(September), 156–167. doi:10.1016/j.compedu.2013.02.019

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third ed.). Thousand Oaks CA: Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS ONE*, *9*(2), e89900. doi:10.1371/journal.pone.0089900

Galuschka, K., & Schulte-Körne, G. (2016). Clinical practice guideline: The diagnosis and treatment of reading and/or spelling disorders in children and adolescents. *Dtsch Arztebl Int.*, *113*(16), 279–86. doi:10.3238/arztebl.2016.0279

Görgen, R., Huemer, S., Schulte-Körne, G., & Moll, K. (2020). Evaluation of a digital game-based reading training for german children with reading disorder. *Computers & Education*, *150*, 103834. doi:10.1016/j.compedu.2020.103834

Goswami, U. (1999). Integrating orthographic and phonological knowledge as reading develops: Onsets, rimes and analogies in children's reading. *Converging methods for understanding reading and dyslexia*, 57–75.

Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing*, *23*(8), 995–1019. doi:10.1007/s11145-009-9186-6

Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., & Leong, V. (2013). Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language*, *69*(1), 1–17. doi:10.1016/j.jml.2013.03.001

Griffiths, M. (2002). The educational benefits of videogames. *Education and Health*, *20*(3), 47–51.

Grissom, R. J., & Kim, J. J. (2012). *Effect Sizes for Research: Univariate and Multivariate Applications* (2nd ed.). doi:10.4324/9780203803233

Groth, K., Hasko, S., Bruder, J., Kunze, S., & Schulte-Körne, G. (2013). Interventionseffekte bei Lese-Rechtschreibstörung: Evaluation von zwei Förderkonzepten unter besonderer Betrachtung methodischer Aspekte [Measurement of treatment in developmental dyslexia under methodological considerations]. *Lernen und Lernstörungen*, *2*(3), 161–175. doi:10.1024/2235-0977/a000038

Grund, M. (n.d.). *GUT1. Grundwortschatz 500* [GUT1. Basic vocabulary 500]. Retrieved from [2020-02-13]https://www.gut1.de/grundwortschatz/grundwortschatz-500/

Grund, M., Leonhart, R., & Nauman, C. L. (2017). *Diagnostischer Rechtschreibtest für 4. Klassen (DRT 4)* [Diagnostic Spelling Test for 4th Grade (DRT 4)] (3rd ed.; M. Hasselhorn, W. Schneider, & U. Trautwein, Eds.). Hogrefe Schultests. Retrieved from https://www.testzentrale.de/shop/diagnostischer-rechtschreibtest-fuer-4-klassen.html

Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., & Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. *Computers & Education*, *102*, 202–223. doi:10.1016/j.compedu.2016.09.001

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? – a literature review of empirical studies on gamification. In *2014 47th hawaii international conference on system sciences* (pp. 3025–3034). IEEE. doi:10.1109/HICSS.2014.377

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, *54*, 170–179. doi:10.1016/j.chb.2015.07.045

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Routledge. doi:10.4324/9780203887332

Hills, M., & Armitage, P. (1979). The two-period cross-over clinical trial. *British Journal of Clinical Pharmacology*, *8*(1), 7–20. doi:10.1111/j.1365-2125.1979.tb05903.x

Holz, H., Beuttler, B., & Ninaus, M. (2018). Design rationales of a mobile game-based intervention for german dyslexic children. In *Proceedings of the 2018 annual symposium on computer-human interaction in play companion extended abstracts* (pp. 205–219). doi:10.1145/3270316.3272053

Holz, H., Brandelik, K., Beuttler, B., Brandelik, J., & Ninaus, M. (2018). How to train your syllable stress awareness – A digital game-based intervention for German dyslexic children. *International Journal of Serious Games*, *5*(3), 37–59. doi:10.17083/ijsg.v5i3.242

Holz, H., Ninaus, M., Beuttler, B., Brandelik, K., & Meurers, D. (unpublished). Effectiveness of a digital game-based spelling training for german primary-school children.

Holz, H., Ninaus, M., Meurers, D., & Kirsch, A. (2018). Validity and player experience of a mobile game for German dyslexic children. In *Proceedings of the 2018 annual symposium on computer-human interaction in play companion extended abstracts* (pp. 469–478). doi:10.1145/3270316.3271523

Ise, E., Engel, R. R., & Schulte-Körne, G. (2012). Effective treatment of dyslexia: A meta-analysis of intervention studies. *Kindheit und Entwicklung*, *21*(2), 122–136. doi:10.1026/0942-5403/a000077

Ise, E., & Schulte-Körne, G. (2010). Spelling deficits in dyslexia: Evaluation of an orthographic spelling training. *Annals of Dyslexia*, *60*(1), 18–39. doi:10.1007/s11881-010-0035-8

Jamieson, J. (2004). Analysis of covariance (ANCOVA) with difference scores. *International Journal of Psychophysiology*, *52*(3), 277–283. doi:https://doi.org/10.1016/j.ijpsycho.2003.12.009

Jessen, M., Marasek, K., Schneider, K., & Claßen, K. (1995). Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German. In *Proceedings of the international congress of phonetic sciences* (Vol. 13, pp. 428–431).

Jiménez-Fernández, G., Gutiérrez-Palma, N., & Defior, S. (2015). Impaired stress awareness in Spanish children with developmental dyslexia. *Research in Developmental Disabilities*, *37*, 152–161. doi:10.1016/j.ridd.2014.11.002

Kargl, R., & Purgstaller, C. (2010). *Morphemunterstütztes Grundwortschatz- Segmentierungstraining* [Morpheme-oriented basic vocabulary and segmenting training]. Göttingen: Hogrefe. Retrieved from http://afs-weiterstadt.de/iWeb/AFS_-_Podcasts/BFZ_undMorpheme/Eintrage/2012/6/30_Literatur-Diagnostik-Medien_files/Morpheus-Testzentrale.pdf

Kargl, R., Purgstaller, C., Weiss, S., & Fink, A. (2008). Effektivitätsüberprüfung eines morphemorientierten Grundwortschatz-Segementierungstrainings (MORPHEUS) bei Kindern und Jugendlichen [Effectiveness of a morpheme-oriented basic vocabulary segmenting training (MORPHEUS) for children and adolescents]. *Heilpädagogische Forschung*, *34*(3), 147–156.

Kast, M., Baschera, G.-M., Gross, M., Jäncke, L., & Meyer, M. (2011). Computer-based learning of spelling skills in children with and without dyslexia. *Annals of Dyslexia*, *61*(2), 177–200. doi:10.1007/s11881-011-0052-2

Katusic, S. K., Colligan, R. C., Barbaresi, W. J., Schaid, D. J., & Jacobsen, S. J. (2001). Incidence of reading disability in a population-based birth cohort, 1976-1982, Rochester, Minn. *Mayo Clinic Proceedings*, *76*(11), 1081–1092. doi:10.4065/76.11.1081

Kazakou, M. N., & Soulis, S. (2015). Feedback and the speed of answer of pupils with dyslexia in digital activities. In *Procedia computer science* (Vol. 67, pp. 204–212). doi:10.1016/j.procs.2015.09.264

Ke, F., & Abras, T. (2013). Games for engaged learning of middle school children with special learning needs. *British Journal of Educational Technology*, *44*(2), 225–242. doi:10.1111/j.1467-8535.2012.01326.x

Klatte, M., Bergström, K., Steinbrink, C., Konerding, M., & Lachmann, T. (2018). Effects of the computer-based training program Lautarium on phonological awareness and reading and spelling abilities in German second-graders. In T. Lachmann & T. Weis (Eds.), *Reading and dyslexia: From basic functions to higher order cognition* (pp. 317–333). doi:10.1007/978-3-319-90805-2_15

Klicpera, C., & Gasteiger Klicpera, B. (2000). Sind Rechtschreibschwierigkeiten Ausdruck einer phonologischen Störung? [Are spelling disorders an indicator of phonological deficits?] *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *32*(3), 134–142. doi:10.1026//0049-8637.32.3.134

Klicpera, C., Schabmann, A., & Gasteiger-Klicpera, B. (2013). *Legasthenie - LRS* [Dyslexia] (4th ed.). München: Reinhardt. Retrieved from http://www.reinhardt-verlag.de/de/titel/52968/Legasthenie_LRS/

Klicpera, C., Weiss, J., & Gasteiger-Klicpera, B. (2013). Erfolg einer schulischen Legastheniker-Förderung für Kinder der 3. und 4. Schulstufe nach dem Programm von Reuter-Liehr [Success of a school support program for dyslexic children of the

3rd and 4th grade based on the training program of Reuter Liehr]. *Heilpädagogische Forschung*, *39*(2), 87–95.

Kohler, K. J. (1986). Invariance and variability in speech timing: From utterance to segment in German. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 268–289). Hillsdale: Erlbaum.

Landerl, K. (2003). Categorization of vowel length in german poor spellers: An orthographically relevant phonological distinction. *Applied Psycholinguistics*, *24*(4), 523–538. doi:10.1017/S0142716403000262

Landerl, K., & Reitsma, P. (2005). Phonological and morphological consistency in the acquisition of vowel duration spelling in Dutch and German. *Journal of Experimental Child Psychology*, *92*(4), 322–344. doi:10.1016/j.jecp.2005.04.005

Landerl, K., & Thaler, V. (2013). Reading and spelling acquisition and dyslexia in german. In R. M. Joshi & P. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 135–148). Routledge.

Lenhard, W., & Lenhard, A. (2016). Evidenzbasierte Förderung schulischer Fertigkeiten am Computer: Lernspiele mit Elfe und Mathis. *Förderprogramme für Vor- und Grundschule*, *2006*, 87–113.

Leong, V., Hämäläinen, J., Soltész, F., & Goswami, U. (2011). Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language*, *64*(1), 59–73. doi:10.1016/j.jml.2010.09.003

Mayringer, H., & Wimmer, H. (2014). *Salzburger Lese-Screening für die Schulstufen 2-9.* [Salzburg reading screening for grades 2–9]. Hogrefe.

McArthur, G., Eve, P. M., Jones, K., Banales, E., Kohnen, S., Anandakumar, T., ... Castles, A. (2012). Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews*(12). doi:10.1002/14651858.CD009115.pub2

Moll, K., Kunze, S., Neuhoff, N., Bruder, J., & Schulte-Körne, G. (2014). Specific learning disorder: Prevalence and gender differences. *PLoS ONE*, *9*(7), e103537. doi:10.1371/journal.pone.0103537

Moll, K., & Landerl, K. (2009). Double dissociation between reading and spelling deficits. *Scientific Studies of Reading*, *13*(5), 359–382. doi:10.1080/10888430903162878

Moll, K., & Landerl, K. (2010). *SLRT II – Lese- und Rechtschreibtest. Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT)* [SLRT II – Salzburg Reading and Spelling Test]. Bern: Verlag Hans Huber. Retrieved from https://www.testzentrale.de/shop/lese-und-rechtschreibtest.html

Moll, K., Wallner, R., & Landerl, K. (2012). Kognitive Korrelate der Lese-, Leserechtschreib- und der Rechtschreibstörung [Cognitive correlates of deficits in reading and/or spelling]. *Lernen und Lernstörungen*, *1*(1), 7–19. doi:10.1024/2235-0977/a000002

Müller, R. (2003a). *Diagnostischer Rechtschreibtest für 2. Klassen (DRT 2)* [Diagnostic Spelling Test for 2nd Grade (DRT 2)]. Göttingen: Beltz. Retrieved from https://www.testzentrale.de/shop/diagnostischer-rechtschreibtest-fuer-2-klassen.html

Müller, R. (2003b). *Diagnostischer Rechtschreibtest für 3. Klassen (DRT 3)* [Diagnostic Spelling Test for 3rd Grade (DRT 3)] (4, Ed.). Göttingen: Beltz. Retrieved from https://www.testzentrale.de/shop/diagnostischer-rechtschreibtest-fuer-3-klassen.html

Ninaus, M., Kiili, K., McMullen, J., & Moeller, K. (2016). A game-based approach to examining students' conceptual knowledge of fractions. In R. Bottino, J. Jeuring, & R. C. Veltkamp (Eds.), *Games and learning alliance. gala 2016* (pp. 37–49). Cham: Springer International Publishing. doi:10.1007/978-3-319-50182-6_4

Ninaus, M., Kiili, K., Mcmullen, J., & Moeller, K. (2017). Assessing fraction knowledge by a digital game. *Computers in Human Behavior*, *70*, 197–206. doi:10.1016/j.chb.2017.01.004

Parker, L. E., Lepper, M. R., Bartholomew, K., Cordova, D., & Mayer, J. (1992). Effects of fantasy contexts on children's learning and motivation: Making learning more fun. *Journal of personality and social psychology*, *62*(4), 625–633.

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction*, *29*, 128–140. doi:10.1016/j.learninstruc.2013.02.006

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, *50*(4), 258–283. doi:10.1080/00461520.2015.1122533

Pompino-Marschall, B. (2009). *Einführung in die Phonetik* [Introduction to phonetics]. Berlin, Boston: De Gruyter. doi:10.1515/9783110224818

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.r-project.org/

Ramus, F. (2003). Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, *13*(2), 212–218. doi:https://doi.org/10.1016/S0959-4388(03)00035-7

Rausch, J. R., Maxwell, S. E., & Kelley, K. (2003). Analytic methods for questions pertaining to a randomized pretest, posttest, follow-up design. *Journal* *of Clinical Child & Adolescent Psychology*, *32*(3), 467–486. doi:10.1207/S15374424JCCP3203_15

Reuter-Liehr, C. (1993). Behandlung der Lese-Rechtschreibschwäche nach der Grundschulzeit: Anwendung und Überprüfung eines Konzeptes [Treatment of dyslexia after primary school: Application and verification of a concept]. *Zeitschrift für Kinder- und Jugendpsychiatrie*, *21*(3), 135–147.

Revelle, W. (2018). psych: Procedures for psychological, psychometric, and personality research [Computer software manual]. Evanston, Illinois. (R package version 1.8.12) Retrieved from https://CRAN.R-project.org/package=psych

Röber, C. (2012). Die Systematik der Orthographie als Basis von Analysen von Kinderschreibungen. Eine empirische Untersuchung zur Schreibung der i-Laute [The systematics of orthography as a basis for analyses of children's spellings. An empirical study of the spelling of the "i" sounds]. In U. Bredel & H. Günther (Eds.), *Orthographietheorie und rechtschreibunterricht* (pp. 71–102). Tübingen: Niemeyer.

Ronimus, M., Kujala, J., Tolvanen, A., & Lyytinen, H. (2014). Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge. *Computers & Education*, *71*, 237–246. doi:10.1016/j.compedu.2013.10.008

Rose, S. J. (2009). *Identifying and teaching children and young people with dyslexia and literacy difficulties.* Department for Children, Schools and Families.

Sauter, K., Heller, J., & Landerl, K. (2012). Sprachrhythmus und Schriftspracherwerb [Speech rhythm skills and literacy]. *Lernen und Lernstörungen*, *1*(4), 225–239. doi:10.1024/2235-0977/a000023

Schmitt, K. L., Hurwitz, L. B., Duel, L. S., Linebarger, D. L. N., & Nichols Linebarger, D. L. (2018). Learning through play: The impact of web-based games on early literacy development. *Computers in Human Behavior*, *81*, 378–389. doi:10.1016/j.chb.2017.12.036

Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015). childLex: aa lexical database of German read by children. *Behavior Research Methods*, *47*(4), 1085–1094. doi:10.3758/s13428-014-0528-1

Schulte-Koerne, G., Deimel, W., Huelsmann, J., Seidler, T., & Remschmidt, H. (2001). Das Marburger Rechtschreib-Training - Ergebnisse einer Kurzzeit-Intervention [The Marburg Spelling Training Program - Results of a short-term intervention]. *Zeitschrift fuer Kinder- und Jugendpsychiatrie und Psychotherapie*, *29*(1), 7–15. doi:10.1024//1422-4917.29.1.7

Schulte-Körne, G., Deimel, W., & Remschmidt, H. (1998). Das Marburger Eltern-Kind-Rechtschreibtraining–Verlaufsuntersuchung nach zwei Jahren [The Marburg parent-child-spelling training – follow up after two years]. *Zeitschrift für Kinder-und Jugendpsychiatrie*, *26*(3), 167–173.

Schulte-Körne, G., & Mathwig, F. (2013). *Das Marburger Rechtschreibtraining: ein regelbegleitetes Förderprogramm für rechtschreibschwache Kinder* [The Marburg Spelling Training: A rule-based training program for children with spelling difficulties]. Hogrefe. Retrieved from https://www.testzentrale.de/shop/das-marburger-rechtschreibtraining.html

Schulte-Körne, G., & Remschmidt, H. (2003). Legasthenie - Symptomatik, Diagnostik, Ursachen, Verlauf und Behandlung [Dyslexia – Symptomatology, diagnosis, etiology, long-term outcome and therapy]. *Dtsch Arztebl International*, *100*(7), 396–406.

Schulte-Körne, G., Schäfer, J., Deimel, W., & Remschmidt, H. (1997). Das Marburger Eltern-Kind-Rechtschreibtraining [The Marburg Parent-Child Spelling Training]. *Zeitschrift für Kinder-und Jugendpsychiatrie*, *25*(3), 151–159.

Schulte-Körne, G. (2010). The prevention, diagnosis, and treatment of dyslexia. *Deutsches Arzteblatt international*, *107*(41), 718–727. doi:10.3238/arztebl.2010.0718

Senn, S. S. (2002). *Cross-over trials in clinical research* (Vol. 5). John Wiley & Sons.

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Topics in Cognitive Science*, *8*(1), 305–321. doi:10.1111/tops.12183

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. doi:10.3102/0034654307313795

Snowling, M. J. (1995). Phonological processing and developmental dyslexia. *Journal of Research in Reading*, *18*(2), 132–138. doi:10.1111/j.1467-9817.1995.tb00079.x

Snowling, M. J. (2001). From language to reading and dyslexia. *Dyslexia*, *7*(1), 37–46. doi:10.1002/dys.185

Squire, K. D. (2008). Video games and education: Designing learning systems for an interactive age. *Educational Technology*, *48*(2), 17–26. Retrieved from http://www.jstor.org/stable/44429558

Staffeldt, S. (2010). *Einführung in die Phonetik, Phonologie und Graphematik des Deutschen. Ein Leitfaden für den akademischen Unterricht* [An introduction to phonetics, phonology, and graphematics of German. A guide to academic instruction]. Tübingen: Stauffenburg.

Tacke, G. (2005). Evaluation eines Lesetrainings zur

Förderung lese-rechtschreibschwacher Grundschüler der zweiten Klasse [Evaluation of a reading training for second-grade primary school children with poor reading and spelling skills]. *Psychologie in Erziehung und Unterricht*, *52*(3), 198–209.

Thomson, J. M., & Jarmulowicz, L. (2016). *Linguistic rhythm and literacy* (No. volume 17). John Benjamins Publishing Company. doi:10.1075/tilar.17

Thomson, J. M., Leong, V., & Goswami, U. (2013). Auditory processing interventions and developmental dyslexia: A comparison of phonemic and rhythmic approaches. *Reading and Writing*, *26*(2), 139–161. doi:10.1007/s11145-012-9359-6

Um, E. R., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology*, *104*(2), 485–498. doi:10.1037/a0026609

Vennemann, T. (2011). *Preference laws for syllable structure: And the explanation of sound change with special reference to german, germanic, italian, and latin*. Walter de Gruyter.

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2004). Discrimination of word stress in early infant perception: Electrophysiological evidence. *Cognitive Brain Research*, *18*(2), 149–161. doi:10.1016/j.cogbrainres.2003.10.001

Wellek, S., & Blettner, M. (2012). On the proper use of the crossover design in clinical trials. *Deutsches Aerzteblatt Online*. doi:10.3238/arztebl.2012.0276

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. doi:10.1111/j.1469-7610.1976.tb00381.x

Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education*, *60*(1), 412–425. doi:10.1016/j.compedu.2012.07.018

Zientek, L., Nimon, K., & Hammack-Brown, B. (2016). Analyzing data from a pretest-posttest control group design: The importance of statistical assumptions. *European Journal of Training and Development*, *40*(9), 638–659. doi:10.1108/EJTD-08-2015-0066

A. Description of the Digital Game-Based Spelling Training

A.1 Linguistic Background

The spelling of long and short vowels is a major difficulty for German children (Klicpera & Gasteiger Klicpera, 2000; Landerl, 2003). The orthographic markers, also known as "*Dehnungs- und Doppelungszeichen*" [lengthening and doubling marks], are graphemes marking long and short vowels and generally occur in stressed syllables (markers for long vowels, such as the bigram *ie* in *BIE-ne* [bee] or in conjunction with stressed syllables (markers for short vowels, such as the ambisyllabic consonant doubling *tt* in *Ge-WIT-ter* [thunderstorm]) (Staffeldt, 2010; Vennemann, 2011).

Short vowels are consistently marked by the following two rules (cf. Ise & Schulte-Körne, 2010): (i) "If the short vowel phoneme is followed by only one consonant in the same morpheme,[4] then this consonant has to be doubled in the spelling (e.g., *rennen* [to run], and *Ball* [ball])", and (ii) "if the short vowel phoneme is followed by two or more consonant phonemes in the same morpheme, then these consonants are not doubled (e.g., *Felsen* [rock] and *Wald* [forest])".

In contrast, the marking of long vowels is more complex and less consistent (cf. Ise & Schulte-Körne, 2010). Long vowel phonemes can be marked (i) by doubling the vowel grapheme (e.g., *Haar* [hair]), (ii) by a diphthong[5] (e.g., *Daumen* [thumb]), by marking the long vowel *i* with the bigram *ie* (e.g., *Biene* [bee]), (iii) by adding a "silent h" (e.g., *fehlen* [to miss]), or (iv) simply by the absence of a consonant doubling (e.g., the grapheme *o* is a long vowel phoneme in *holen* [to fetch sth.] but a short vowel phoneme in *wollen* [to want sth.]). However, the rules of long vowel marking are more complex and have many exceptions. For example, marking of the long vowel phoneme *i* follows the rule that "if *i* is a long vowel phoneme, then it is spelled with the bigram *ie* (e.g., *Biene* [bee])", with the exception of words that are not of German origin (e.g., *Kino* [cinema]), words in which the long vowel *i* is not preceded by a consonant (e.g., *Igel* [hedge-

[4] A morpheme is the smallest meaningful unit in written language. For example, the root of a word is a morpheme and *renn* is the root of *rennen* [to run].

[5] Diphthongs are double sounds formed by the combination of two different vowels in a single syllable. Typical German diphthongs are *ei*/*au* (e.g., *weinen* /ˈvaɪ̯ən/ [to cry] and *Kaiser* /ˈkaɪ̯zə [emperor], *eu*/*äu* (e.g., *freuen* /ˈfʀɔɪ̯ən/ [to be pleased] and *Bäume* /ˈbɔʏmə/ [trees]), and *au* (e.g., *Daumen* /ˈdaʊ̯mən/ [thumb]).

hog]), words that are untypical for German as they have more than two syllables (e.g., *Maschine* [machine] or *Mandarine* [tangerine]), pronouns (e.g., *mir, dir, wir* [mine, yours, we] and *ihr, ihm, ihn* [her, him, his]), and others (Röber, 2012).

The same phenomenon of vowel length marking can also be explained on the syllable level. Short vowels are marked orthographically "if the phonological word features an ambisyllabic consonant, a so-called syllabic joint. Then, the grapheme, which phonographically corresponds to the ambisyllabic consonant, is doubled"(Eisenberg, 2013, p. 266). According to syllable rules, an ambisyllabic consonant can function as the final sound of the first stressed syllable or as the initial sound of the following unstressed syllable (Eisenberg, 1998). For example, the consonant *n* in the words *REN-nen* [to run], *KEN-nen* [to know sb. or sth.], or *NEN-nen* [to name sb. or sth.] is ambisyllabic. According to a syllable rule stating that stressed syllables with short vowels are always closed,[6] it functions as the final sound of the first stressed syllable. According to a syllable rule stating that simple consonants between two vowels always belong to the syllable of the second vowel, it functions as the initial sound of the unstressed vowel (Eisenberg, 1998).

As such, vowel length markers express phonological characteristics that are generally connected to syllable stress (Eisenberg, 1998). They express a long and loud syllable rhyme that is typically filled by a stressed long vowel (e.g., the long vowel /ˈeː/ in *NEH-men* [to take]) or by a stressed short vowel which is connected with an ambisyllabic coda (e.g., the short vowel /ˈɛ/ + ambisyllabic coda /n/ → /ˈɛn/ in *REN-nen* [to run]). Thus, the phonological origin of orthographic markers is connected to syllable stress. However, this phonological origin can be superimposed by morphological processes. For example, the ambisyllabic consonant structure can vanish in inflected words (e.g., *RENNT* [he/she/it runs], or *ge-RANNT* [I/we/they/he/she/it ran]), or word formation processes can shift the primary stress to another, unmarked syllable (e.g., *AB-fall* [trash]). However, each of these orthographically marked words can be traced back to the basic form of the trochee – the German disyllabic standard word in which the first syllable is stressed and the second syllable is unstressed (e.g., *FAL-len* [to fall], *REN-nen* [to run], *FEL-sen* [rock], *SE-geln* [to sail]). The phonological origin of orthographic markers lies in this basic form that consists of a stressed and an unstressed syllable.

Further, German orthography, just like in English, closely adheres to the principle of morpheme consistency (Landerl & Reitsma, 2005), i.e., "the spelling of morphemes is preserved in different word forms (e.g., *fahren* [to drive], *Fahrer* [diver], *Gefährt* [vehicle])" (Landerl & Thaler, 2013, p. 136). The orthographic spelling rules are only applicable to the word stem, which is consequently spelled with high consistency. Thus, once the spelling of a certain word stem is stored, it can be applied to all word forms (Landerl & Reitsma, 2005). Moreover, with regard to word stress, German words usually adhere to stem stress (Bußmann, 2008, p. 22), i.e., the stress falls on the first syllable of the stem of the word.

A.2 Educational Content

To date, the first module of Prosodiya has been published that focuses on syllable stress awareness, syllable segmentation, vowel length distinction, orthographic marking of long and short vowels, and spelling. Further modules that focus on, among others, morphological skills (e.g., identifying word stems), are subject to development.

A.2.1 Curriculum and Difficulty Adjustment

The educational curriculum is divided into five curriculum units and is designed on four individual levels whose difficulties increase at different rates throughout the game, see Figure A1. The difficulty addresses task-specific characteristics, i.e., changing the complexity of a task, and the orthographic complexity of words.

At the top level, different linguistic or orthographic skills are covered in individual units. These skills range from syllable stress awareness to vowel length distinction, identification of orthographic markers for long and short vowels, and finally applying spelling rules.

On the second level, units consist of one or more chapters, depending on the scope of the unit. For example, the third unit "orthographic markers" is split into two chapters, whereas the first chapter deals with the orthographic marking of open syllables (long vowels) and the second chapter with the orthographic marking of closed syllables (short vowels).

At mid-level, subchapters within a chapter

[6] Syllables that end with a single or cluster of consonant phonemes (the coda) are called closed syllables, i.e., the syllable is closed by the consonant phoneme(s). In contrast, open syllables are coda-less and end with a vowel phoneme.

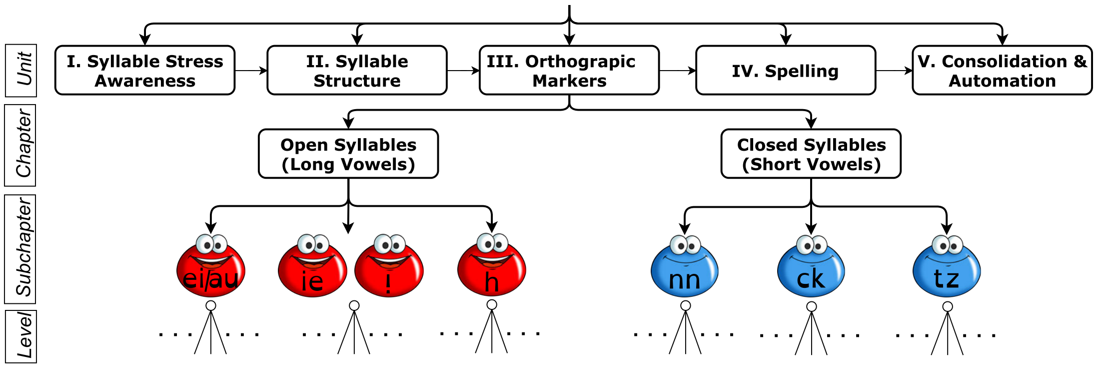


Figure A1. Overview of the pedagogical structure of the present version of Prosodiya. The game increases in complexity and difficulty on four levels at individual rates: units, chapters, subchapters, and levels.

deal with different linguistic or orthographic subcompetencies. For example, the chapter on the orthographic marking of long vowels first deals with diphthongs, then with the spelling of the long *i* (i.e, the bigram *ie* and exceptions), and finally with the "silent h".

Lastly, levels within a subchapter increase in difficulty of the words' structures and complexities as well as in task complexity. For example, the orthographic complexity of words increases as follows: First, phonetically accurate words are trained, i.e., words that are spelled exactly how you hear them (each letter represents one spoken sound). Then, word length and number of syllables increases. Third, words with consonant clusters are practiced, and lastly words with vowel length markers are covered. On the other hand, task complexity increases by decreasing hints and support provided to the children. For example, the game "stress pattern" starts displaying target words syllabified and reveals the number of syllables to the children. Later, the written word is replaced by a corresponding image and/or the number of syllables is not revealed to the children, which results in tasks that also include syllable segmentation.

The word selection as well as unlocking of new content adapts to the individual proficiency level of each child.

A.2.2 Unit I "Syllable Stress Awareness"

In the first unit, children train their syllable stress awareness by identifying stress patterns of given words, see Figure 2a on page 6. We provide three different sound files for each word that increase with regard to the intensity of the intonation. If children request help or submit a wrong answer, the word is spoken in the next stronger intonation level to give scaffolding feedback.

This unit continuously increases in difficulty in that the word length and complexity of the orthographic structures of the target words increases and less frequent stress pattern are practiced. Additionally, the number of syllables is not always revealed to the children and the displayed written word may be replaced by a corresponding image.

As we received feedback in the present study that children wished for more variety in the tasks during the first unit of the game, we also implemented a task of syllable counting for the public version, see Figure A2a. Additionally, easy spelling games (cf. Section A.2.5) are also introduced in the first unit of the public version of the game.

A.2.3 Unit II "Syllable Structure" or "Vowel Length Distinction"

In the second unit, children work on perceiving and distinguishing the length of the vowel of the stressed syllable. For this, we implemented a novel variant of the commonly used vowel length distinction task that builds upon the competence of stress pattern recognition. In addition to detecting syllable stress, the children have to decide whether the stressed syllable is open (the syllable ends with a long vowel, big red blob with its mouth open) or closed (the vowel is closed by a consonant, big blue blob with closed mouth), see Figure 2b. Again, due to the feedback received in the present study to add more variety to the intervention, we implemented an additional simplified version of this game in which children only need to identify the vowel length, without rebuilding the stress pattern, see Figure A2b.

We provide sound files of minimal pairs for each word to support the learner when they require

(a) Game 5: "syllable counting". Children count the number of syllables by pressing the "+" and "–" buttons. The trisyllabic target word is *er-IN-nern* [to remember].

(b) Game 6: "simple vowel length distinction". Children need to decide whether the stressed syllable of the word contains a long vowel (red) or short vowel (blue) by touching the respective Kugellicht. In the given word *BIE-ne* [bee], the stressed syllable contains the long vowel *i*.

Figure A2. Games teaching syllable segmentation and vowel length distinction.

help or submit wrong answers. The minimal pairs consist of the correct pronunciation of the word and a pseudoword counterpart for which the vowel length of the stressed syllable was changed to the contrary.

In this unit, we also address mouth motor activities by teaching the children that at the end of open syllables, they can continuously lengthen the vowel, which keeps the mouth open. At the end of closed syllables, however, the consonant is "stopping" and "squeezing" the vowel and the mouth is closed at the lips, the teeth, or by the tongue. The wording of "open" and "closed" is also reflected in the features of the mouth of the blobs. As children with dyslexia have difficulties permeating the sound level of a language in order to improve letter-sound correspondence on the segmental level (Moll, Wallner, & Landerl, 2012), mouth motor activities can be used to facilitate learning of letter-sound correspondence (Boyer & Ehri, 2011). The difficulty of this unit increases similarly to the first unit.

A.2.4 Unit III "Orthographic Markers"

After acquiring the knowledge about syllable stress and the structure of the stressed syllables, children learn the rules that underlie the spelling of open and closed syllables in the third part of the intervention. This part includes two different game types in which children first learn to recognize the orthographic markers that belongs to the vowel of the stressed syllables, see Figure 2c on page 6, and then spell out the word in a simplified spelling game, see Figure 2d on page 6.

First, children learn about the orthographic marking of long vowels and later about the marking of short vowels. They learn that long vowels can be (i) not marked orthographically (e.g., *RA-ten* [to guess]), (ii) marked with a diphthong (double vowel, e.g., *DAU-men* [thumb]), (iii) marked with the bigram *ie* in case the vowel is a long *i* (e.g., *BIE-ne* [bee]), or by adding a "silent h" (e.g., *FEH-len* [to miss]). In case for the long *i*, unmarked exceptions are also taught (e.g., *TI-ger* [tiger] or *Man-da-RINE* [tangerine]). Words that are marked by adding a silent h are exceptions that do not follow explicit rules and must be memorized and learned by heart with memos such as "*Das stumme h, das ist nicht schwer, steht meist vor l, m, n, und r*" [the silent h precedes mainly but not necessarily the letters l, m, n, and r after a long vowel phoneme]. For the children to better memorize words with a silent h, all words that are marked with a silent h that will be practiced in a level (e.g., *KOH-le* [coal], *FOH-len* [foal], and *FAH-ren* [to drive]) are shown and read out successively at the very beginning of the level, before the first word is practiced.

In the second part of this unit, they learn about the two rules that underlie the spelling of closed syllables. They learn that (i) "if the short vowel phoneme of the stressed syllable is followed by two or more consonants, the "stopper" (the consonant closing the syllable) is not doubled in the spelling (e.g., *FEL-sen* [rock])", and (ii) "if the short vowel phoneme of the stressed syllable is followed by only one consonant phoneme, then the stopper has to be doubled in the spelling as well" (e.g., *REN-nen* [to run]). The ambisyllabic consonant doubling has two special cases that are also trained: *ck* is written instead of *kk* (e.g., *HA-cke* [pick]) and *tz* is written instead of *zz* (e.g., *HIT-ze* [heat]).

The orthographic marking of short vowels is taught using the phonetic rules that originate in the stressed syllable of typical German trochees (see Section A.1) and is explained children-friendly as follows: "if you can hear no other consonant after the stopper of the closed syllable before you hear the next vowel, then the stopper must be doubled! For example, in the word *REN-nen* [to run], you can only hear one consonant after the vowel of the closed syllable, the stopper. You can hear a vowel directly after the stopper! In such cases, you can pronounce the stopper twice. If you can pronounce the stopper twice, then you also have to spell it twice!".

The difficulty increases in the phonetic similarity of choices. For example, the chapter about the long vowel *i* starts with comparing words that have an unmarked long vowel with words whose long *i* is marked by the bigram *ie*. Later on, exception words with a long vowel *i* that are not marked orthographically (e.g., *TI-ger* [tiger]) and words with a short vowel *i* (e.g., *WIN-ter* [winter]) are added to the pool of words.

In the course of these chapters, the two games "orthographic markers" and "spelling" are used alternately so that the children first learn about the respective orthographic markers and then foster their knowledge by spelling out the words. At this point, the "spelling" game only offers the exact letters of a target word to spell it, resulting in a letter arrangement task.

The different orthographic markers and their linguistic characteristics are introduced in individual tutorials. For example, ambisyllabic consonant doubling (e.g., *nn*, *ck*, *tz*) is explained as follows: "if you can hear no other consonant after the stopper of the closed syllable before you hear the next vowel, then the stopper must be doubled! For example, in the word *REN-nen* [to run], you can only hear one consonant after the vowel of the closed syllable, the stopper. You can hear a vowel directly after the stopper! In such cases, you can pronounce the stopper twice. If you can pronounce the stopper twice, then you also have to spell it twice!".

This unit is particularly important as the training to recognize orthographic markers is crucial for spelling acquisition (Galuschka et al., 2014; Landerl, 2003), and the inclusion of algorithms of spelling rules to detect and apply orthographic marking has been successfully shown to improve spelling (e.g., Ise & Schulte-Körne, 2010; Kargl & Purgstaller, 2010) and is recommended by clinical practical guidelines (Galuschka & Schulte-Körne, 2016). However, the algorithms to determine orthographic marking of vowel length have not been related to syllable stress in other computer-based interventions before.

A.2.5 Unit IV "Spelling"

The fourth unit primarily focuses on spelling words to foster children's previously acquired knowledge. In spelling games, children pick letters from the letter area and arrange them in the spelling line, see Figure 2d on page 6. The letter area contains a predefined set of letters that each can be used once to write the word.

Easy Spelling Game. In easy spelling games, no distracting letters are used, resulting in a letter arrangement task. In addition, syllable arcs are drawn underneath the spelling line in some conditions to help link the awareness of orthographic markers to the stressed syllables and to help in syllable segmentation. The colors of the syllable arcs refer to syllable stress and vowel length: yellow for unstressed syllables, red for open stressed syllables, and blue for closed stressed syllables.

Difficult Spelling Game. In comparison to the spelling games practiced earlier, this chapter increases the difficulty by adding distracting letters to the set of available letters. These distracting letters are either not part of the written word or duplicates of present letters. This unit of the game increases the difficulty of the spelling game in terms of adjusting the phonological similarity of distracting letters to actual letters of the word. First, distracting letters that do not share phonological similarities to any letter of the word are used, resulting in a letter discrimination task, see Figure 2d on page 6. Later on, distracting letters that can lead to phonologically very similar or even homophonic misspellings are used. Homophonic words sound alike but are misspelled or have a different meaning. For example, the letters {*ä*, *h*, *l*, *m*} are added to the word *FEL-sen* [rock] that may lead to homophonic misspellings, such as *FEL-lsen* or *FÄL-sen*, or to phonologically very similar misspellings such as *FEL-sem* or *FEH-lsen*. To make the chapter more varied, the other games are also practiced.

Figure A3. In-game map. All regions except for the final chapter – the *Magic Forest* – have been successfully freed from the mysterious fog that is haunting the lands of Prosodiya.

To support scaffolding feedback, individual letters can be solved or distracting letters can be deleted after the children entered a misspelled word.

A.2.6 Unit V "Consolidation and Automation"

In the fifth unit, children consolidate their previously acquired linguistic knowledge about German orthography. For this, all games of the previous units are practiced in medium or hard difficulties to automate reading and spelling processes.

A.2.7 Word Material

The trained word material of the experimental version consists in total of 399 words taken from the *Grundwortschatz GUT1* (Basic Vocabulary GUT1; Grund, n.d.), the *Marburger Rechtschreibtraining* (Marburg Spelling Training; Schulte-Körne & Mathwig, 2013), the *Kieler Leseaufbau* (Kiel Reading Training; Dummer-Smoch & Hackethal, 2011), and the *childLex* (Schroeder, Würzner, Heister, Geyken, & Kliegl, 2015).

As the orthographic regularities trained in the program generally apply to the trochaic word form,[7] the experimental version only included words in their base forms and non-compound nouns. Plural is used in case of monosyllabic nouns (e.g., the plural form *Bäu-me* [trees] is trained instead of *Baum* [tree]). Morphological inflection, i.e., conjugation and declension, is not yet covered. Exercises to deduce the orthographic marking of inflected words, such as to learn that the inflected word form *rennt* [he/she/it runs] is spelled with an ambisyllabic consonant doubling as it is derived from the orthographically marked base form *rennen* [to run], are currently being developed.

A.3 Game Design Elements

Game design elements are used in learning environments to positively engage the learner and to invoke position emotions in order to positively affect learning (Hamari et al., 2016; Plass, Heidig, Hayward, Homer, & Um, 2014) and to increase motivation, satisfaction, and perception towards the learning material (Um, Plass, Hayward, & Homer, 2012). In the following, we briefly describe our approach to keep children engaged with the game and to enable the training to be used unassisted. We refer the reader to (Holz, Beuttler, & Ninaus, 2018) for detailed explanations of the rationales behind and evaluation of the game design elements.

A.3.1 Narrative, Environment, and Game Progress

The training is embedded in a fantasy-themed setting that features narrative and environmental elements, which has been shown to be beneficial for motivation, involvement, and learning (Cordova & Lepper, 1996; Parker, Lepper, Bartholomew, Cordova, & Mayer, 1992). The fantasy world is haunted by a mysterious fog, see Figure 4b on page 9 and Figure A3, that covers all the peaceful land. Little inhabitants called *"Kugellichter"* ["spherical lights"], the game's protagonists and pedagogical agents, seek the children's help as they themselves are too weak to help their homeland. Only the children, guided by the Kugellichter through the world of syllables and orthography, can free the land from its dreadful destiny. In order to decipher the mysteries of German orthography and obtain the "*wisdom of words*", they need to understand and use the "*power of the stressed syllable*".

We implemented a weekly and daily progression system in form of cutscenes, a world map, and changes of environment as well as atmosphere. In the game version used in the present study, only a prologue of the story was implemented to raise the children's interest. More cutscenes were added after the study.

The narrative, environment, and game progress is designed to match the progression of the three lower levels (chapters, subchapters, and levels) of the training's curriculum and difficulty system explained in Section A.2.1. Each chapter is embedded in a unique environment and has an epony-

[7] This also includes trisyllabic words with an unstressed prefix, such as *ver-LIE-ren* [to loose] (/fɛɐ̯.ˈliː.ʁən/)

mous landmark that needs to be freed by the fog, which is reflected by the map and level-based environments of subchapters, see Figure A3.

The children's journey starts at the *Waterfall* – the source of the stressed syllable's power – before it takes them through the *Hovi-Village* to rescue its inhabitants, all the way to the *Glass-Blossom Lake* for its purification. Subsequently, the *Dragon's Stronghold* leads the children to higher grounds, past the *East Mountain* and across *The Great River*, before the journey ends in the *Magic Forest*.

We use three game elements implemented in a weekly and daily progression system to convey the progress of the game: the world map, cutscenes, and change of background environments and atmosphere. While the story is explicitly told in cutscenes narrated by the Kuggelichter (see Figure A4), the deliverance of regions is also reflected on the map (see Figure A3) and in changes of background environments used in levels . We implemented this multilevel progress, which also implicitly tells the story by progressing through the level's backgrounds, to increase the children's self-perception of progression, their perception of positive affect and immersion, and to maintain motivation over longer periods of time. In the following, I will explain each of these game elements in more detail.

Map. We designed the in-game map of the game as the "main scene" of the game from the children's perspective, see Figure 4b on page 9. Each time children progress through the game, corresponding regions on the map are redeemed from the fog and adjacent areas call for their help, awaiting them with new challenges.

On the map, children can either play new levels to make progress and unlock new content, or play old levels to beat their previous high scores and gain stars. We used glass blossoms as level symbols, the yellow Kugellicht to indicate cutscenes, and individual icons for each tutorial. Additionally, flags corresponding to the game's chapters indicate the linguistic challenges that are practiced in the area.

Cutscenes. In cutscenes, the Kugellichter continue the narration of the story. To support the storytelling, corresponding images are displayed in a wooden frame. For example, in the cutscene displayed in Figure A4, children made their way from the *Hovi-Village* and arrived at the shores of the *Glass-Blossom Lake*. After clearing the path,

Figure A4. Cutscene "At the shores of the Glass-Blossom Lake". Kugellichter narrate the story and tell the children about the secrets of the glass blossoms.

they are now asked to clear the fog from the lake so that the inhabitants of Prosodiya can dive for glass blossoms to regain their power and strength that was lost due to the fog.

In our effectiveness study (cf. Holz, Ninaus, Beuttler, Brandelik, & Meurers, unpublished), we received the feedback that cutscenes to explicitly tell the game's story and progression are very motivating and were missed in the study version. In the study version, only a prologue of the story was implemented to raise the children's interest. In the current version, each chapter provides multiple cutscenes.

A.3.2 Interactive Tutorials and Feedback

In order for the intervention to be used by primary school children without the extra help from adults, the two most important design elements are instruction and feedback.

A.3.3 Tutorials and Tooltips

We implemented interactive tutorials for each featured game or linguistic characteristic. The tutorials are kept short, simple, and fun and we tried to ensure that children understand the game mechanics as well as the linguistic background. In order to proceed within a tutorial, children are frequently asked to actively solve the current step following the instructions of the pedagogical agents, see Figure A5. We focused on a high level of interactivity to increase the children's participation and to ensure that they understand new game mechanics and linguistic principles. Besides the instructional support, the tutorials also continue the storyline.

Based on observations in pilot studies, one detailed and comprehensive tutorial in the beginning of a chapter is not enough. Children may forget

Figure A5. Tutorial on the use of *ck*: The yellow Kugellicht explains with the example word *Wecker* [alarm clock] that, instead of doubling the letter *k*, the grapheme *ck* is used in spelling. It then asks the children to move the *ck*-Kugellicht onto the leaf.

Figure A6. Exemplary tooltip for the game "orthographic marker" briefly explaining game mechanics and, in this case, the use of the consonant doubling *ck*.

about the objective of the game, its game mechanics, or about linguistic and orthographic characteristics, especially when they take a longer break from the game. Hence, we also implemented short and spot-on task explanations, so-called tooltips, that appear at the start of each level and that can be accessed manually during play, see Figure A6. The spot-on content consists of a spoken explanation with the voice of the yellow Kugellicht and a simple image of the level's objective and challenges. Depending on the degree of difficulty, the children may also get additional hints on what has changed in the gameplay or what to pay attention to.

A.3.4 Feedback

Besides instructions, feedback in an educational context is crucial for knowledge improvement and skill acquisition and might affect motivation of learners (cf. Shute, 2008). Our game uses scaffolding and so-called knowledge of correct response (KRC) feedback. Scaffolding feedback may help dyslexic children to solve exercises faster (Kazakou & Soulis, 2015) and KCR feedback has been shown to support memorization and deeper learning (e.g., Corbalan, Kester, & J.G. van Merriënboer, 2009; Erhel & Jamet, 2013).

The feedback depends on the children's answers and is as follows: if the answer given to a task is correct, a positive sound is played, stars are collected and added to the current score, the progress bar is adjusted, and game elements respond positively, e.g., Kugellichter happily bounce up and down. A different, more sophisticated sound is played if the task is solved at the first go. In the case of wrong answers, children are encouraged to try again. Affective encouragement may also positively affect their performances (e.g., Schmitt, Hurwitz, Duel, Linebarger, & Nichols Linebarger, 2018).

In addition, scaffolding feedback facilitates solving a task when they fail to do so. In this regard, scaffolding feedback is defined as hints or information on areas that exceed the children's current knowledge that enable them to solve a task they can not complete without extra help (Wood, Bruner, & Ross, 1976). For example, words are replayed with increasingly emphasized intonation when children fail to identify the stress patterns. Or, in the case of spelling exercises, children may delete distracting letters, i.e., letters not found in the target word, or get individual letters solved automatically.

If children are not able to solve a word within three trials, the solution is displayed. When present, the pedagogical agents give spoken feedback as their empathetic responses may positively impact learning (Plass et al., 2015).

A.3.5 Rewards and Incentives

We designed different rewards for Prosodiya. Children can collect points when answering correctly. They get more points if they solve a task at the first go to avoid trial-and-error behavior. Upon finishing a level, children are rewarded with a summary, see Figure A7. Depending on their performances, the level might have been successfully mastered, unlocking subsequent game content. To account for poorer-performing children and to avoid frustration, subsequent content is also unlocked after dynamically adapted number of level repetitions. To provide a high replay value and to increase training effects, we use a 1-3 star rating (i.e., more stars for higher performance) for each level, displayed underneath the level symbol on the world map, see Figure 4b on page 9. In the current version, collected points cannot be redeemed and only reflect in-game achievement.

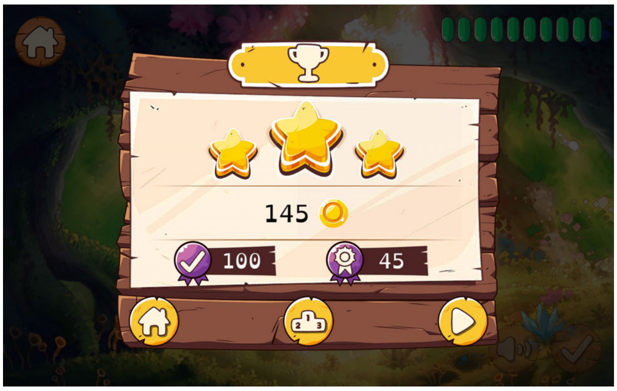


Figure A7. Exemplary summary of a level.

Part VI

Appendix

Appendix A

Materials

A.1 Syllable Stress Awareness Test

We designed a syllable stress awareness test for the effectiveness study of Prosodiya. The test was conducted using an individually administered paper version of the game “Stress pattern” (Figure 5.3 on page 47), in which children had to reproduce the stress pattern of 15 words that were included in the training. Each word was individually read out and displayed as a picture in a PowerPoint presentation. The children had to produce the stress pattern with printed versions of the *Kugellichter* for stressed and unstressed syllables. Test version A was used at T1, version B at T2, and version C at T3. Order of items was randomized prior testing.

	List A	**List B**	**List C**	**Solution**[1]
1	lesen	sagen	rufen	DEEdee
2	Minute	Karate	Banane	deeDEEdee
3	Frage	Blume	Bruder	DEEdee
4	radieren	verlieren	verwelken	deeDEEdee
5	sagen	geben	reden	DEEdee
6	gesund	Gericht	Geschenk	deeDEE
7	verbinden	versinken	verwelken	deeDEEdee
8	Wiese	Liebe	Riese	DEEdee
9	Gedächtnis	Gefängnis	Geschichte	deeDEEdee
10	Stiefel	Spiegel	fliegen	DEEdee
11	beginnen	gewinnen	bekommen	deeDEEdee
12	halten	helfen	Felsen	DEEdee
13	verlassen	Gewitter	Kartoffel	deeDEEdee
14	lernen	turnen	merke	DEEdee
15	sammeln	schnurren	schütteln	DEEdee

[1] DEE = stressed syllable, dee = unstressed syllable.

A.2 Self-Designed Spelling Test

We designed a spelling test to assess more specific learning categories in the effectiveness study of Prosodiya. The self-designed spelling test was assessed at T2 (version A) and T3 (version B) and was the same for all grades.

Version A

	Item	Sentence frame	Category
1.	ankommen	Wir werden morgens ________.	¬T
2.	Brücke	Gehen wir über die ________.	T
3.	ausfallen	Heute werden zwei Stunden ________.	¬T
4.	austrinken	Ich will den Becher austrinken ________.	¬T
5.	backen	Sie ________ einen Kuchen.	T
6.	bekommst	Du ________ ein Geschenk von mir.	IT
7.	biegen	Gummi kann man ________.	¬T
8.	einstecken	Mama muss noch Geld ________.	¬T
9.	flickt	Ben ________ den Reifen.	IT
10.	fliegen	Die Vögel ________.	T
11.	hilft	Hassan ________ Tine bei den Hausaufgaben.	IT
12.	hissen	Piraten ________ die Flagge.	¬T
13.	juckt	Der Stich ________.	IT
14.	kennt	Jana ________ das Mädchen.	IT
15.	klagen	Traurige Menschen ________ oft.	¬T
16.	Krücke	Mit gebrochenen Beinen braucht man eine ________.	¬T
17.	lebt	Der Tiger ________ im Dschungel.	IT
18.	liegt	Der Hase ________ im Gras.	IT
19.	notieren	Auf einem Zettel kann man etwas ________.	¬T
20.	stellt	Karl ________ seinen Ranzen in die Ecke.	IT
21.	schlafen	Wir ________ im Bett.	T
22.	besiegen	Wir wollen die andere Mannschaft ________.	T
23.	schminken	An Fasching ________ wir uns.	¬T
24.	spielst	Du ________Flöte.	IT
25.	treffen	Die Freunde ________ sich.	T
26.	verlassen	Alle ________ das Klassenzimmer.	T
27.	verliert	Anna ________ bei UNO.	IT
28.	versinkt	Das Schiff ________.	IT
29.	Wiese	Kinder spielen auf der ________.	T
30.	wissen	Ich will das ________!	T

[1] T= training word (no transfer learning), ¬T= untrained word (near transfer learning), IT= inflected training word (far transfer learning).

Version B

	Item	Sentence frame	Category[1]
1.	aufwecken	Papa wird die Kinder ________.	¬T
2.	radieren	Bleistift kann man ________.	T
3.	betrinken	Wir wollen uns mit Cola ________.	¬T
4.	Biene	Auf der Blüte sitzt eine ________.	T
5.	Bissen	Lass uns einen ________ vom Brot nehmen.	¬T
6.	dient	Die Magd ________ der Königin.	IT
7.	entkommen	Der Dieb konnte ________.	¬T
8.	frieren	Im Schnee ________ wir schnell.	T
9.	lernt	Ole ________ gerade.	IT
10.	gefallen	Die Bilder ________ mir.	¬T
11.	besiegt	Der Ritter ________ den Angreifer.	IT
12.	gewinnst	Du ________ ja sowieso immer!	IT
13.	knackt	Im Wald ________ es.	IT
14.	leckt	Der Hund ________ am Arm.	IT
15.	packen	Wir ________ unsere Koffer.	T
16.	plagen	Die Mücken ________ uns.	¬T
17.	rasieren	Männer ________ sich morgens.	T
18.	legt	Die Henne ________ Eier.	IT
19.	rennt	Sina ________ schnell.	IT
20.	schiebst	Du ________ den Kinderwagen.	IT
21.	schlagen	Tim will auf den Boxsack ________.	T
22.	sieben	Die Woche hat ________ Tage.	¬T
23.	stellen	Wir ________ die Gläser auf den Tisch.	T
24.	schafft	Martin ________ es, 10 km zu laufen.	IT
25.	Stücke	Wir schneiden den Kuchen in viele ________.	¬T
26.	stinken	Die Socken ________!	¬T
27.	trocken	Das Handtuch ist ________.	T
28.	vermessen	Der Schreiner hat sich ________.	T
29.	verwelkt	Die Blume ________.	IT
30.	Wasser	Wir trinken ________.	T

[1] T = training word (no transfer learning), ¬T = untrained word (near transfer learning), IT = inflected training word (far transfer learning).

A.3 Training Experience Questionnaire

Table A.1: Questions of different categories covered in the Training Experience Questionnaire (TEQ). The number refers to the question's number in the TEQ (S denotes unnumbered Smileyometer questions listed at the very beginning of the TEQ).

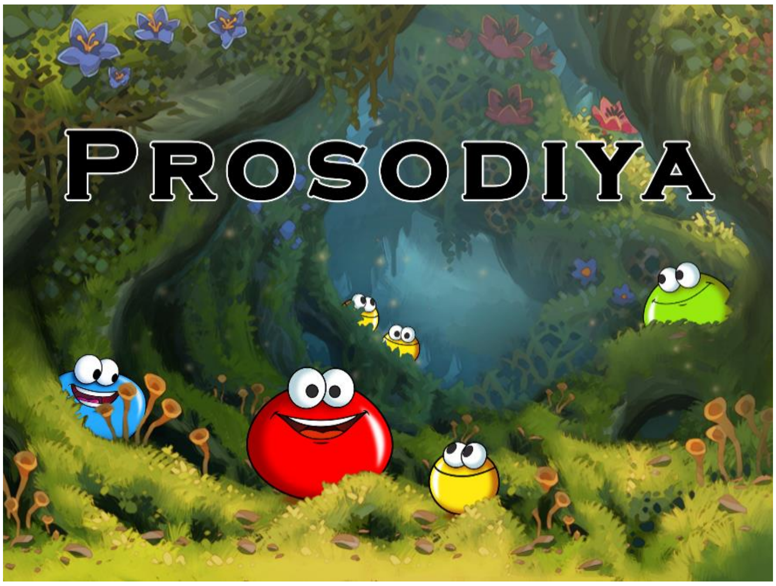
Category		Question	Number
Usability	U1	Did you quickly understand how to play Prosodiya?	27
	U2	Do you think the game Prosodiya is easy to use?	30
	U3	Did you always know what do to while playing the game?	37
	U4	In the different exercises of Prosodiya, was it always clear to you what you had to do?	41
Self-Efficacy	SE1	How much have you learned in Prosodiya that helps you to improve your reading and spelling?	2
	SE2	Did Prosodiya help you to learn to read?	28
	SE3	Did Prosodiya help you to learn to spell?	31
	SE4	Did the game increase your confidence in German classes?	40
	SE5	How often do you think about the things that you learned in Prosodiya when you don't know how to spell a word?	47
Linguistic Awareness	LA1	Do you now know what stressed syllables are?	42
	LA2	Do you now know what open syllables are?	43
	LA3	Do you now know what closed syllables are?	44
	LA4	Do you now know how to spell open syllables?	45
	LA5	Do you now know how to spell closed syllables?	46
Likelihood to Recommend	LTR	Would you go tell a friend Prosodiya is a good game?	25
Intention to Use	ITU	Would you like to continue playing with Prosodiya?	26
Game or Homework	GoH	Do you think Prosodiya is more like homework or more like a game?	1
Kugellichter	K	How did you like the Kugellichter?	S2
	K1	Did you enjoy playing with the Kugellichter?	48
	K2	Did you feel that the Kugellichter were your friends?	49
	K3	Do you think the Kugellichter are stupid?	50

Continued on next page

Table A.1 – TEQ continued from previous page.

Category		**Question**	**Number**
Narrative	S1	Do you think the game's story was interesting?	5
Map	M1	How did you like the map of Prosodiya?	S4
	M2	Did the map give you the feeling of being *in* the world of Prosodiya?	33
	M3	How did you like the fog that covered the map?	S5
	M4	Did you enjoy dispelling the fog from the map?	34
	M5	Were you annoyed by the fog that covered the map?	35
Environment	E1	How did you like the background images?	S6
	E2	How did you like the fog in the exercises?	S7
	E3	Did you enjoy dispelling the fog in the exercises so that the background gets uncovered?	38
	E4	Do you find the fog in the exercises stupid?	39
Tutorials	T1	How well did the Kugellichter explain the “secret of words”?	51
	T2	Did you enjoy the tutorials?	52
	T3	Did the tutorials help you to understand the “secrets of words”?	53
	T4	Do you think the tutorials were too long?	54
Tooltips	TT1	Did you understand the short explanations prior to each level well?	55
	TT2	Did the tooltips help you to better understand and solve the exercises?	56
	TT3	Were you annoyed by the tooltips?	57
Individual Games	G1	How much did you enjoy the game “Stress pattern”?	58
	G2	How much did you enjoy the game “Open and closed syllables”?	59
	G3	How much did you enjoy the game “Orthographic markers”?	60
	G4	How much did you enjoy the game “Spelling”?	61
	Fav.	What was your favorite game?	62

A.3 Training Experience Questionnaire



Prosodiya – Spielerfragebogen

Sag uns deine Meinung!

Datum: ________

Teilnehmer-Code:

1. und 2. Buchstabe deines Vornamens

1. und 2. Buchstabe des Vornamens deiner Mutter

1. und 2. Buchstabe der Straße, in der du wohnst

Bist du Links- oder Rechtshänder? ☐ Linkshänder ☐ Rechtshänder

Bitte gib an, was du über Prosodiya denkst.
Kreuze dafür immer eins der folgenden Bilder an

überhaupt nicht	ein wenig	mittel	ziemlich	sehr

Kreuze das passende Bild an:

Wie sehr mochtest du …	**überhaupt nicht**	**ein wenig**	**mittel**	**ziemlich**	**sehr**
… Prosodiya:					
… die Kugellichter:					
.. das Stickerheft					
… die Landkarte: 					
… den Nebel auf der Landkarte: 					

Wie sehr mochtest du …	**überhaupt nicht**	**ein wenig**	**mittel**	**ziemlich**	**sehr**
… die Hintergrundbilder (die Orte in den Übungen):					
… den Nebel in den Übungen:					

Bitte gib an, was <u>du über Prosodiya denkst.</u>
Kreuze dafür immer einen Kreis an.

1. Findest du, dass Prosodiya eine Hausaufgabe ist, oder ist es ein Spiel?

O	O	O	O	O
Hausaufgabe	eher Hausaufgabe	weder noch	eher Spiel	Spiel

2. Wie viel hast du in Prosodiya gelernt, was dir beim Lesen und Schreiben hilft?

O	O	O	O	O
überhaupt nichts	kaum etwas	ein bisschen	ziemlich viel	sehr viel

3. Wie schwer war Prosodiya?

O	O	O	O	O
sehr schwer	schwer	mittel	einfach	sehr einfach

4. Wie lang findest du eine Übung in Prosodiya? In einer Übung musstest du immer 10 Worte lösen.

O	O	O	O	O
sehr lang	lang	genau richtig	kurz	sehr kurz

Bitte gib an, wie du dich beim Spielen mit Prosodiya gefühlt hast.
Kreuze dafür immer eins der folgenden Kästchen an:

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

Kreuze das passende Kästchen an:

	überhaupt nicht	ein wenig	mittel	ziemlich	sehr
5. Ich fand die Geschichte von Prosodiya interessant	☐	☐	☐	☐	☐
6. Ich habe mich erfolgreich gefühlt	☐	☐	☐	☐	☐
7. Ich habe mich gelangweilt	☐	☐	☐	☐	☐
8. Ich fand Prosodiya beeindruckend	☐	☐	☐	☐	☐
9. Ich habe alles um mich herum vergessen	☐	☐	☐	☐	☐
10. Ich war frustriert	☐	☐	☐	☐	☐
11. Ich fand spielen mit Prosodiya ermüdend	☐	☐	☐	☐	☐
12. Ich habe mich geschickt gefühlt	☐	☐	☐	☐	☐
13. Ich habe mich nervös gefühlt	☐	☐	☐	☐	☐

	überhaupt nicht	ein wenig	mittel	ziemlich	sehr
14. Ich war ganz mit dem Spielen beschäftigt					
15. Ich habe mich zufrieden gefühlt					
16. Ich habe mich herausgefordert gefühlt					
17. Ich musste mich sehr anstrengen					
18. Ich habe mich gut gefühlt					
19. Prosodiya hat mir Spaß gemacht					
20. Ich habe vom Spielen schlechte Laune bekommen					
21. Das Spiel Prosodiya sah schön aus					
22. Ich war gut in Prosodiya					
23. Ich war genervt					
24. Ich fand Prosodiya schwierig					

Was denkst du über Prosodiya?
Kreuze das passende Kästchen an:

25. Würdest du einem Freund von Prosodiya erzählen und ihm sagen, dass Prosodiya ein gutes Spiel ist?

☐	☐	☐	☐	☐
auf keinen Fall	eher nicht	vielleicht	wahrscheinlich	auf jeden Fall

26. Würdest du gerne weiter mit Prosodiya spielen?

☐	☐	☐	☐	☐
auf keinen Fall	eher nicht	vielleicht	wahrscheinlich	auf jeden Fall

Kreuze das passende Kästchen an:

27. Hast du schnell verstanden, wie man Prosodiya spielt?

☐	☐	☐	☐	☐
überhaupt nicht	ein wenig	mittel	ziemlich	sehr

28. Hat dir Prosodiya dabei geholfen, Lesen zu lernen?

☐	☐	☐	☐	☐
überhaupt nicht	ein wenig	mittel	ziemlich	sehr

29. Hast du dich immer darauf gefreut, Prosodiya weiter zu spielen?

☐	☐	☐	☐	☐
überhaupt nicht	ein wenig	mittel	ziemlich	sehr

30. Findest du, dass Prosodiya einfach zu benutzen ist?

☐	☐	☐	☐	☐
überhaupt nicht	ein wenig	mittel	ziemlich	sehr

31. Hat dir Prosodiya dabei geholfen, Schreiben zu lernen?

überhaupt nicht | ein wenig | mittel | ziemlich | sehr

32. Fandst du es schade, dass du auf die neuen Übungen immer bis zum nächsten Montag warten musstest?

überhaupt nicht | ein wenig | mittel | ziemlich | sehr

33. Hat dir die Landkarte von Prosodiya das Gefühl gegeben, in der Welt von Prosodiya zu sein?

überhaupt nicht | ein wenig | mittel | ziemlich | sehr

34. Hat es dir Spaß gemacht, den Nebel von der Landkarte zu vertreiben?

überhaupt nicht | ein wenig | mittel | ziemlich | sehr

35. Hat dich der Nebel auf der Landkarte gestört?

überhaupt nicht | ein wenig | mittel | ziemlich | sehr

36. Findest du es schade, dass das Training mit Prosodiya vorbei ist?

überhaupt nicht | ein wenig | mittel | ziemlich | sehr

37. War dir beim Spielen mit Prosodiya klar, was du machen musst?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

38. Hat es dir Spaß gemacht, den Nebel in den Übungen zu vertreiben damit der Hintergrund frei wird?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

39. Fandst du den Nebel in den Übungen doof?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

40. Fühlst du dich durch das Training mit Prosodiya in Deutsch sicherer?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

41. War dir in den verschiedenen Übungen klar, was du machen musst?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

Kreuze das passende Kästchen an:

42. Weißt du jetzt, was eine betonte Silbe ist?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

43. Weißt du jetzt, was eine offene Silbe ist?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

44. Weißt du jetzt, was eine geschlossene Silbe ist?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

45. Weißt du jetzt, wie man offene Silben schreibt?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

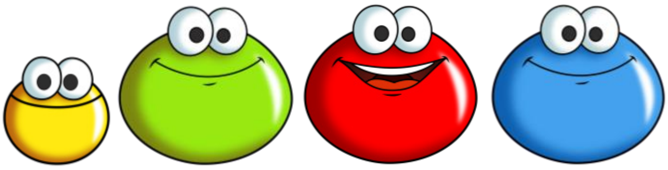
46. Weißt du jetzt, wie man geschlossene Silben schreibt?

überhaupt nicht	ein wenig	mittel	ziemlich sicher	sehr sicher
☐	☐	☐	☐	☐

47. Wenn du mal nicht weißt, wie du etwas schreiben sollst, wie oft denkst du dann an Prosodiya und überlegst, was du dort gelernt hast?

nie	selten	manchmal	oft	immer
☐	☐	☐	☐	☐

Bitte beantworte ein paar Fragen zu den Kugellichtern.



48. Hast du gerne mit den Kugellichtern gespielt?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

49. Hattest du das Gefühl, dass die Kugellichter deine Freunde sind?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

50. Findest du die Kugellichter doof?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr
☐	☐	☐	☐	☐

Bitte beantworte ein paar Fragen zu den <u>langen Erklärungen</u> der Kugellichter, in denen du die Übungen und Geheimnisse kennengelernt hast.

51. Wie gut haben die Kugellichter die Geheimnisse der Wörter erklärt?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr

52. Haben dir die langen Erklärungen Spaß gemacht?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr

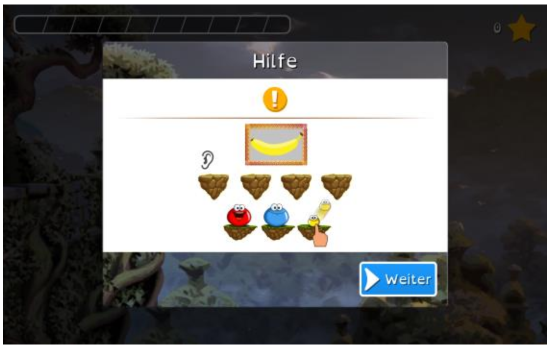
53. Haben dir die langen Erklärungen der Kugellichter dabei geholfen, die Geheimnisse der Wörter zu verstehen?

überhaupt nicht	ein wenig	mittel	ziemlich	sehr

54. Findest du, dass die Erklärungen der Kugellichter zu lang waren?

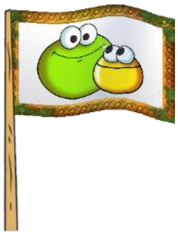
überhaupt nicht	ein wenig	mittel	ziemlich	sehr

Bitte beantworte ein paar Fragen zu den kurzen Erklärungen, die am Anfang von jedem Level vorgelesen wurden



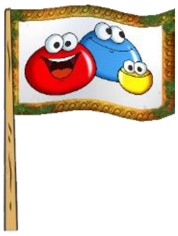
55. Hast du die kurzen Erklärungen vor jedem Level gut verstanden?				
überhaupt nicht	ein wenig	mittel	ziemlich	sehr
56. Haben dir die kurzen Erklärungen dabei geholfen, die Aufgaben besser zu verstehen und zu lösen?				
überhaupt nicht	ein wenig	mittel	ziemlich	sehr
57. Haben dich die kurzen Erklärungen vor jedem Level genervt?				
überhaupt nicht	ein wenig	mittel	ziemlich	sehr

Bitte beantworte ein paar Fragen <u>zu den Übungen von Prosodiya.</u>
Bitte beantworte die Fragen nur zu den Übungen, wenn du im Spiel soweit gekommen bist.



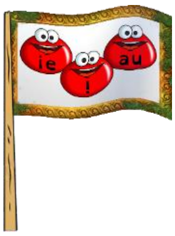
58. Hat dir die Übung „Die betonte Silbe" Spaß gemacht?

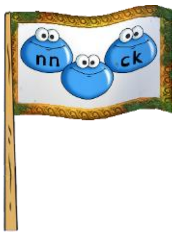
überhaupt nicht | ein wenig | mittel | ziemlich | sehr



59. Hat dir die Übung „Offene und geschlossene Silben" Spaß gemacht?

überhaupt nicht | ein wenig | mittel | ziemlich | sehr





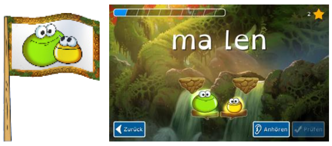
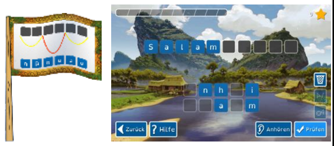
60. Hat dir die Übung „Wie schreibt man die betonte Silbe?" Spaß gemacht?

überhaupt nicht | ein wenig | mittel | ziemlich | sehr

61. Hat dir Übung „Schreiben" Spaß gemacht?

überhaupt nicht — ein wenig — mittel — ziemlich — sehr

62. Welches war deine Lieblingsübung? Kreuze eine an.

☐ „Die betonte Silbe" 	☐ „Offene und geschlossene Silben" 	☐ „Wie schreibt man betonte Silben?" 	☐ „Schreiben" 

Bitte beantworte ein paar Fragen <u>zu den kleinen Bildern</u>, die manchmal zu den Worten im Bilderrahmen angezeigt wurden

63. Wie schön findest du die kleinen Bilder, die zu den Worten angezeigt wurden?

überhaupt nicht — ein wenig — mittel — ziemlich — sehr

64. Wie oft haben die kleinen Bilder zu den Wörtern gepasst?

nie — selten — manchmal — oft — immer

Was hat dir an Prosodiya besonders gefallen?

Hast du noch Tipps für uns, wie wir Prosodiya besser / interessanter / witziger machen können?

A.4 NASA-TLX for Children

Table A.2: The mental demand, physical demand, effort, and frustration subscales of the adapted NASA-TLX for children. Verbal descriptions are listed for the touch interaction style. For drag-and-drop and point-and-touch, "touch the letters" is replaced by "dragging and dropping the letters" and "first selecting and then placing the letters", respectively.

Subscale Simple Title		**Adapted TLX description** (verbal description translated from German)	
Mental Demand Nachdenken (Thinking)		How much did you have to think when you wrote the words by touching the letters? Did you have to do a lot of thinking and figuring out or not very much when you wrote the words by touching the letters? Look at the girl in the pictures. Over here (in the right picture), the girl looks like she is thinking very hard. On this side (in the left picture), it looks like the girl did not have to think a lot. How did you feel when you wrote the words by touching the letters? Show me on the slider how much thinking or figuring out you had to do.	
Physical Demand Bewegen (Moving)		How much did you need to move or strain your budy when you wrote the words by touching the letters? Did you have to move your arm, hand, or fingers much or little when your wrote the words by touching the letters? Look at the boy in the pictures. Over here (in the right pictgure), the boy looks like he had to move his arm, his hand, or his finger very much when he wrote words by touching the letter. On this side (in the left picture), it looks like the boy did not have to move his body a lot. How did you feel when you wrote the words by touching the letters? Show me on the slider how much you you had to move your arm, your hand, and your fingers.	
Effort Anstrengung (Effort)		How hard did you have to work when you wrote the words by touching the letters? Did it take a little or a lot of effort when you wrote the words by touching the letters? Look at the girl in the pictures. Over here (in the left picture), she looks like she had not to use a lot of effort. On this side (in the right picture), it looks like she had to use a lot of effort to write the words by touching the letters. How did you feel when you wrote the words by touching the letters? Show me on the slider how much effort you Show me on the slider how hard you had to work.	
Frustration Ärgern (Annoyance)		How frustrated did you feel when you wrote the words by touching the letters? Did you feel stressed out or did you feel relaxed? Overall, did you feel annoyed or not very annoyed when you wrote the words by touching the letters? Look at the boys in the pictures. Over here (in the right picture), he looks very frustrated and annoyed. On this this (in the left picture), he does very relaxed and not look frustrated or annoyed at all. How did you feel when you wrote the words by touching the letters? Show me on the slider if you felt frustrated or relaxed when you wrote the words by touching the letters.