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] *Leibniz-Institut für Wissensmedien (IWM), Tübingen*
[c] *Department of Psychology, University of Tübingen, Germany*
[c] *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 & June, 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

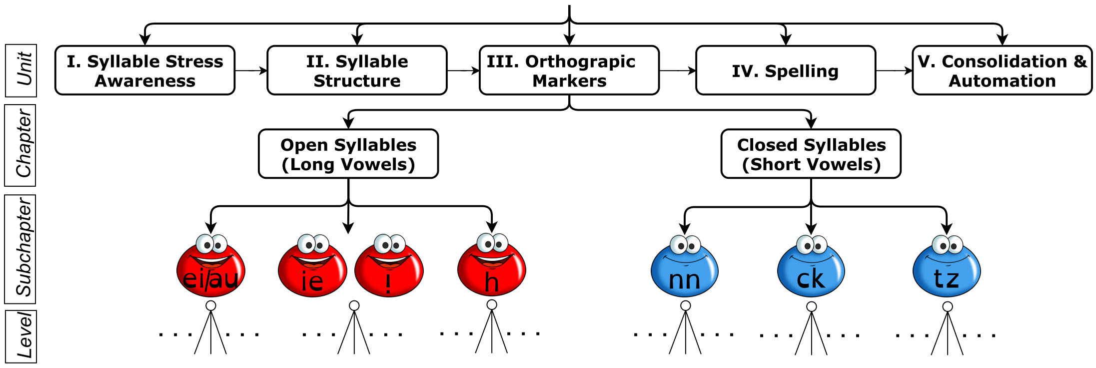


Figure 1. Overview of the pedagogical structure of the present version of Prosodiya. The game increases in complexity and difficulty on four levels at individual rates: units, chapters, subchapters, and levels.

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*.

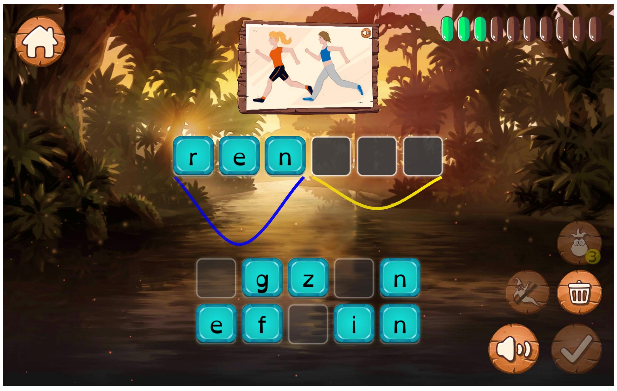
In the first unit, children learn to identify the stress pattern of words and to segment words into syllables, see Figure 2a. In the second unit, children learn to distinguish vowel lengths and to identify open (ends with a vowel) and closed (ends with a consonant) syllables, see Figure 2b. In the third 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 DTG 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

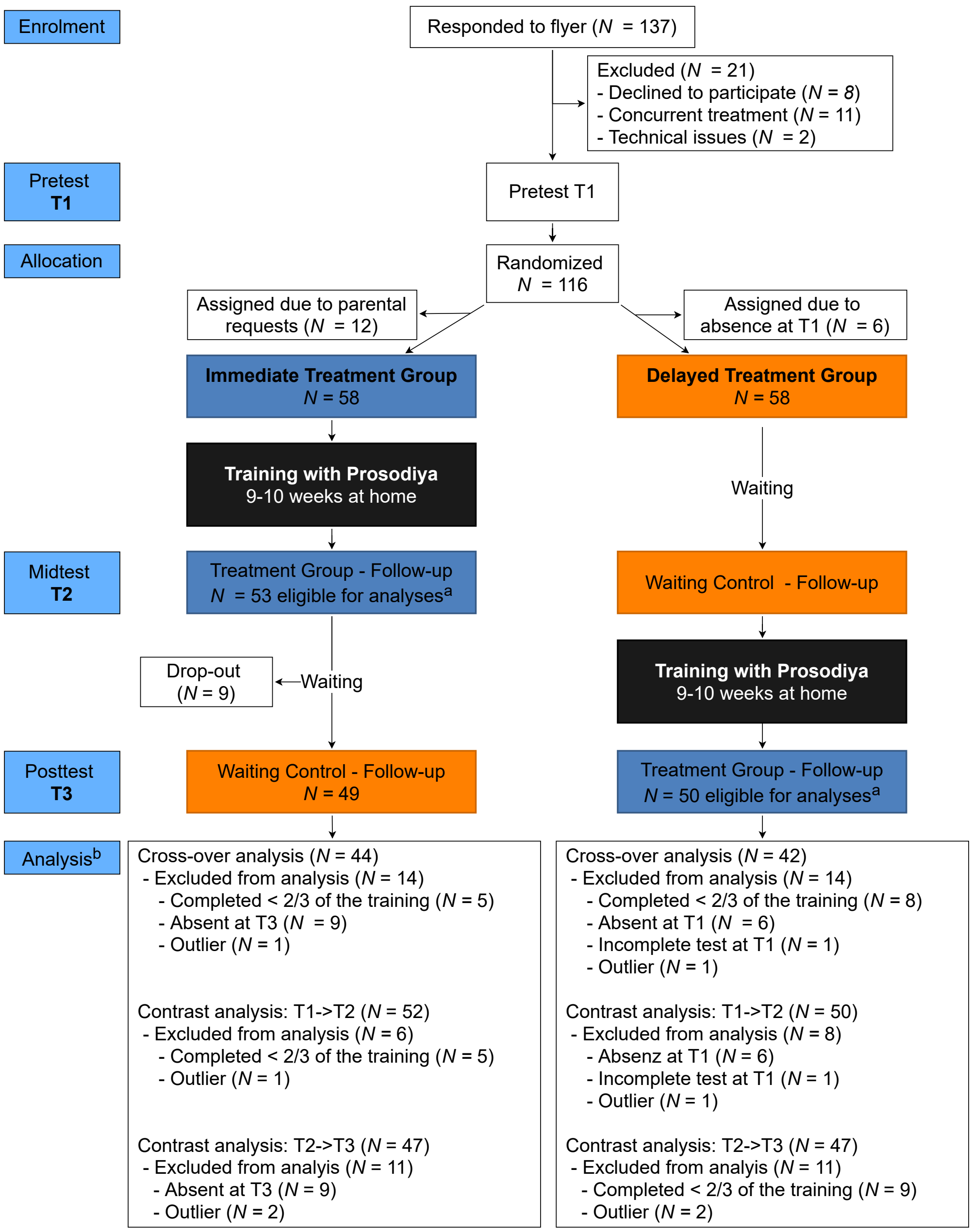


Figure 3. Flow diagram of the two-period, wait-list controlled crossover treatment design for the randomized controlled field trial of the training program.
Note. [a] Only children from the active training group who completed at least two-thirds of the training program were included in the analyses of 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.

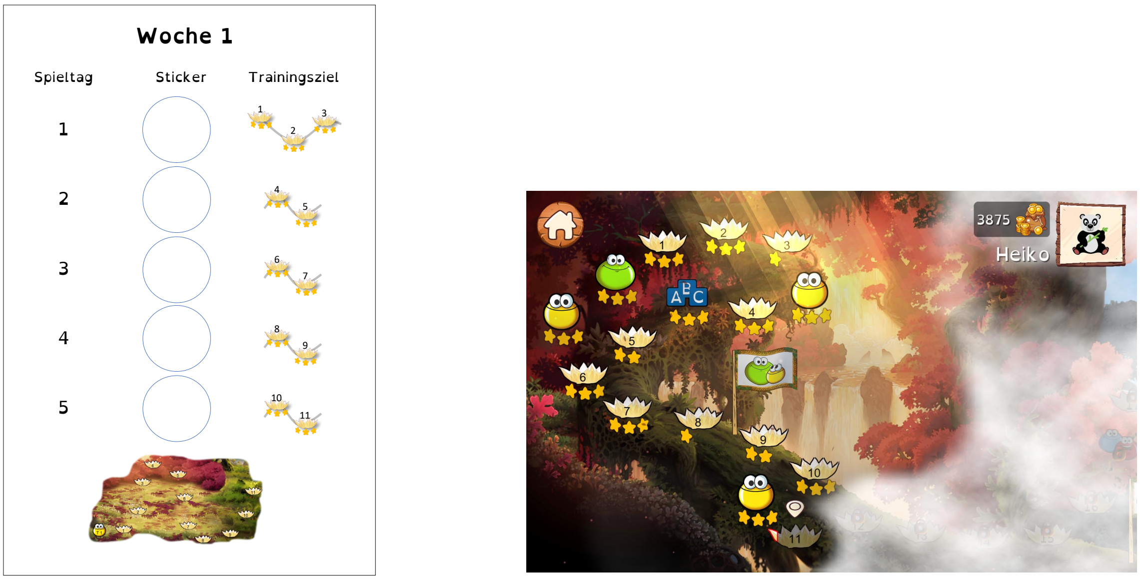
As listed in Table 1, the reading and spelling abilities of the participants were significantly below average and ranged between very poor and below average, with very few exceptions of average performance.

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, 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). 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 cross-over trials when controlling for possible time effects (Senn, 2002; Wellek & Blettner, 2012). The cross-over 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 ITGand 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 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 cross-over 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 psueodoword 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 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 cross-over 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 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 cross-over 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 cross-over 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 significant 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.

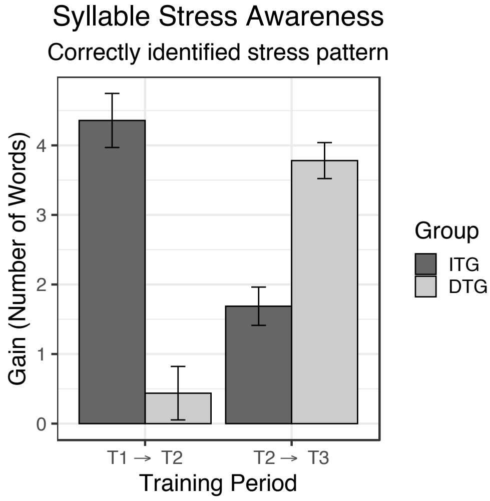


Figure 5. Estimated marginal means of learning gains in syllable stress awareness during by group (ITG = immediate treatment; DTG = delayed treatment) and training period. Bars represent the standard errors of the mean.

Standardized Spelling Test. The cross-over analysis revealed a large significant training effect on standardized spelling scores, $t(80.82) = 2.79$, $p = .007$, $d = 0.60$. 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, $F(1,\ 95) = 7.13$, $p = .009$, $\hat{d} = 0.60$, 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 vowel length marking than children from the control group.

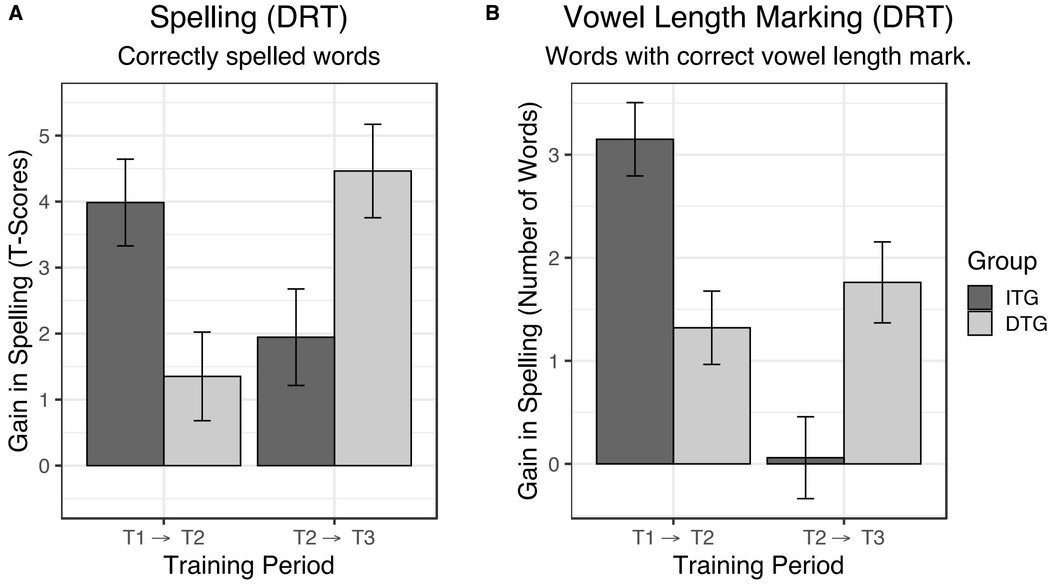


Figure 6. Estimated marginal means of spelling gains on the standardized spelling test (DRT) by group (ITG = immediate treatment; DTG = delayed treatment) and training period. Bars represent the standard errors of the mean. **A**: Gain in the standardized total score; **B**: Gain in vowel length marking (.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.

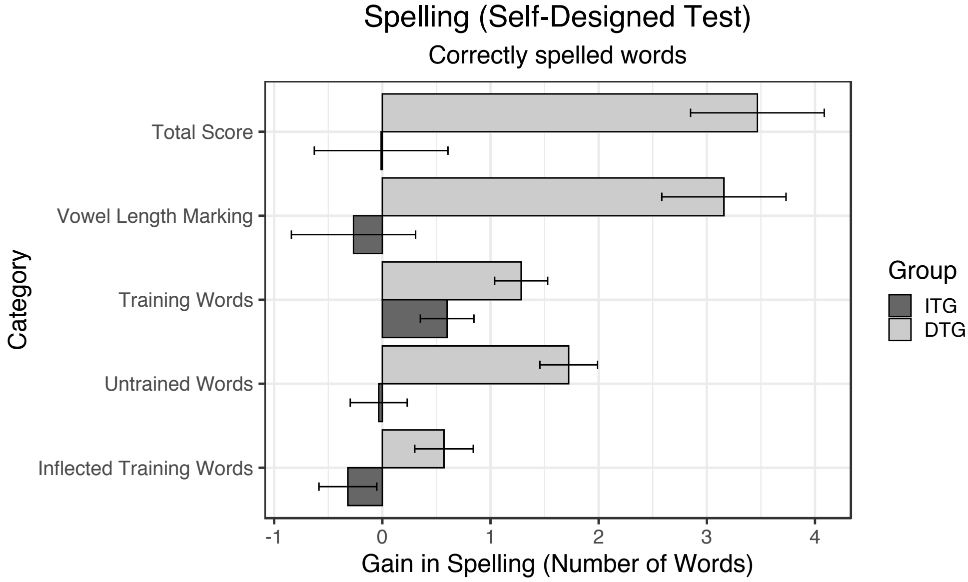


Figure 7. Estimated marginal means of spelling gains on our self-designed spelling test by group during the second training period (T2→T3), in which the DTG received the training. Categories from top to bottom: total score in the test; reduction of spelling mistakes in vowel length marking; training 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 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 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 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 $F(1,\ 75) = 0.36$, $p = .551$.

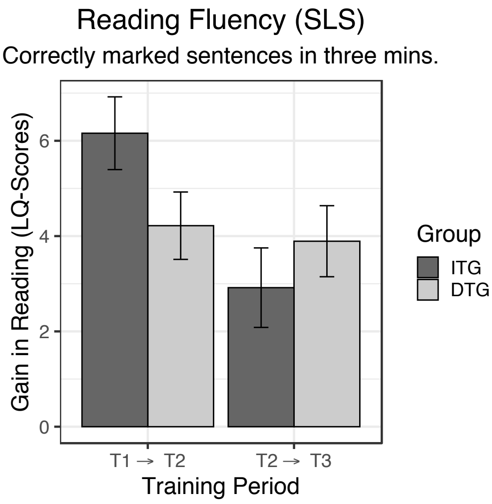


Figure 8. Estimated marginal means of learning gains in reading fluency by group (ITD= immediate treatment; DTG = delayed treatment) and training period. Bars represent the standard errors of the mean.

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.

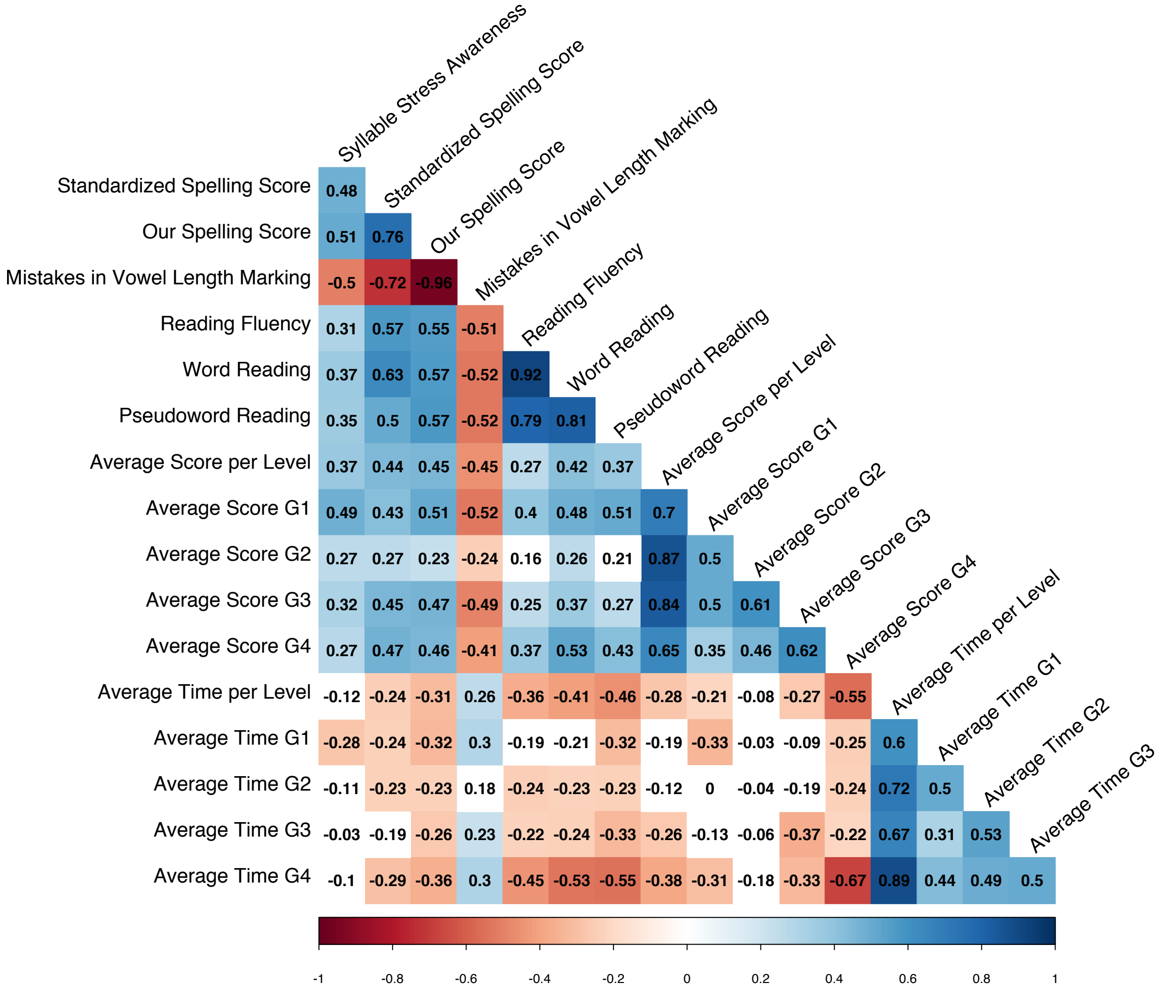


Figure 9. Correlations between literacy skills assessed at T3 (syllable stress awareness, standardized spelling score, our spelling score, words with incorrect vowel length marking, reading fluency, and (pseudo-) word reading) and average in-game scores and times per game type (G1 = "stress pattern", 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 speudoword 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 cross-over analysis. This absolute improvement was subjected to a stepwise linear regression analysis with pre-treatment score (T1 for the iITG 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 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, 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 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., & June, F. (2009). *Identifying and teaching children and young people with dyslexia and literacy difficulties* (No. June). Retrieved from http://www.bdadyslexia.org.uk/news/rose-review.html

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* /ˈfʁɔɪ̯ən/ [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.

[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.

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

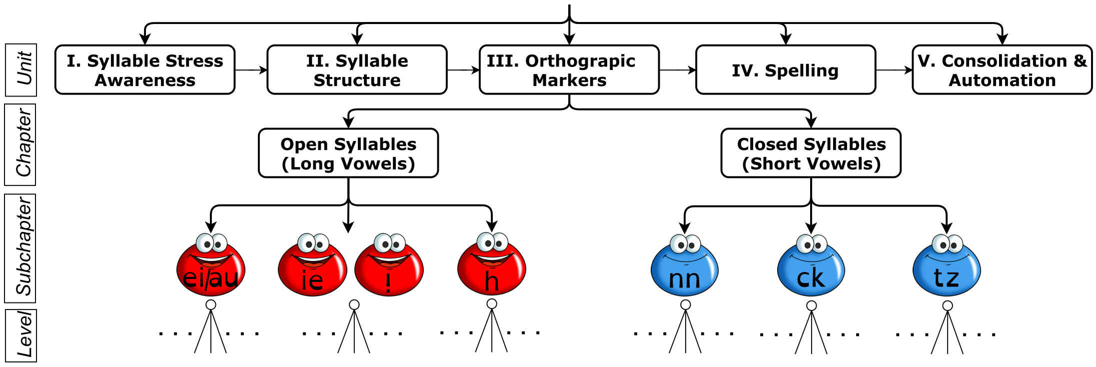


Figure A1. Overview of the pedagogical structure of the present version of Prosodiya. The game increases in complexity and difficulty on four levels at individual rates: units, chapters, subchapters, and levels.

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.