

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Ronimus, Miia; Eklund, Kenneth; Pesu, Laura; Lyytinen, Heikki

Title: Supporting struggling readers with digital game-based learning

Year: 2019

Version: Published version

Copyright: © The Author(s) 2019

Rights: CC BY 4.0

Rights url: <https://creativecommons.org/licenses/by/4.0/>

Please cite the original version:

Ronimus, M., Eklund, K., Pesu, L., & Lyytinen, H. (2019). Supporting struggling readers with digital game-based learning. *Educational Technology Research and Development*, 67, 639-663. <https://doi.org/10.1007/s11423-019-09658-3>



Supporting struggling readers with digital game-based learning

Miia Ronimus¹ · Kenneth Eklund² · Laura Pesu³ · Heikki Lyytinen⁴

Published online: 20 February 2019
© The Author(s) 2019

Abstract

This study investigates the effectiveness of a digital game—GraphoLearn (GL)—in supporting second-grade students who have persistent difficulties with acquiring accurate and fluent reading skills. The participants ($N=37$) were randomly assigned either to a 6-week intervention including sessions with GL, in addition to school-provided support, or a control group receiving only school-provided support. The intervention took place at the students' homes and schools under the supervision of their parents and teachers. The results showed that the children who received the GL intervention developed significantly faster in word reading than the control group. Moreover, their reading development was significantly faster during the GL intervention compared with that of the follow-up period, which included only typical school-provided support. No transfer effects on reading fluency, reading comprehension, or spelling were found. Furthermore, the children who, according to the observations of their parents and teachers, showed higher cognitive engagement during the intervention had higher gains in word reading and sentence reading fluency than the children who appeared less cognitively engaged. Higher emotional engagement was related to increased playing time but not to larger gains in reading. The study indicates that a short digital game-based intervention training of letter–sound correspondences and word-level reading can give a boost to the reading development of struggling readers. Inspecting the engagement and in-game performance during gameplay provides important information that can be used for further development of the game to respond to the needs of the learners with severe difficulties.

Keywords Digital game-based learning · Serious game · Engagement · Motivation · Reading disability · GraphoLearn

✉ Miia Ronimus
miia.ronimus@jyu.fi

¹ Niilo Mäki Institute, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland

² Faculty of Education and Psychology, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland

³ Department of Teacher Education, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland

⁴ Department of Psychology, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland

Introduction

As learning tools, digital games are thought to have enormous potential because they can captivate children's engagement for long periods of time (e.g., Gee 2007; Prensky 2001; Shute et al. 2009). Digital games designed for learning (also known as "serious games") can be considered a distinct form of educational technology; these games aim to pair various game features, such as fantasy, rules, goals, sensory stimuli, challenge, mystery, and control, with instructional content to increase the motivational appeal of learning (Garris et al. 2002). So far, studies have not found consistent evidence that serious games would indeed fulfill their potential as engaging learning tools. The meta-analysis of Wouters et al. (2013) showed that serious games are more effective than conventional instruction in terms of learning, but contrary to popular beliefs, serious games do not appear to be more motivating than conventional instruction. The review of Girard et al. (2012) also indicated that serious games are promising learning tools, but the researchers stressed that firm conclusions about the effects of serious games cannot be made until more experimental studies comparing games with other forms of training (or no training) are conducted.

Additional research is also needed concerning the effectiveness of serious games on specific user groups, such as students who have learning difficulties (Ke and Abras 2013). Serious games could offer an alternative method of practice for these students because these games can provide individually adaptive training on specific skills and experiences of success in tasks which these students would likely fail to achieve in a conventional classroom education. Also, students with learning difficulties often have motivational problems (Chapman et al. 2000; Morgan and Fuchs 2007), and serious games could engage these students in learning more effectively than classroom instruction (Ke and Abras 2013; Rosas et al. 2003).

The present study extends the current knowledge of the potential of a serious game designed for preventing and remediating reading difficulties, specifically here in supporting second graders who have difficulty acquiring accurate and fluent reading skills. Supporting the engagement of these children is especially important because they are often poorly motivated to engage in reading-related activities (e.g., Mol and Bus 2011). We focus on GraphoLearn (GL), the design of which is based on the results of the Jyväskylä Longitudinal Study of Dyslexia (JLD). The JLD has followed the development of language and reading skills of 200 Finnish children—half of whom have a familial risk for dyslexia—from birth to early adulthood, providing ample evidence of the predictors of reading difficulties (e.g., Lyytinen et al. 2008). GL is the only Finnish serious game designed specifically for children who have a risk for dyslexia, making it a natural choice for the present study. In addition to studying the effectiveness of the game, we evaluated the role of engagement in the learning of reading-disabled children.

Literature review

Reading disability, or dyslexia, is one of the most common learning problems and is estimated to affect between 5–10% of the population (e.g., Pennington and Bishop 2009). The process of reading acquisition varies between languages, which has implications regarding reading instruction and remediation of reading difficulties. In transparent orthographies such as Finnish, learning to read is straightforward. Finnish has an almost one-to-one

correspondence between the spoken and written language at the level of phonemes and graphemes, meaning that to learn to decode, one only needs to learn to recognize and sound out each of the 23 letters (and the two-letter grapheme) and to assemble the sounds from sequences of letters into syllables and words. Consequently, more than 90% of Finnish children learn to read accurately at the sentence level during the first months of first grade (Lerikkanen et al. 2004). In opaque orthographies, such as English, achieving an accurate decoding skill is significantly slower (Aro and Wimmer 2003).

Despite the transparency of Finnish orthography, there are students who struggle at achieving accurate and fluent reading skills. Typically, accuracy problems are more pronounced during the first grade of primary school, after which Finnish dyslexic children tend to have more problems in reading fluency (Eklund et al. 2015). Difficulties in decoding often correlate with difficulties in reading comprehension (Juel et al. 1986) and may result in poor school achievement and choosing a career that does not require an extended education (Savolainen et al. 2008).

According to the predominant phonological theory, dyslexic individuals have difficulties in the representation, storing, or retrieval of speech sounds (e.g., Ramus et al. 2003). The results of the JLD indicate that one of the core deficits in dyslexia is related to phonological sensitivity, which appears as a difficulty in learning to differentiate speech sounds from each other and connecting them to corresponding letters (e.g., Lyytinen et al. 2008). The meta-analysis of Galuschka et al. (2014) provided further support for the phonological theory by showing that only reading interventions that include phonics instruction (the systematic teaching of letter–sound correspondences and decoding strategies based on blending or segmenting individual letters or phonemes) produce significant effects on reading performance.

Dyslexia often co-occurs with other learning difficulties, such as attention difficulties (Willcutt and Pennington 2000) and language impairment (Pennington and Bishop 2009), which may restrain a child's responsiveness to generally effective reading interventions. Moreover, children with reading difficulties are unlikely to voluntarily engage in behaviors that would promote the development of reading skills, such as seeking books and other materials to read during their leisure time (e.g., Mol and Bus 2011). Children with reading difficulties may also have low academic self-concepts (Chapman et al. 2000) and suffer from socio-emotional problems, such as stress, anxiety, and depression (Mugnaini et al. 2009). Therefore, children with reading disabilities often need individualized training that is not available in a typical classroom education.

Researchers have stressed the importance of early interventions in the prevention of the accumulation of difficulties (e.g., Torgesen 2004). Children who have difficulties with reading often become reluctant to practice reading, leading to more persistent problems (Morgan and Fuchs 2007). Therefore, interventions should be delivered before the negative motivational consequences of poor reading skills start to emerge.

Supporting reading skill development with digital game-based learning

Previous studies have indicated that educational technology in general has a small to moderate effect on the improvement of the reading skills of struggling readers (see the meta-analysis of Cheung and Slavin 2013). However, less is known of the specific effects of games in supporting children with reading difficulties. A benefit of serious games is that the instructional content can be tailored to the child's individual needs and be dynamically adapted to the child's skill level. Having an optimal level of challenge is crucial for players'

learning and engagement (Abdul Jabbar and Felicia 2015; Garris et al. 2002; Ke and Abras 2013; Lepper and Malone 1987). Also, learning fundamental skills, such as grapheme–phoneme correspondences, requires a sufficient amount of repetition to automatize the skill. A serious game can provide the amount of drilling each learner needs, potentially without creating boredom. Similar practice would be difficult to implement using conventional methods of instruction.

A few earlier studies have investigated the effects of digital games on the development of children’s reading skills. The study of Van de Ven et al. (2017) involved 8-year-old Dutch children ($N=60$) with mild learning disabilities. The children participated in a short (9×15 min) reading intervention. The training (e.g., grapheme-to-phoneme conversion and semantic categorization) was embedded in an adventure game (Letter Prince) that included several motivational elements. The short intervention improved children’s pseudoword and text-reading fluency (immediate and long term) but not the children’s self-reported reading motivation. Van Gorp et al. (2016) investigated the effectiveness of a game called Reading Race on 8-year-old Dutch children ($N=64$) with poor word-decoding skills (below the 25th percentile). The game aimed at improving reading efficiency by giving tasks that included reading words and pseudowords and making semantic categorizations. Gaming elements were used to encourage the player to produce fast and correct responses. The game—used in the classrooms for a total of 5 h over 5 weeks—was effective at increasing readers’ word decoding efficiency, and the effects were retained 5 weeks after the intervention.

In the present article, we focus on GL (formerly known as GraphoGame), which aims to improve children’s phoneme discrimination and grapheme–phoneme correspondence skills. In GL, the player hears a speech sound from headphones and tries to find and click with the mouse on the corresponding letter among two or more alternatives shown on the screen (see Fig. 1). The player receives immediate feedback, and in the case of an incorrect response, the trial is repeated, and the correct choice is highlighted to reinforce the correct connection between the sound and letter. Each trial is short—each

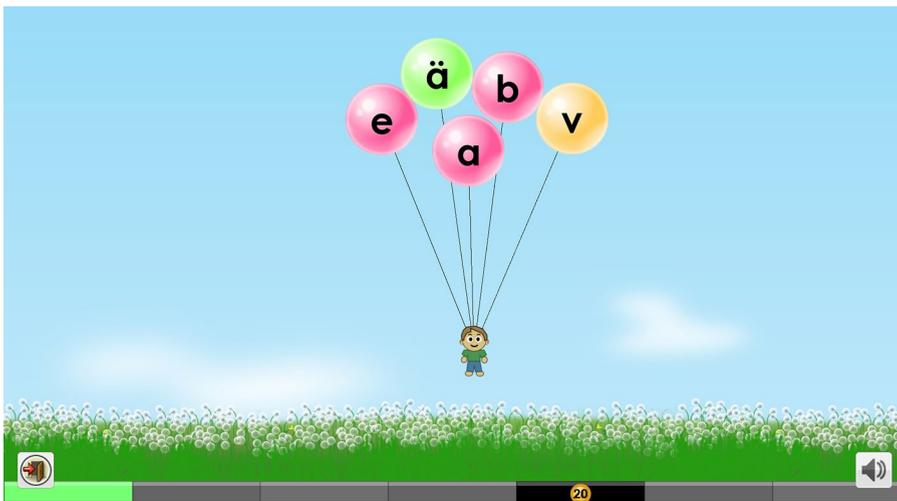


Fig. 1 An example of a GL learning task. The player has heard a speech sound and is expected to select the corresponding letter from the alternatives shown on the screen. After selecting the correct letter, the balloon pops. The goal is to pop all the balloons and help the avatar land safely

being only a few seconds—and the trials follow each other at a fast pace, resulting in an intensive gaming experience where a session of few minutes may include more than a hundred trials. As the player advances in the game, the trials start to include longer units, such as syllables and short words. It is expected that exposure to trials such as these will improve the players' proficiency in letter–sound correspondences and identification of large units (syllables and short words), leading to better reading accuracy and speed.

GL is designed as an early intervention of children between the ages of 6 and 8. The user interface is simple, and all instructions are spoken aloud so that using the game does not require having any reading skill. The difficulty level of the game adapts dynamically to a child's performance. The game starts from the letters and sounds that are considered the easiest to learn (and are taught first in the first-grade curriculum), and depending on the level of the child's performance, the game gradually moves on to more difficult items. The game repeats each connection until the adaptation determines that the child has learned it and that practice is no longer required. Typically, a player's success rate is between 80% and 90%, which is assumed to be high enough to keep the player engaged without undermining the experience of challenge. For a thorough description of the GL method, see Richardson and Lyytinen (2014). Information of the international GL research and images of the game are available on the following webpage: <http://grapholearn.info>.

Saine et al. (2010) have documented the effects of GL training when it is given as a part of remedial reading instruction sessions aimed at first graders with mild difficulties in pre-reading skills (below the 30th percentile). The intervention was carried out over a period of 28 weeks as 4 weekly sessions of 45 min each (of which 15 min was used for GL training). The results showed that the intervention had a significantly more powerful effect on GL players' ($N=25$) letter knowledge, reading accuracy, spelling, and reading fluency compared with children ($N=25$) who participated in similar remedial training but without GL.

Two other studies focused on the effects of GL on reading speed. Heikkilä et al. (2013) studied second- and third-grade students ($N=150$) who had poor reading speed and found that a short intervention consisting of ten 5–10-min sessions for 2–3 weeks increased the children's reading speed of the syllables which were trained in the GL. Also, Hintikka et al. (2008) studied second- and third-grade students ($N=39$) with a low reading speed (below the 25th percentile) and found that a short GL intervention (six 15–20-min sessions provided over a period of 8–10 days) improved the reading speed of the trained sublexical items, but the training effect was not generalized to pseudoword reading or general reading speed. These two studies did not find the systematic transfer effects on reading speed, which may be because of the short duration of the training. The study of Saine et al. (2010) showed that longer interventions may be more effective in producing effects also in skills not directly trained by the game, such as reading fluency and spelling. Further evidence of the effects of GL have been obtained in brain imaging studies involving nonreading, 6-year-old kindergarten students. These studies have showed changes in children's neural processing after short interventions with GL (Brem et al. 2010; Lovio et al. 2012).

In addition to the Finnish version, GL has been implemented in several other languages, and its effects on children's reading skills have been studied internationally (e.g., Jere-Folotiya et al. 2014; Ojanen et al. 2015). Kyle et al. (2013) studied the effects of English GL on the learning of 6–7-year-old second-grade students ($N=31$) who had poor reading skills, per a teacher's evaluation. The results indicated that GL-based training of consistent connections between spoken and written items works also in the inconsistent orthography of English, and the best results are achieved by a training that includes both letter–sound connections and longer units (i.e., orthographic rime units).

The present study extends the knowledge gathered in previous GL studies by focusing on children who had persistent difficulties with reading, that is, second graders whose reading skill fell below the 14th percentile. So far, there are few findings concerning the feasibility of GL—or digital game-based learning in general—for users who have moderate to severe learning difficulties (Ke and Abras 2013). Because those with reading difficulties are often not motivated to read (Chapman et al. 2000; Morgan and Fuchs 2007), the present study also investigated the role of engagement in the effectiveness of GL training, which has not been addressed in previous GL studies.

The potential of games in engaging struggling learners

One of the potential benefits of educational games is that they tend to engage students in learning more effectively than traditional instruction (e.g., Annetta et al. 2009; Rosas et al. 2003; Wrzesien and Raya 2010); however, there are contradictory findings on this as well (Wouters et al. 2013). The features considered to make digital games engaging include challenge, control (opportunities for making choices), clear goals, feedback, fantasy, and immersion (Abdul Jabbar and Felicia 2015; Garris et al. 2002; Ke and Abras 2013; Lepper and Malone 1987). Engagement here can be defined as a focused involvement in an activity that is accompanied by a positive emotional tone (Skinner and Belmont 1993). Many studies conducted in classroom contexts have indicated that student engagement is related to positive academic outcomes (e.g., Finn and Zimmer 2012; Fredricks et al. 2004), including the development of reading skills (e.g., Guthrie et al. 2012).

According to Fredricks et al. (2004), engagement has behavioral, emotional, and cognitive components. Behavioral engagement refers to the behaviors directly related to the learning process, such as attentiveness and persistence in completing the given assignments (Finn and Zimmer 2012). Cognitive engagement is conceptually close to behavioral engagement, but while the latter refers to observable behaviors, cognitive engagement covers internal investment in learning, that is, the learner's efforts and persistence in attempting to understand and master the given tasks, especially when these tasks are challenging, as well as self-regulation skills and the use of cognitive strategies (Finn and Zimmer 2012; Fredricks et al. 2004). Emotional or affective engagement refers to positive and negative emotions (such as joy, interest, boredom, etc.) the individual may experience while performing the activities (Fredricks et al. 2004). Affective engagement can be seen as the source of motivation for the students to be behaviorally and cognitively involved in activities (Finn and Zimmer 2012). However, there is less research on its importance in achievement than there is of the importance of behavioral and cognitive engagement (Fredricks et al. 2004).

Although it seems that the gaming elements that increase player engagement also have positive effects on learning, there is not much evidence from experimental studies that this is the case (Abdul Jabbar and Felicia 2015). Some studies have even indicated that children's engagement in learning with digital environments does not lead to better learning outcomes (Annetta et al. 2009; Kim et al. 2017; Ronimus et al. 2014; Wrzesien and Raya 2010). This may be because of shortcomings in the game design, but it is also possible that the entertaining features of games distract children from focusing on the subject matter (Wrzesien and Raya 2010; Zheng and Spires 2014). Kim et al. (2017) found no associations between students' engagement (emotional/cognitive and behavioral) in playing a fractions game and their mathematics performance. The researchers reasoned that this may have been because of the difficulty of the learning content; the students were not

able to improve their math achievement despite the level of engagement. An earlier GL study found that although a rewarding system embedded in the game seemed to increase children's engagement (measured by playing time) during the first sessions of gameplay, the rewards had no overall positive effect on children's learning (Ronimus et al. 2014). The features used to increase student engagement may only have a novelty effect that wears off during extended practice. It is difficult to draw conclusions of this matter because of the paucity of previous studies, but the studies mentioned above have indicated that the relationship between engagement and learning may be more complicated in game-based learning than in school environments. Games may be engaging (especially emotionally) because of their entertaining features, but these features may distract the players from learning the content (Wrzesien and Raya 2010; Zheng and Spires 2014).

The present version of GL supports player's engagement with a labyrinth-like fantasy world and an avatar (besides the adaptation system that aims to achieve an optimal challenge level). The player moves the avatar in the game world (Fig. 2), and each encountered learning task is a "block" in the labyrinth, so completing the task clears the block away, allowing the avatar to move forward. Therefore, the player cannot progress in the game world without completing the learning tasks, hence integrating the mechanics of gameplay and learning, which is expected to support skill development (Ke and Abras 2013). In addition, GL rewards the player for completing the learning tasks by giving virtual accessories and clothes that the player can use to personalize his or her avatar. Virtual characters and avatars have been found to have several positive effects in serious games, such as increase of motivation, immersion, and cognitive engagement (Abdul Jabbar and Felicia 2015).

In the present study, we used the concept of cognitive engagement to describe effort and persistence (mental investment) during training with GL, whereas behavioral engagement referred to the time the child was involved in completing the learning tasks of the game (exposure time). The concept of emotional engagement was used to describe the child's



Fig. 2 In GL, the player, utilizing an avatar, navigates through a labyrinth. The goal is to reach the magic door on the bottom of the screen and gain access to the next level by completing learning tasks embedded in various fantasy contexts. The player is rewarded by accessories that can be used to personalize the avatar. The accessories are hidden in blocks marked by a question mark, and the found accessories appear on the panel on the right side of the screen

attitude toward playing and willingness to continue using GL. Because self-report measures are somewhat problematic with young children, for example, because of children's difficulties in understanding the terminology (Fulmer and Frijters 2009) and suggestibility (Borgers et al. 2000), we used both self-report measures and teachers' and parents' evaluations to assess the children's cognitive and emotional engagement. We also utilized the information recorded in the game logs about the child's in-game performance, namely the success rate (percentage of correct trials) to complete the assessments. The children's success rate in the game could be an important mediator between engagement and learning; if engagement actualizes as good performance in game tasks, the learning could be more effective, potentially via increased experiences of success.

Research questions

The rationale behind the current study was to find out whether a short intervention of 6 weeks with GL influences the reading development of struggling readers. The present study extends the previous studies conducted with GL and other serious games by focusing on second graders who, first, have moderate to severe disabilities in reading accuracy and/or fluency (i.e., belonging to the lowest 14th percentile of their grade level; see Galuschka et al. 2014) and, second, have not been able catch up with the other students with the help of the standard support provided by school (remedial reading lessons and part-time special education) in first grade. Previous studies (Heikkilä et al. 2013; Hintikka et al. 2008; Lovio et al. 2012; Saine et al. 2010; Van de Ven et al. 2017; Van Gorp et al. 2016) have shown that games, like reading interventions in general (Galuschka et al. 2014), can have a positive effect on the reading skills of children who have mild reading difficulties. However, children with severe reading disabilities have been underexamined in previous studies addressing the effects of game-based learning.

To evaluate the effectiveness of GL, we used a randomized controlled trial design with a treatment group (GL training vs. control) as the independent measure, and pre- and posttest word reading, reading fluency, reading comprehension, and spelling skills as the dependent measures. We were especially interested in the group \times time interaction effect on skills, which, if significant, would suggest different rates of development in reading and spelling skills in the training and control groups. As a second measure of the effectiveness of GL in enhancing reading and spelling skills, we compared the pretest-posttest development of the training group to the posttest-follow-up development of the training group. In this comparison, progress in reading and spelling skills during the intervention was compared with progress during the period of normal school-provided support within the same individuals.

More specifically, we sought answers to the following questions:

1. Does playing GL result in an improvement in reading and spelling skills in struggling readers? More specifically,
 - a. Are there differences in the development of the trained word reading skill between the GL training group and the control group who received only typical school-provided support?
 - b. Is there a transfer effect to skills not directly trained in the game, namely reading fluency, reading comprehension, and spelling?

- c. How do the reading and spelling skills of the intervention participants develop during a follow-up period compared with during the intervention period?

Based on earlier studies about the direct and transfer effects of GL (Heikkilä et al. 2013; Hintikka et al. 2008; Saine et al. 2010), we created the following hypotheses:

H1 The children in the GL training group will develop faster in word reading than the children in the control group.

H2 If H1 is true, there is a transfer effect on reading fluency, reading comprehension, and spelling.

H3 The development of reading and spelling skills during the intervention period is faster than the development during the follow-up period.

In addition to studying the effects of GL training on reading, we sought answers to the following explorative questions concerning the role of engagement in GL-based learning (without setting specific hypotheses because of the paucity of previous research):

2. Is engagement during gameplay related to children's performance in GL and learning during the intervention? More specifically,
 - a. Is behavioral engagement (exposure time) related to GL performance and gains in reading and spelling skills?
 - b. Are self-reported and observed cognitive engagement related to GL performance and gains in reading and spelling skills?
 - c. Are self-reported and observed emotional engagement related to GL performance and gains in reading and spelling skills?
 - d. Does performance in GL mediate the possible effect of engagement on learning?

Method

Participants

Permission to conduct the research was obtained from the education units of the cities of Helsinki, Espoo, and Vantaa. The participants were recruited through the principals of the elementary schools located in these three cities, per the research plan reviewed by the University of Jyväskylä Ethical Committee. The participants were sought among children for whom—according to the observations of special education teachers—reading acquisition had been especially challenging and who needed a lot of support for their difficulties. Children with severe cognitive deficits (who attended special schools) were excluded.

Children whose parents returned a signed consent form ($N=49$) participated in the screening test at the end of first grade. Children who fell at or below the 14th percentile in the word list reading test (described in more detail in the measures section) according to the data collected in the First Steps Study, which followed the reading development of about 2,000 Finnish children from kindergarten to the ninth grade (e.g., Niemi et al.

2011), were selected as participants ($N=40$). These children read 15 words or less on the word list reading test; the grade level mean according to the First Steps data is 28.20 words ($SD=12.04$).

One of the participants withdrew from the study because of technical problems in using GL at home. In addition, the data of two children were discarded from the final analyses: one child did not use GL at all during the intervention, and the other used GL for 5 h also during the follow-up period when GL was not supposed to be used. Therefore, the final sample comprised 37 children (23 boys and 14 girls). All children were native speakers of Finnish. The mean age of the participants at the beginning of the second grade was 8.23 years ($SD=0.34$). The children came from 25 schools, and the number of participants from each school ranged from one to four.

According to the background questionnaire filled out by the parents, the most typical form of support that the children had received was remedial reading lessons (55.6% of children), which is a temporary form of support that is given when needed. One-third of the children had attended part-time special education, which is a more regular and long-term form of support. Three of the children had received full-time special education (8.4%), which is typically given to children with learning difficulties in several areas. Parental reports also indicated that 48.6% of the children had comorbid learning-related difficulties, most commonly in language development, attention, or motor skills.

The parents' educational level was representative of families in Finland: 18.9% of the mothers and 16.2% of the fathers had a master's degree or higher, 48.6% of the mothers and 35.1% of the fathers had a polytechnic or vocational college degree, 21.6% of the mothers and 27.0% of the fathers had a degree from vocational school or high school, and 10.8% of the mothers and 21.6% of the fathers had not obtained a degree after basic comprehensive education.

Procedure

The selection process of the participants and procedure are represented in Fig. 3. Each school was randomly assigned either to the GL intervention or control group. The assignment was made at the school level because it was considered easier for the teachers if all the participants in their school belonged to the same group. Moreover, by this choice, we tried to minimize the potential envy between the children and the teachers' or parents' inclination to give compensatory support to children not selected for the intervention. The number of children in the intervention group was 17 (five girls, 29%) and in the control group 20 (nine girls, 45%). The groups did not differ from each other in gender distribution $\chi^2(1, N=37)=0.95, p > .05$. The 6-week intervention took place at school (four children), at home (three children), or at both places (10 children), depending on the preferences of the teachers and parents. It was necessary to allow the teachers and parents to decide the place of the training sessions because of limitations in computer access in some schools and homes.

Because GL is freely available through an online service, most participants had already used an earlier version of GL before the current study. Three children in the intervention group and four children in the control group had no prior experience with GL. A lack of prior experience was not considered a problem because GL has a child-friendly user interface, and children typically grasp immediately how to navigate through the game. Also, the older version of GL did not include the labyrinth fantasy world, so the game version introduced in the current study had novel features for all participants.

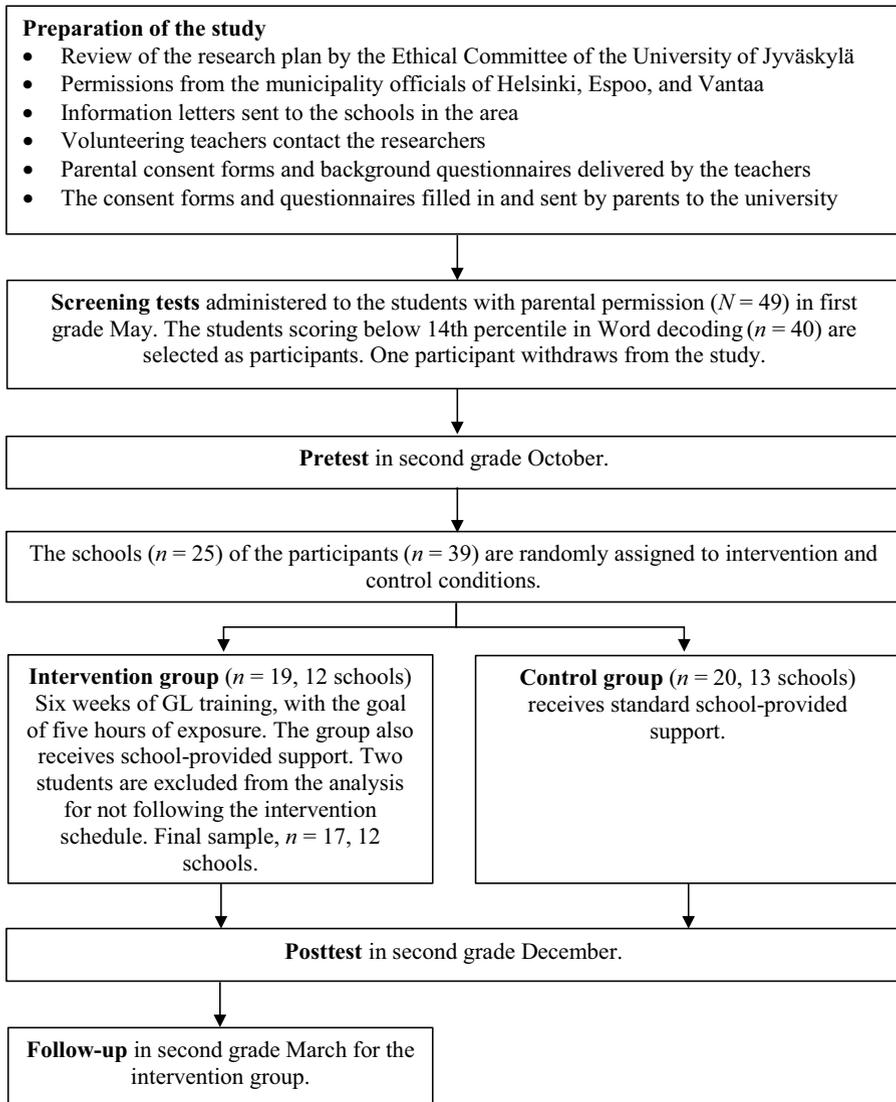


Fig. 3 A flow chart demonstrating the procedure of the study

The parents and teachers of the intervention group were sent instructions by email for installing and using GL. Technical help was also provided via email or phone when needed. The teachers and parents of the children assigned to the control group were asked not to use GL during the intervention period. Children's GL usage was monitored on the online server by the researchers to ensure that both groups followed the given instructions during the intervention.

It was recommended that each child in the intervention group should play GL at least 5 h during the 6-week intervention, that is, at least 50 min each week for sessions of about

10 min. Short sessions were recommended because of the high amount of repetition in the game, which could lead to boredom if continued for too long. Exposure times were monitored via the GL server, and feedback on the accumulated time was sent by email once a week; this was done to help the teachers and parents follow the playing schedule. If the child's accumulated exposure time was less than expected, the teacher and parents were asked to encourage the child to play more. During the intervention, the children continued to receive their usual school-provided reading support. Teachers were advised to use GL as a supplement to regular reading lessons, not to replace reading lessons with GL.

The present study included four measurement points. The screening was conducted in May (final month of first grade). The pretest was conducted in late September or early October, the posttest in December, and the follow-up in March (all in the second grade). The follow-up results are available for the intervention group only; because of ethical reasons, the control group children could begin training with GL after the posttest. The tests were administered individually at school in a separate room during school hours by trained research assistants.

Description of GraphoLearn

The GL version used in the present study focused on training early reading skills up to the level of decoding single words. The game was used with a computer. Most of the game tasks involved matching a sound (of a letter, syllable, or word) to its written equivalent shown on the screen (see Fig. 1). The learning content was organized into three categories varying in their level of difficulty. The adaptation mechanism chose the content of each trial from one of the categories based on the player's previous responses. The training started from the easiest category (connecting individual sounds and letters) and gradually—depending on the player's performance—moved on to the second category (matching spoken and written syllables) and finally on to the third category (matching spoken and written words). Each category included a specific set of items that were trained until the adaptation determined that the player had learned them. The second and third categories also included tasks in which the player built words from the given letters, but most of the training consisted of matching spoken items with their written equivalents. The game supported child's engagement with a rewarding system and an avatar (see Fig. 2).

Measures

The assessments of word reading and spelling skills were administered at the pretest, posttest, and follow-up. Engagement in GL training was assessed at the posttest.

Word reading skill

Word reading skill was assessed by three tasks to increase reliability. First, word recognition was measured with a picture–word matching task selected from a standardized reading achievement test battery (Lindeman 1998). In the task, the child worked independently with a test sheet and a pencil according to the instructions given by the test administrator. The child selected the correct word from four phonologically similar alternatives and connected the word to the matching picture by drawing a line. The task included 80 items, and the time limit was 5 min. Second, an oral word list reading was measured with a word list selected from a standardized reading and spelling test battery: Lukilasse (Häyrynen et al.

1999). In this task, the child was asked to read aloud as many words as possible from a 105-item list of increasingly difficult words. The time limit for the task was 45 s. Third, pseudoword decoding was measured with a subtest from the Finnish version of the Test of Word Reading Efficiency (TOWRE; Torgesen et al. 1999). In this task, the child was asked to read aloud as many pseudowords as possible from a 90-item list of increasing difficulty. The time limit was 45 s. In each of the three tasks, the number of correctly matched and decoded items within the time limit was the final score. For the analysis, a composite score of the word reading skill was calculated from the standardized scores of the three tasks. Cronbach's alpha reliabilities for these data were .92, .92, and .95 at the pre-, post-, and follow-up tests, respectively.

Spelling

Spelling ability was measured with a task selected from the Lukilasse test battery (Häyrinen et al. 1999). In this task, the child heard, one at a time, 20 increasingly difficult words and was asked to write them on an answer sheet. Two points were given if the word was spelled correctly. One point was given if the word contained a small error, such as missing the dot above "i," and 0 points were given if the error was obvious, such as an incorrect or missing letter. The maximum score was 40. The reliability of this task is .86 according to the test manual.

Reading fluency

The reading fluency test was based on the fluency subtasks of the Woodcock–Johnson III Tests of Achievement (Woodcock and Johnson 1989). In this test, the child was asked to read silently a list of short sentences that are either true (such as "A ball is round") or untrue ("Blueberries are yellow"). The child marked each sentence as either true or untrue. The list includes 70 sentences, and the time limit was 3 min. Because of the 50% chance of getting a correct answer by guessing, incorrect responses were subtracted from the correct responses to form the final score. The children who were unable to perform the task because of poor reading skills were given a score of 0. Cronbach's alpha reliabilities for this task were .93, .94, and .95 at the pre-, post-, and follow-up tests, respectively.

Reading comprehension

The 12-item reading comprehension task (Lerkkanen et al. 2006) included two stories with pictures. Both stories consisted of six pictures and 24 sentences. Next to each picture were four sentences telling the story, but one of the sentences included a word that was inconsistent with the picture. The child's task was to read the sentences silently and identify the inconsistent word (e.g., if the picture included a guinea pig in a box, and one of the four sentences read "The guinea pig is in a cage," the child was expected to mark the word "cage"). The time limit was 10 min. If the child marked the correct word, 2 points were given, and if the child marked other words in the correct sentence or the whole sentence, 1 point was given. The maximum score was 24. The children who were unable to perform the task because of poor reading skills were given a score of 0. Cronbach's alpha reliabilities for these data were .90, .88, and .91 at the pre-, post-, and follow-up tests, respectively.

GraphoLearn engagement

Before conducting the study, the engagement assessment scale was piloted to ensure the children understood the statements as expected, and an explorative factor analysis (using a larger dataset collected from different GL interventions) was used to determine the items assessing emotional and cognitive engagement. The assessment was administered by reading each item aloud to the child, who would respond by pointing one of five squares of different sizes printed on a sheet of paper (largest square = every time, second largest = often, middle = every now and then, second smallest = rarely, and smallest = never). The following items measured emotional engagement: (a) I enjoy playing GL, (b) I would like to play GL even more, (c) it is fun to practice reading with GL, (d) I could play GL forever, and (e) playing GL makes me happy. The following items were used to assess cognitive engagement: (a) I try my best when I play GL, (b) I choose my responses carefully, (c) I like it when a difficult task appears in the game, (d) I concentrate hard when I play, and (e) I like to read even the difficult words in GL. The Cronbach's alphas for self-reported emotional and cognitive engagement were .79 and .69, respectively.

The teachers' and parents' evaluations of the child's engagement were assessed using an online survey after the intervention. The parents and teachers were asked to respond to the questions only if they had personally observed the child's playing of the game. The items were rated using a scale from 1 (not at all) to 5 (very much). If both the teacher and parent ratings were available for the same child, the mean of the ratings was used in the analysis. The following four items were used to measure emotional engagement: (a) The child enjoyed playing GL, (b) the child would have liked to play GL more often, (c) the child played the game on his or her own initiative, and (d) the child thought GL was boring (reversed). The following three items were used to evaluate cognitive engagement: (a) The child concentrated well while playing the game, (b) the child liked the challenging tasks of the game, and (c) the child persisted even when the game tasks became difficult for him or her. The Cronbach's alphas for emotional and cognitive engagement were .77 and .82, respectively.

Behavioral engagement was measured by the exposure time retrieved from the game logs after the intervention. The exposure time consisted of the time the child was actively engaged in completing the learning tasks of the game. The time spent with the other features of the game, such as moving in the labyrinth or personalizing the avatar and pause times, were excluded.

Parent's educational level

The parents were asked to report their educational level in the background questionnaire sent to them at the beginning of the study. The educational level of the mothers and fathers was measured with a 7-point scale: 1 = unfinished comprehensive school, 2 = comprehensive school, 3 = vocational school/high school, 4 = vocational college, 5 = polytechnic, 6 = master's degree, 7 = PhD degree.

Data analysis

We began the analysis by inspecting the exposure times and parent's educational level, which might affect the interpretations of the results. Because of the small and unequal

group sizes, we used the Kruskal–Wallis test to investigate the potential differences in exposure time between the children who trained in different places, namely, only at home ($n=3$), only at school ($n=4$), or both at home and school ($n=10$). Independent samples *t* tests were used to compare the parents' educational level and initial reading and spelling skills between the training and control groups.

To compare the differences in the development of word reading skill between the GL training group and the control group, a mixed design ANOVA—with time (pre- and post-assessment) as the within-subject factor, group (training group vs. control group) as the between subject factor, and the word reading composite score as the dependent measure—was performed. Next, to see whether progress in trained word reading had a transfer effect on other reading and spelling skills, separate mixed design ANOVAs using spelling, reading fluency, and reading comprehension as the dependent measures were performed. In all these analyses, time (pre- and postassessment) was used as the within-subject factor and group (training group vs. control group) as the between subject factor. When a significant group \times time interaction was found, a paired samples *t* test was performed separately for the two groups to examine in which of the two groups the change in skill was significant. Moreover, if the change in skill was significant in both groups, a difference score was calculated by subtracting the pretest level from the posttest level, and the magnitude of the change in skill was compared between the two groups using an independent samples *t* test. Third, to compare the development in reading and spelling between the training period and the follow-up period within the training group, gain scores for word decoding, spelling, reading fluency, and reading comprehension were calculated, one for each of the periods (pre-post and post-follow-up), after which the scores were compared using paired samples *t* tests. To rule out the potential confounding effect of the training place, we used the Kruskal–Wallis test to compare the learning gains of the three groups when training at different places.

Finally, Pearson correlations were used to examine whether engagement or GL performance were related to progress in reading and spelling skills during the intervention period. When engagement was significantly associated with both GL performance and reading or spelling skills, a hierarchical linear regression analysis was used to examine whether GL performance was a mediating factor between engagement and reading or spelling skill.

Results

Exposure times

Exposure times in GL were retrieved from the server after the intervention period. On average, GL was used for 324.68 min ($SD=102.25$ min) during the 6-week intervention. The exposure times ranged from 116.18 to 549.98 min. Fourteen children (82%) reached the target exposure time of 5 h. The mean exposure time for home-only players was 323.71 min, for school-only players 321.03 min, and for players who used the game at both places 326.43 min (Kruskal–Wallis $H=.350$, $df=2$, $p=.839$).

Parents' educational level

The training group ($M=3.94$, $SD=1.68$) and the control group ($M=3.65$, $SD=1.42$) did not differ regarding fathers' education ($t=0.97$, $p=.577$), but mothers' education was

higher in the training group ($M=4.76$, $SD=1.20$) than in the control group ($M=3.70$, $SD=1.34$; $t=2.52$, $p=.016$, Cohen's $d=0.83$). However, no significant associations were found between mothers' educational level and children's word reading, spelling, reading fluency, or reading comprehension skills ($r=-.02-.05$, $p=.769-.996$). Therefore, and because of the small sample size and to avoid loss of statistical power, mother's educational level was not included as a covariate in the following analyses.

GL training and development of reading and spelling skills

No differences were found between the two groups in initial reading and spelling skills before the intervention (see Table 1). The development of the training and control groups in reading and spelling is presented in Fig. 4 and Table 1. The results showed that in the mixed design ANOVA for the word reading composite, the main effect of time was significant, $F(1, 35)=122.03$, $p<.001$, $\eta_p^2=.78$, as was the time \times group interaction, $F(1, 35)=5.46$, $p=.025$, $\eta_p^2=.14$. In both groups, the mean level of the word reading composite increased during the follow-up period (paired samples t test separately by group, $t(16)=-8.16$, $p<.001$ and $t(19)=-6.77$, $p<.001$, for the training and the control group, respectively). An independent samples t test—where the difference score measured the change between the pre- and posttest was used as the independent variable—showed that the training group developed faster than the control group $t(35)=2.37$, $p<.05$.

Next, we ran several mixed design ANOVAs to see if the progress in trained word reading had a transfer effect to other reading and spelling skills. First, in the mixed design ANOVA for spelling, the main effect of time was significant, $F(1, 35)=32.07$, $p<.001$,

Table 1 Descriptive statistics and group comparisons between the training and control groups in reading and spelling skills

Assessed skill	Training group ($n=17$)		Control group ($n=20$)		$t(35)$	p	Effect size Cohen's d^a
	M	SD	M	SD			
Pretest (September)							
Word reading	-.90	.73	-.66	.73	-1.00	.32	-0.33
Reading fluency	5.35	7.91	8.50	7.74	-1.22	.23	-0.40
Spelling	14.76	10.77	15.70	8.65	-0.29	.77	-0.10
Reading comprehension	6.00	7.12	6.60	6.27	-0.27	.79	-0.09
Posttest (December)							
Word reading	-.20	.98	-.20	.77	0.02	.99	0.00
Reading fluency	9.47	10.34	12.35	9.22	-0.90	.38	-0.29
Spelling	20.65	11.38	21.10	8.84	-0.14	.89	-0.04
Reading comprehension	11.35	7.90	10.30	6.51	0.44	.66	0.15
Follow-up test (March)							
Word reading	-.10	1.01					
Reading fluency	12.18	10.90					
Spelling	20.29	10.04					
Reading comprehension	11.65	8.13					

^aEffect sizes were estimated with Cohen's d computed using a pooled standard deviation

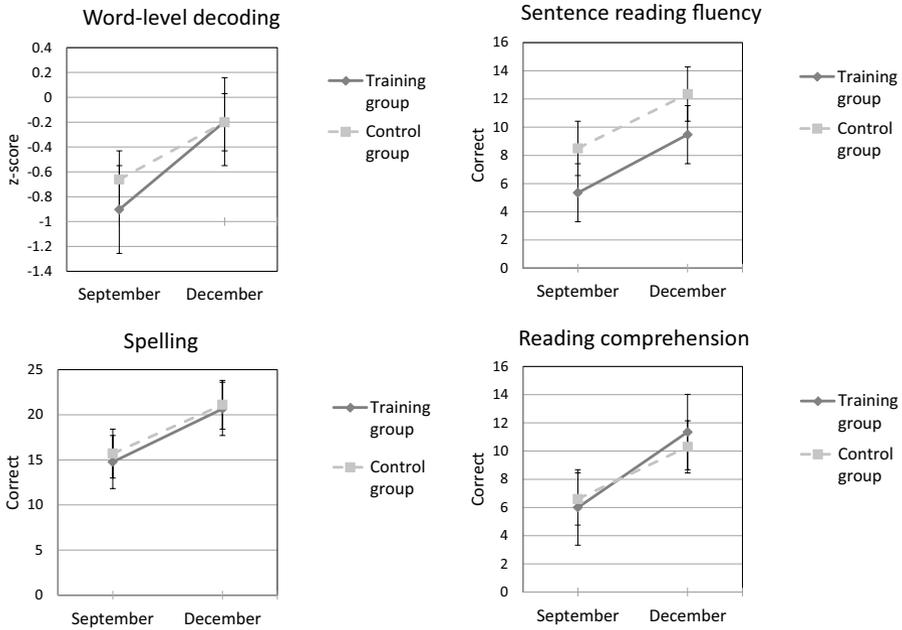


Fig. 4 The development of reading and spelling skills in the training and control groups

$\eta_p^2 = .48$, but the time \times group interaction was not, $F(1, 35)=0.06, p=.810, \eta_p^2 = .002$. Likewise, in the mixed design ANOVAs for reading fluency and reading comprehension, the main effects of time were significant, $F(1, 35)=29.92, p<.001, \eta_p^2 = .46$ and $F(1, 35)=21.46, p<.001, \eta_p^2 = .38$, respectively, whereas the time \times group interactions were not, $F(1, 35)=0.03, p=.855$, and $F(1, 35)=0.72, p=.403, \eta_p^2 = .02$, respectively. Both the training and control groups improved their performance in these three skills, but there were no differences in the rate of improvement between the groups.

The development of the training group was assessed again in the follow-up measurement about 3 months after the end of intervention. To compare the development in reading and spelling between the training period and the follow-up period, two gain scores for each

Table 2 The training group’s gain scores in reading and spelling during the training and follow-up periods

Assessed skill	Training group (n=17)				t(16)	p	Effect size Cohen’s <i>d</i> ^a
	Training period		Follow-up period				
	M	SD	M	SD			
Word reading	.71	.34	.09	.47	3.80	.002	1.51
Reading fluency	4.12	3.94	2.71	5.11	0.79	.441	0.31
Spelling	5.88	5.24	-0.35	4.54	2.84	.012	1.27
Reading comprehension	5.35	6.80	0.29	5.30	1.97	.066	1.01

^aEffect sizes were estimated with Cohen’s *d* computed using a pooled standard deviation

measure were calculated, one for each of the periods. The means, standard deviations, and effect sizes of the gains are presented in Table 2. First, the gain scores were close to zero in all measures except reading fluency during the follow-up period. The paired samples t test for word reading showed that the gain during the training period was significantly larger than during the follow-up period, and the effect size was above 0.80, which can be considered large according to Cohen (1988). In addition, in spelling, the gain during the training period was significantly larger than during the follow-up period, and the effect size in spelling was large as well. No significant difference in gain scores was found in reading fluency or reading comprehension; however, in the latter, the difference was close to significant ($p=.066$) in favor of the training period. The effect size between the gains of the training and follow-up period in reading comprehension was large.

To rule out the potential confounding effect of the training place on the learning outcomes, we compared the training period gains scores of children training at home, at school, or at both places. The Kruskal–Wallis test found no significant differences in the four gain scores between the three groups trained at different places: word reading ($H=4.22, df=2, p=.121$), spelling ($H=1.43, df=2, p=.489$), reading fluency ($H=2.68, df=2, p=.262$), and reading comprehension ($H=4.54, df=2, p=.103$).

Engagement and development of reading and spelling skills

Based on parent and teacher evaluations, there was a variation in the level of children’s emotional engagement ($M=3.57, SD=0.70, range=2.25–4.75$) and cognitive engagement ($M=3.67, SD=.80, range=2.00–5.00$) during the use of GL. Similarly, there was a variation in the level of children’s self-reported emotional engagement ($M=3.69, SD=0.98, range=2.2–5.00$) and cognitive engagement ($M=4.07, SD=1.2–5.00$). Children’s mean success rate in the game was 87.71% ($SD=6.44, range=71.22–95.08$).

First, the correlation analysis (see Table 3) revealed a few significant correlations between the different aspects of engagement. Adult-observed emotional engagement and GL exposure time were related, suggesting that children who seemed to enjoy playing GL

Table 3 Correlations between GL engagement, GL success rate, and reading and spelling gains during the training period

	1	2	3	4	5	6	7	8	9
1. GL exposure time	–								
2. GL emotional engagement (observed)	.60*	–							
3. GL cognitive engagement (observed)	-.20	-.13	–						
4. GL emotional engagement (self-reported)	.18	.54*	-.48	–					
5. GL cognitive engagement (self-reported)	-.17	.03	-.06	.59*	–				
6. GL success rate	-.17	-.24	.70**	-.32	-.08	–			
7. Word reading gain	.26	.15	.55*	-.25	-.26	.73**	–		
8. Reading fluency gain	.12	-.18	.60*	-.19	.04	.64**	.69**	–	
9. Spelling gain	.22	-.01	-.19	-.07	-.02	.16	.16	.08	–
10. Reading comprehension gain	.22	.07	.02	.16	-.27	.24	.19	.23	.10

$n=17$, except in associations related to observed GL engagement, where $n=16$

* $p < .05$, ** $p < .01$

played the game more than children with a lower level of emotional engagement. Also, adult-observed and self-reported emotional engagement were significantly associated, as were self-reported emotional and cognitive engagement. The only aspect of engagement significantly related to learning gains was adult-observed cognitive engagement: a higher cognitive engagement was related to larger gains in word decoding and reading fluency. Adult-observed cognitive engagement was also associated with a higher success rate in GL.

A hierarchical regression analysis using the gain score in word decoding as the dependent measure showed that first, cognitive engagement observed by adults explained 37.4% of the gain in word decoding, $F(1, 14)=8.35, p<.05$. Second, the GL success rate did not significantly raise the percentage of explained variance in the gain of word decoding when entered into the model as the second step, $F(1, 13)=0.81, p=.38$. The standardized beta coefficients showed that neither of the independent predictors were significant when entered simultaneously in the model ($\beta=0.42, p=.18$ and $\beta=0.27, p=.38$, for cognitive engagement observed by adults and GL success rate, respectively). In addition, the cognitive engagement observed by adults explained 30.6% of the gain in reading fluency, $F(1, 14)=6.16, p<.05$. Moreover, the GL success rate significantly raised the percentage of explained variance in the gain of reading fluency when entered into the model at the second step, $F(1, 13)=6.52, p<.05$. Together, these two measures explained 53.8% of the variance in the gain in reading fluency, $F(2, 13)=7.56, p<.01$. The standardized beta coefficients showed that the GL success rate fully mediated the effect of cognitive engagement observed by adults on the gain of reading fluency ($\beta=0.08, p=.78$ and $\beta=0.68, p<.05$, for cognitive engagement observed by adults and the GL success rate, respectively).

Discussion

The aim of the current study was to examine the potential of a digital game-based reading intervention designed to train letter–sound correspondence and word-decoding skills in supporting second graders who experience persistent reading difficulties. The results support our first hypothesis by showing that a 6-week intervention with GL carried out by teachers and parents (here with a mean exposure time of 5 h) seems to improve the children’s word reading skill. The results are in line with the earlier studies that have found evidence of positive effects of short game-based interventions on the reading skills of children (Heikkilä et al. 2013; Hintikka et al. 2008; Lovio et al. 2012; Van de Ven et al. 2017; Van Gorp et al. 2016). The current study extends these previous studies by showing that serious games that train reading skills can be effective also with children who have moderate to severe reading disabilities, which are often laborious to remediate using conventional instruction. We did not, however, find support for our second hypothesis. Despite the positive effect on word-level reading skill, we did not find transfer effects on spelling, sentence-level reading fluency, or reading comprehension, suggesting that in this group of children, the intervention was not more effective than school-provided support at improving skills not directly trained by the game.

The third hypothesis was partially supported. During the follow-up period, the development of word reading and spelling skills of the intervention participants was significantly slower than during the intervention. The children were nevertheless able to maintain the achieved level in reading and spelling over the 3-month period after the intervention. The within-group development implies that the training period with GL gave a boost to the

children's word reading and spelling skills, but to continue this positive development, a longer intervention would probably be required. However, because of the lack of a control group to analyze during the follow-up period, we are not able to draw conclusions regarding the long-term effects of GL interventions on children who have moderate to severe reading difficulties. This needs to be addressed in future studies.

Concerning our second research question, we found that adult-observed cognitive engagement, success rate in GL, and gains in word reading and reading fluency during the intervention were associated with each other. The success rate seemed to mediate the effect of cognitive engagement to reading fluency gain, suggesting that the children who were able to focus and persist while playing tended to have higher success rates, and having a high success rate further contributed to the development of reading fluency. In the case of word reading, the in-game success rate seemed to only partially mediate the effect of cognitive engagement on the gain in word decoding. Because the game was effective in training word decoding, it is possible that this training effect confounded the process by which engagement contributes to learning. It is important to study with larger samples the mediating role of in-game performance in future studies. Overall, the findings are in accordance with earlier research about the importance of engagement to achievement (Finn and Zimmer 2012; Fredricks et al. 2004, Guthrie et al. 2012). However, emotional engagement (adult-observed and self-reported) was not associated with learning gains. It is possible that the experience of fun in GL was mostly affected by the gaming aspect, such as rewards and avatars, which may not have contributed to learning the content. Similar findings have been made in earlier studies about game-based learning (Wrzesien and Raya 2010; Zheng and Spiers 2014).

The results also indicated that behavioral engagement, when measured by exposure time, is not a predictor of learning. This could partially be because the amount of exposure was not fully controlled by the researchers, but instead, the choices made at homes and schools affected the regularity of playing. It cannot be ruled out that children who were particularly slow learners were encouraged to play more than others, which would explain the absence of a correlation here. However, the findings concerning the importance of cognitive engagement seem to support the notion that exposure alone is not sufficient if the child's mind is not in the learning of the content. Exposure time was related to higher emotional engagement (observed by adults), suggesting that the amount of playing may be an indicator of how much the child enjoyed playing the game. In the present study, the reverse may also be true. Teachers and parents may have interpreted higher exposure time as a sign of a child's enjoyment of the game because they received regular feedback of the accumulation of the exposure time by email throughout the intervention.

We did not find any associations between children's self-reported engagement and learning gains. This may be related to problems associated with using self-report measures with young children (Fulmer and Frijters 2009; Borgers et al. 2000). Children's self-reported cognitive engagement did not correlate with adult-observed cognitive engagement, indicating that young children may not be able to reliably assess their level of concentration and effort. The correlation between children's self-reported cognitive and emotional engagement also shows that children may have difficulties in differentiating these two aspects of engagement from each other. The children may have considered themselves cognitively engaged also when working on other aspects of gaming (such as collecting rewards and personalizing the avatar) irrelevant to the learning of the content. In the case of teachers and parents, cognitive and emotional engagement did not correlate, indicating that they may see gameplay as having two separate aspects: educational and entertaining and that the child may be more oriented toward one or the other.

Practical implications

Based on the current study, GL can be recommended for teachers and parents as a supplemental tool for children's reading remediation. The game was effective in training the word decoding of children who have moderate and severe reading difficulties, but it is important to realize that these children need additional support to achieve transfer and long-term effects. GL could be used as the initial step, providing a boost to the development of basic decoding skills, and the further development of more advanced reading skills could be supported with other methods.

The study also shows that assessing engagement and in-game performance can provide valuable information for further development of the game. The present study shows that in-game performance is a potential mediator between the player's engagement and learning, and this process should be researched further using larger samples. It seems that if the game supports players' cognitive engagement, encouraging them to have good in-game performance, the training effects could be stronger. In the case of GL, a potential way to increase players' cognitive engagement could be the elaboration of the feedback the game gives the players. The present version of GL provided only immediate correct/incorrect feedback after each trial, but no information of the overall level of performance or skill development. The players progressed in the game world by completing the learning tasks, regardless of the level of performance, which may not have encouraged them to try their best. In the case of struggling learners, who often have motivational problems, preventing progress in the case of poor performance is not an ideal approach. Instead, the provision of encouraging, informative feedback showing what the players have accomplished and which areas need more practice could help direct their efforts toward mastering the learning content, as has been suggested by previous research (Butler and Winne 1995; Hattie and Timperley 2007).

Limitations

The low number of participants is the main limitation of the present study. However, we consider the sample to be generally representative of Finnish second graders who have moderate to severe reading difficulties because the participants came from 25 schools and from three cities. Another limitation is the lack of a control group in the follow-up period. The results of the training group indicate that learning was slower during typical instruction than during the intervention, but because we did not have a comparison group showing how much children who have reading difficulties normally develop during the same period of typical instruction (from second grade January to March), it is difficult to draw conclusions concerning the stability of the training effect.

Conclusion

The current study showed that a game-based intervention designed to train grapheme–phoneme correspondence skills and word decoding and that is carried out by parents and teachers can be effective in supporting children who have moderate and severe reading difficulties. This finding is encouraging because it indicates that this specific

group of children, who often show poor responses to typical school-provided support methods, can benefit from digital game-based training of reading skills. The effect emerged specifically for the word-level reading skill the game was designed to teach. The results also highlight the importance of studying different aspects of engagement in addition to skill development because an analysis of engagement can provide information that can be used to advance game development.

Acknowledgements Open access funding provided by University of Jyväskylä (JYU). We wish to thank all the children, teachers, families, and research assistants who participated in the study.

Funding This study was funded by grants from The Social Insurance Institution of Finland (Kela) and The Ella and Georg Ehrnrooth Foundation.

Compliance with ethical standards

Ethical approval All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abdul Jabbar, A. I., & Felicia, P. (2015). Gameplay engagement and learning in game-based learning: A systematic review. *Review of Educational Research*, *85*, 740–779. <https://doi.org/10.3102/0034654315577210>.
- Annetta, L. A., Minogue, J., Holmes, S. Y., & Chen, M.-T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education*, *53*, 74–85. <https://doi.org/10.1016/j.compedu.2008.12.020>.
- Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*, 621–635. <https://doi.org/10.1017/S0142716403000316>.
- Borgers, N., de Leeuw, D., & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin de Méthodologie Sociologique*, *66*, 60–75. <https://doi.org/10.1177/075910630006600106>.
- Brem, S., Bach, S., Kucian, K., Guttorm, T., Martin, E., Lyytinen, H.,... Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *PNAS*, *107*(17), 7939–7944. <https://doi.org/10.1073/pnas.0904402107>
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*, 245–281. <https://doi.org/10.3102/00346543065003245>.
- Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (2000). Early reading-related skills and performance, reading self-concept, and the development of academic self-concept: A longitudinal study. *Journal of Educational Psychology*, *92*, 703–708. <https://doi.org/10.1037/0022-0663.92.4.703>.
- Cheung, A. C. K., & Slavin, R. E. (2013). Effects of educational technology applications on reading outcomes for struggling readers: A best-evidence synthesis. *Reading Research Quarterly*, *48*, 277–299. <https://doi.org/10.1002/rrq.50>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Erlbaum.

- Eklund, K., Torppa, M., Aro, M., Leppänen, P., & Lyytinen, H. (2015). Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and 8. *Journal of Educational Psychology, 107*, 126–140. <https://doi.org/10.1037/a0037121>.
- Finn, J. D., & Zimmer, K. S. (2012). Student engagement: What is it? Why does it matter? In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 97–131). New York: Springer. <https://doi.org/10.1007/978-1-4614-2018-7>.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59–109. <https://doi.org/10.3102/00346543074001059>.
- Fulmer, S. M., & Frijters, J. C. (2009). A review of self-report and alternative approaches in the measurement of student motivation. *Educational Psychology Review, 21*, 219–246. <https://doi.org/10.1007/s10648-009-9107-x>.
- Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS ONE, 9*(2), e89900. <https://doi.org/10.1371/journal.pone.0089900>.
- Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*, 441–467. <https://doi.org/10.1177/1046878102238607>.
- Gee, J. P. (2007). *Good video games and good learning: Collected essays on video games, learning and literacy*. New York: Peter Lang Publishing.
- Girard, C., Ecalte, J., & Magnant, A. (2012). Serious games as new educational tools: How effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning, 29*, 207–219. <https://doi.org/10.1111/j.1365-2729.2012.00489.x>.
- Guthrie, J. T., Wigfield, A., & You, W. (2012). Instructional contexts for engagement and achievement in reading. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 601–634). New York: Springer. <https://doi.org/10.1007/978-1-4614-2018-7>.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112. <https://doi.org/10.3102/003465430298487>.
- Häyriäinen, T., Serenius-Sirve, S., & Korkman, M. (1999). *Lukilasse. Lukemisen, kirjoittamisen ja laskemisen seurantatutkimus peruskoulun ala-asteen luokille 1–6* [Lukilasse. Screening test of reading, writing and arithmetic for grades 1–6]. Helsinki, Finland: Psykologien Kustannus Oy.
- Heikkilä, R., Aro, M., Närhi, V., Westerholm, J., & Ahonen, T. (2013). Does training in syllable recognition improve reading speed? A computer-based trial with poor readers from second and third grade. *Scientific Studies of Reading, 17*, 398–414. <https://doi.org/10.1080/10888438.2012.753452>.
- Hintikka, S., Landerl, K., Aro, M., & Lyytinen, H. (2008). Training reading fluency: Is it important to practice reading aloud and is generalization possible? *Annals of Dyslexia, 58*, 59–79. <https://doi.org/10.1007/s11881-008-0012-7>.
- Jere-Folotiya, J., Chansa-Kabali, T., Munachaka, J. C., Sampa, F., Yalukanda, C., Westerholm, J. ... Lyytinen, H. (2014). The effect of using a mobile literacy game to improve literacy levels of grade one students in Zambian schools. *Educational Technology Research and Development, 62*, 417–436. <https://doi.org/10.1007/s11423-014-9342-9>.
- Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*, 243–255. <https://doi.org/10.1037/0022-0663.78.4.243>.
- Ke, F., & Abras, T. (2013). Games for engaged learning of middle school children with special learning needs. *British Journal of Educational Technology, 44*, 225–242. <https://doi.org/10.1111/j.1467-8535.2012.01326.x>.
- Kim, S., Chang, M., Deater-Deckard, K., Evans, M. A., Norton, A., & Samur, Y. (2017). Educational games and students' game engagement in elementary school classrooms. *Journal of Computers in Education, 4*, 395–418. <https://doi.org/10.1007/s40692-017-0095-4>.
- Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG Rime and GG Phoneme. *Reading Research Quarterly, 48*, 61–76. <https://doi.org/10.1002/rrq.038>.
- Lepper, M. R., & Malone, T. W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction: III. Cognitive and affective process analyses* (pp. 223–253). Hillsdale, NJ: Erlbaum.
- Lerkanen, M.-K., Poikkeus, A.-M., & Ketonen, R. (2006). *ARMI. Luku- ja kirjoitustaidon arviointimateriaali 1. Luokalle* [Reading and writing assessment material for the 1st grade]. Helsinki, Finland: WSOY.

- Lerkkanen, M.-K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J.-E. (2004). Predicting reading performance during the first and the second year of primary school. *British Educational Research Journal*, *30*, 67–92. <https://doi.org/10.1080/01411920310001629974>.
- Lindeman, J. (1998). *ALLU—Ala-asteen lukutesti [ALLU—Reading test for elementary school]*. Turku: Turun yliopisto, Oppimistutkimuksen keskus.
- Lovio, R., Halttunen, A., Lyytinen, H., Näätänen, R., & Kujala, T. (2012). Reading skill and neural processing accuracy improvement after a 3-hour intervention in preschoolers with difficulties in reading-related skills. *Brain Research*, *1448*, 42–55. <https://doi.org/10.1016/j.brainres.2012.01.071>.
- Lyytinen, H., Erskine, J., Ahonen, T., Aro, M., Eklund, K., Guttorm, T., ... Viholainen, H. (2008). Early identification and prevention of dyslexia: Results from a prospective follow-up study of children at familial risk for dyslexia. In G. Reid, F. Manis, & L. Siegel (Eds.), *The Sage handbook of dyslexia* (pp. 121–146). Thousand Oaks, CA: Sage.
- Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*, 267–296. <https://doi.org/10.1037/a0021890>.
- Morgan, P. L., & Fuchs, D. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation? *Exceptional Children*, *73*, 165–183. <https://doi.org/10.1177/001440290707300203>.
- Mugnaini, D., Lassi, S., La Malfa, G., & Albertini, G. (2009). Internalizing correlates of dyslexia. *World Journal of Pediatrics*, *5*, 255–264. <https://doi.org/10.1007/s12519-009-0049-7>.
- Niemi, P., Nurmi, J.-E., Lyyra, A.-L., Lerkkanen, M.-K., Lepola, J., Poskiparta, E., et al. (2011). Task avoidance, number skills and parental learning difficulties as predictors of poor response to instruction. *Journal of Learning Disabilities*, *44*, 459–471. <https://doi.org/10.1177/0022219411410290>.
- Ojanen, E., Ronimus, M., Ahonen, T., Chansa-Kabali, T., February, P., Jere-Folotiya, J., ... Lyytinen, H. (2015). GraphoGame—A mobile solution to global illiteracy crisis. *Frontiers in Psychology*, *6*(671), 1–13. <https://doi.org/10.3389/fpsyg.2015.00671>
- Pennington, B. F., & Bishop, D. V. M. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology*, *60*, 283–306. <https://doi.org/10.1146/annurev.psych.60.110707.163548>.
- Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.
- Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., et al. (2003). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain*, *126*, 841–865. <https://doi.org/10.1093/brain/awg076>.
- Richardson, U., & Lyytinen, H. (2014). The GraphoGame method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read. *Human Technology*, *10*, 39–60. <https://doi.org/10.17011/ht/urn.201405281859>.
- Ronimus, M., Kujala, J., Tolvanen, A., & Lyytinen, H. (2014). Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge. *Computers & Education*, *71*, 237–246. <https://doi.org/10.1016/j.compedu.2013.10.008>.
- Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., ... Salinas, M. (2003). Beyond Nintendo: Design and assessment of educational video games for first and second grade students. *Computers & Education*, *40*, 71–94. [https://doi.org/10.1016/S0360-1315\(02\)00099-4](https://doi.org/10.1016/S0360-1315(02)00099-4)
- Saine, N. L., Lerkkanen, M., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2010). Predicting word-level reading fluency outcomes in three contrastive groups: Remedial and computer-assisted remedial reading intervention, and mainstream instruction. *Learning and Individual Differences*, *20*, 402–414. <https://doi.org/10.1016/j.lindif.2010.06.004>.
- Savolainen, H., Ahonen, T., Aro, M., Tolvanen, A., & Holopainen, L. (2008). Reading comprehension, word reading and spelling as predictors of school achievement and choice of secondary education. *Learning and Instruction*, *18*, 201–210. <https://doi.org/10.1016/j.learninstruc.2007.09.017>.
- Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). New York: Routledge.
- Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*, 571–581. <https://doi.org/10.1037/0022-0663.85.4.571>.
- Torgesen, J. K. (2004). Avoiding the devastating downward spiral. The evidence that early intervention prevents reading failure. *American Educator*, Fall 2014. Retrieved from <https://www.aft.org/periodical/american-educator/fall-2004/avoiding-devastating-downward-spiral>.
- Torgesen, J., Wagner, R., & Rashotte, C. (1999). *Test of word reading efficiency. TOWRE*. Austin, TX: PRO-ED.

- Van de Ven, M., de Leeuw, L., van Weerdenburg, M., & Steenbeek-Planting, E. G. (2017). Early reading intervention by means of a multicomponent reading game. *Journal of Computer Assisted Learning, 33*, 320–333. <https://doi.org/10.1111/jcal.12181>.
- Van Gorp, K., Segers, E., & Verhoeven, L. (2016). Enhancing decoding efficiency in poor readers via a word identification game. *Reading Research Quarterly, 52*, 105–123. <https://doi.org/10.1002/rrq.156>.
- Willcutt, E. G., & Pennington, B. F. (2000). Comorbidity of reading disability and attention-deficit/hyperactivity disorder: Differences by gender and subtype. *Journal of Learning Disabilities, 33*, 179–191. <https://doi.org/10.1177/002221940003300206>.
- Woodcock, R., & Johnson, M. B. (1989). *Woodcock–Johnson test of academic achievement—revised*. New York: Riverside Publishing.
- Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*, 249–265. <https://doi.org/10.1037/a0031311>.
- Wrzesien, M., & Raya, M. A. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-junior project. *Computers & Education, 55*, 178–187. <https://doi.org/10.1016/j.compedu.2010.01.003>.
- Zheng, M., & Spires, H. A. (2014). Fifth graders' flow experience in a digital game-based science learning environment. *International Journal of Virtual and Personal Learning Environments, 5*, 69–86. <https://doi.org/10.4018/ijvple.2014040106>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Miia Ronimus is a Post-Doctoral Researcher at the Niilo Mäki Institute, Jyväskylä, Finland. Her research interests include digital game-based learning, student motivation, and dyslexia.

Kenneth Eklund is a Senior Lecturer in the Faculty of Education and Psychology at the University of Jyväskylä, Finland. His research interests include learning difficulties, especially dyslexia, development of cognitive skills, and the associations between cognitive skills and learning difficulties.

Laura Pesu is a Project Researcher at the Department of Teacher Education, University of Jyväskylä, Finland. Her research interests include educational psychology and student motivation.

Heikki Lyytinen is an emeritus professor of Developmental Neuropsychology and now UNESCO Chair on Inclusive Literacy Learning for All (2015–2019; <https://agoracenter.jyu.fi/projects/unesco-chair>), at the University of Jyväskylä, Finland. His main research interest is the development of scientifically validated digital learning environments to support acquisition of basic scholastic skills, most specifically reading skills.