Muotoiltu: Ei väliä, Riviväli: 1

Abstract

We used a randomised controlled trial to investigate if a mobile game, GraphoLearn (GL) could effectively support the learning of first graders (*N* = 70), who have severe difficulties in reading and spelling. We studied the effects of two versions of the game: GL Reading, which focused on training letter-sound correspondence and word reading, and GL Spelling, which included additional training in phonological skills and spelling. Children trained with tablet computers which they could carry with them during the six-week intervention. The average exposure time to training was 5 hours 44 minutes. The results revealed no differences in the development of reading or spelling skills between GL players and the control group. However, children's pre-training self-efficacy moderated the effect among GL Reading players: children with high self-efficacy developed more than the control group in word reading fluency, whereas children with low self-efficacy developed less than the control group in spelling.

Keywords: serious game, mobile learning, self-efficacy, reading, GraphoLearn

A mobile game as a support tool for children with severe difficulties in reading and spelling

Introduction

Kommentti [A1]: Why this study matters?

Serious games have become increasingly popular in supporting children's preliteracy (e.g., Kegel, van der Kooy-Hofland, & Bus, 2009; Samur, 2019) and reading skills (e.g., van Gorp, Segers, & Verhoeven, 2016; van de Ven, de Leeuw, Weerdenburg, & Steenbeek-Planting, 2017). In game-based learning the content and level of difficulty can be designed to adapt to each child's individual needs, and child-friendly, easy-to-use interfaces enable independent practice, which is expected to save teachers' resources. In practice, serious games often fail to meet the expectations (McTigue & Uppstad, 2019), and the scarcity of Randomized Controlled Trials (RCT) makes it difficult to draw conclusions about the effects of games on children's learning (Hainey, Connolly, Boyle, Wilson, & Razak, 2016). Moreover, despite the growing popularity of mobile learning apps, very few studies so far have evaluated the impact of mobile games on children's reading skills; in the review of Jamshidifarsani, Garbaya, Lim, Blazevic, and Ritchie (2019) only two out of 32 reviewed programs utilized a tablet computer or a smartphone.

Persistent difficulties in reading do not only affect child's performance in classroom but may also lead to lower educational attainment and higher risk for unemployment in adulthood (Eloranta, Närhi, Eklund, Ahonen, & Aro, 2019). Therefore, interventions for prevention of reading difficulties and their negative consequences are of utmost importance. Although technology-based approaches are potentially effective in supporting struggling readers, more research is needed of the effects of mobile devices and of the appropriate design of game-based interventions (Jamshidifarsani et al., 2019). Also, mobile learning requires ability and motivation to self-regulate one's learning (Grant, 2019; Sha, Looi, Chen, & Zhang, 2012), which raises a need for research of the role of learner characteristics in this kind of learning setting.

The present study aims to extend knowledge concerning the impact of mobile game-based learning on the reading skill of first graders with severe difficulties in reading acquisition~~children's reading skill~~. In this study learning does not occur in a predetermined location, but the children use the game in various times and places with their personal tablet computers. Regarding learner characteristics, we focus on self-efficacy, which has been associated with persistence and effort in achievement situations (Bandura, 1997; Zimmerman, 2005), and investigate its effect on the learning outcomes.

~~This kind of mobile learning requires ability and motivation to self-regulate one's learning (Grant, 2019; Sha, Looi, Chen, & Zhang, 2012), and is, which may increase likely to the role of learner characteristics such as self-efficacy. Previous studies have not, however, addressed its importance in mobile game-based learning. The present study explores this issue The role of learner characteristics is investigated by evaluating children's self-efficacy beliefs and how they affect the outcomes of mobile game-based learning.~~

Literature review

Kommentti [A2]: it would be helpful if after each synthesis they also add a few lines explicitly making a case for why a certain gap matters.

GraphoLearn (GL, previously known as GraphoGame), currently implemented in more than 20 countries (see grapholearn.info), was originally designed as a support tool for Finnish-speaking children who have a risk for dyslexia (a specific reading disability). According to the phonological theory, the main reason for dyslexia is an impairment in the representation, storage, and/or retrieval of speech sounds (Ramus, 2003). GL aims to improve the quality of these representations by presenting speech sounds pronounced by human voice, and by helping the learner associate them with corresponding letters. Intensive repetition is expected to improve the storage and retrieval of speech sounds. The game (see Figure 1) uses the synthetic phonics approach, starting from the phonemes and letters that are visually and phonetically distinct (such as a, s, t), and gradually progressing to more confusable phonemes (e.g., m, n, l) (Richardson & Lyytinen, 2014). After the player has mastered the correspondences between sounds and letters, the game starts to introduce larger units, that is, syllables and words.

Insert Figure 1 here

However, a recent meta-analysis including 19 experimental studies found no overall effect of GL on children's word reading skill (McTigue, Solheim, Zimmer, & Uppstad, 2020). Although several GL studies have reported effects on sublexical skills (Heikkilä, Aro, Närhi, Westerholm, & Ahonen, 2013; Hintikka, Landerl, Aro, & Lyytinen, 2008; Huemer, Landerl, Aro, & Lyytinen, 2008; Lovio, Halttunen, Lyytinen, Näätänen, & Kujala, 2012; Patel, Torppa, Aro, Richardson, & Lyytinen, 2018; Rosas, Escobar, Ramírez, Meneses, & Guajardo, 2017), it seems that transfer to word-level reading is rare. More specifically, transfer effect seems especially weak to words and sublexical items which have not been trained by the game, whereas as improvement has been observed in trained items (Heikkilä et al., 2013; Hintikka et al., 2008; Huemer et al., 2008).

Kommentti [A3]: e.g., p. 3 para 2 regarding the meta-analysis of 19 experimental studies. Tell the reader explicitly why those gaps are important and then share how your study is addressing them.

Kommentti [A4]: it would be helpful to mention studies that are contrary to your hypotheses in the literature review, and then explain why you chose to go a certain way. For e.g., on p. 16 the authors do mention that there wasn't a significant finding for the research question which was consistent with McTigue et al. (2020).

The versions of GL used in the previous studies did not explicitly teach the player how to combine phonemes into words, or how to segment words into phonemes, which may at least partially explain why GL has generally not been effective at improving word-level decoding skill. To find out whether this is the case, the present~~This~~ study investigates if the effect of GL on word-level reading fluency could be enhanced by tasks that provide explicit training in phonological awareness (PA). PA refers to the understanding that words consist of individual phonemes and that the phonemes can be blended, segmented and manipulated to create new words. PA is an important predictor of the development of reading and spelling skills (e.g., Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh, & Shanahan, 2001; Landerl & Wimmer, 2008; Melby-Lervåg, Lyster, & Hulme, 2012; Puolakanaho et al., 2008). The new game version, “GL Spelling”, includes tasks involving completion of a word missing the initial or final letter, completion of a word with a missing syllable, or building of a syllable/word by using the letters shown on the screen (see Figure 2). The tasks are expected to improve children’s phonemic manipulation and segmentation skills. From the child’s perspective, the tasks resemble spelling tasks, which is why the game version is called “GL Spelling”, in contrast to the standard version, which we call “GL Reading” in this paper. Both game versions include “traditional” GL tasks where the player learns to associate spoken and written items in multiple-choice trials (Figure 1), but only GL Spelling also includes PA tasks, which require active manipulation of individual phonemes and letters. Several earlier studies conducted in non-game-based learning environments suggest that including PA and spelling training in early reading instruction is effective at improving children’s reading and spelling skills (Ehri & Wilce, 1987; Santoro, Coyne, & Simmons, 2006; Uhry & Shepherd, 1993; Vandervelden & Siegel, 1997). Concerning game-based studies, Elimelech & Aram (2019) found that a digital spelling game improved preschoolers early literacy skills, especially when the tasks included auditory and visual support, which helped children hear and see how words can be segmented into smaller phonological units. Also, Görgen, Huemer, Schulte-Körne, and Moll (2020) evaluated the impact of a multicomponent digital game including training in PA, phoneme-grapheme mapping, and word reading, on the reading skills of second and third graders with poor reading skills. The training took place at home with tablet computers, and it improved the reading of the words trained by the game, and marginally also the reading of untrained words.

Taken together, previous studies suggest that GL requires further development to improve its effect on word reading (outside the game context). For the present study we designed a new game version including training phonological skills and spelling, to see if it would enhance transfer to word-level reading. We compare this version to the standard version of the game and to typical support offered by the school. Unlike in previous studies, we use mobile versions of the game, to see if there is an additional benefit in the opportunity to use the game without restrictions related to time and place of learning.

Insert Figure 2 here

Self-efficacy and game-based learning Self-efficacy refers to confidence in one's ability to succeed in prospective performance situations (Bandura, 1997). It affects one's thoughts and feelings in these situations, and the consequent level of effort and perseverance (Bandura, 1997; Zimmerman, 2005). Self-efficacy can be seen as a part of the self-regulative cycle, where self-motivational beliefs precede self-control and self-monitoring (such as attention focusing and task strategies) during performance, and after performance self-reflection further modifies self-efficacy and future courses of action (Zimmerman, 2005). According to a recent meta-analysis, self-efficacy has a reciprocal relationship with achievement, with self-efficacy both predicting achievement and being modified by past experiences in achievement situations (Talsma, Schüz, Schwarzer, & Norris, 2018). A recent Finnish study has shown that self-efficacy consistently predicts primary school children's reading fluency development across grade levels (Peura et al., 2019).

Muotoiltu: Sisennys:Ensimmäinen rivi: 0 cm

Students who face difficulties in learning are likely to receive negative feedback and experience failure, which are known to undermine self-efficacy (Bandura, 1997). Previous studies suggest that poor readers tend to have more negative self-efficacy beliefs than normal readers (Carroll & Fox, 2017; Lee & Zentall, 2012) and the effect of self-efficacy on achievement may be stronger for learning disabled students than for normal learners (Multon, Brown, & Lent, 1991). Self-efficacy may also affect how children respond to reading interventions, possibly because children with high self-efficacy are more persistent and willing to try than children with low self-efficacy (Cho, Roberts, Capin, and Roberts, 2015).

The role of children's self-efficacy in game-based learning settings is not well known. Some previous studies indicate that self-efficacy probably has a similar, predictive role in computer-based environments as in non-technological environments. These studies indicate that self-efficacy predicts achievement in a computer game-based assessment (Bergey, Ketelhut, Liang, Natarajan, & Karakus, 2015), and the effectiveness by which students use a guidance system in a virtual learning environment and consequently benefit from the training (Nelson & Ketelhut, 2008).

Previous research suggests that learner characteristics related to self-efficacy, namely self-regulation and executive functions, affect the outcomes of children's game-based learning. Kegel et al. (2009) found that children who had regulatory difficulties did not benefit from independent training with a web-based program training phoneme skills (Living Letters). These children demonstrated more problematic computer behaviour (clicking often, manipulating the mouse, making mistakes) during training than other children. Children with normal regulatory skills benefited from the program. In another study Kegel and Bus (2012) found that a tutor providing individualized feedback increased the effectiveness of Living Letters in 4-year-old children who had low inhibitory control. Also, Van de Sande, Segers, and Verhoeven (2016) found that adding support for executive functions seemed to benefit children who practiced preliteracy skills independently with a computer game. Finally, Ronimus, Eklund, Pesu, & Lyytinen (2019) found that high cognitive engagement while playing GraphoLearn was associated both with higher response accuracy in the game and higher gains in reading fluency.

Taken together, previous studies suggest that learner characteristics related to ability to self-regulate during training, can significantly impact the outcomes of independent game-based practice of reading-related skills. Learner characteristics may predict what kind of response strategies the learner uses and what is his/her level of performance in the game, which then contributes to the learning outcomes. These characteristics could be especially relevant when students have learning difficulties: only those who persist in the face of failure and negative feedback, may be able to focus on the training and gain from it. The present study investigates this issue by evaluating children's self-efficacy before the intervention and by testing a model, where self-efficacy acts as a moderator of learning outcomes. Information about children's learning performance during the intervention is retrieved from game log files and the potential mediating role of game performance explaining the relationship between self-efficacy and learning outcomes is tested.

Present study This study aims to advance knowledge concerning the effects of mobile game-based learning in supporting first graders who have severe difficulties in learning to read and spell. The present study differs from the earlier GL studies especially by increased flexibility in playing times and places enabled by the use of personal tablet computers during the intervention. By this arrangement we aimed to increase the intensity of playing, which could positively affect the learning gains (McGinty, Breit-Smith, Fan, Justice, & Kaderavek, 2011; Ukrainetz, 2009).

Previous GL studies suggest that the standard version of GL focusing on training letter-sound connections and reading may not be effective at improving word-level reading skill (see McTigue et al., 2020). In the present study we prepared a new version, GL Spelling, to see if more explicit phonological training would facilitate transfer to word-level reading. Therefore, the first purpose of the study is to investigate if playing either the new version (GL Spelling) or the standard version (GL Reading) for six weeks during the spring semester of first grade would speed up the development of reading and spelling skills, in comparison to typical support (not including GL) provided by the school.

The second purpose of the study is to examine the effect of self-efficacy on children’s responsiveness to the GL training. To our knowledge, previous studies have not addressed the role of self-efficacy in children’s game-based learning of literacy skills. However, previous studies concerning self-regulation and engagement (Kegel et al., 2009; Kegel & Bus, 2012; Ronimus et al., 2019; van de Sande et al., 2016) suggest that factors related to persistence and effort may be important in this kind of training. We also aim to identify potential mediators between self-efficacy and learning by utilizing the game log data, namely information about response times and response accuracy.

In sum, we aim to respond to the following research questions, specifically for each game version:

1. Does a game-based intervention with GL on tablet computers improve first graders’ reading and/or spelling skills in comparison to typical support?
2. Does self-efficacy moderate the effect of GL on the improvement of reading and spelling skills?
3. Do mean response time and response accuracy mediate the potential effects of self-efficacy on the development of reading and spelling skills during the intervention?

Method

This study is a part of a larger research project called “Dyslexia: Genes, brain functions and interventions” (DysGeBra), coordinated by the Niilo Mäki Institute in Jyväskylä, Finland. The project focuses on the causes and remediation of dyslexia, combining gene study, brain imaging and behavioural assessments. The research plan of the DysGeBra study was reviewed and approved by the Ethical Board of the Central Finland Health Care District. Permissions for conducting the research at schools were obtained from the cities of Jyväskylä, Helsinki, Espoo, Vantaa and Järvenpää and from the principals of the participating schools. The present article reports a sub-study of DysGeBra conducted in spring 2017.

Participants

The recruitment of the participants started in the fall semester, when first grade teachers were approached by an information letter sent to GL users' and Niilo Mäki Institute's newsletter email lists, both of which include thousands of teachers across Finland. The information letter told about the upcoming DysGeBra study and encouraged teachers to use a non-mobile computer version of GL with students who they considered as being at severe risk for dyslexia. The aim was to exclude children whose difficulties could be resolved by a brief period of GL training.

The teachers who had used the GL version or who had otherwise responded to the information letter, were sent information letters and consent forms in the end of fall semester to be delivered to the guardians of these students. Exclusion criterions were hearing, sight and severe cognitive deficits, and Finnish not being the mother tongue. Consent forms signed by the guardian and the child were required from all participants.

Pretests were administered to all participants (*N*=100) in January and February of Grade 1. The selection criterion for the intervention was a score in the lowest 5% of the first grade age group in either reading or spelling. This criterion was fulfilled by 70 children (48 boys, 22 girls). Boys were clearly overrepresented in the sample, which is probably because reading difficulties are more common among boys than girls (Rutter et al., 2004). Data collected in another Finnish study, ReadAll (Hautala et al., 2020), was used as the reference in defining the criterion scores. The mean age of the participants at the time of the pretest was 7.64 years (*SD* = 0.37).

The children came from 29 schools, with the number of children per school ranging from 1 to 6. Half of the children (51.4%) lived in the Jyväskylä region in Central Finland and the other half (48.6%) in South Finland, in Helsinki or nearby cities. Parents' educational level was fairly representative of the educational level of Finnish parents (Official Statistics of Finland, 2018): 26.7% of the mothers and 15.5% of the fathers had a master's or doctoral degree, 38.4% of the mothers and 29.3% of the fathers had a bachelor's degree or equivalent, 23.3% of the mothers and 36.2% of the fathers had an upper secondary education degree, and 11.7% of the mothers and 18.9% of the fathers had not obtained a degree after the basic education.

Based on teacher's reports (available for 65 of 70 children), all children received additional support for their reading and spelling difficulties besides the regular classroom instruction. Most common form of support was part-time special education, received by 69.2% of the children. The rest of the children were in full-time special education classrooms (9.2%), or received other forms of support such as remedial reading lessons or otherwise differentiated instruction in reading and spelling. Majority of the children (89.2 %) had used GL before the intervention of the present study. The intervention was supplemental and did not replace the support provided by the school.

Procedure

The children first participated in the Pretest of self-efficacy and reading and spelling skills in January or February. The Posttest was administered in May. Research assistants, who were students of special education from the Universities of Jyväskylä and Helsinki, were trained to administer the tests at schools during regular school hours.

Based on the results of the Pretest, the children fulfilling the criteria described above ($N = 70$) were selected into the GL intervention study. Because of the relatively low sample size, randomized block design was used to assign the participants in the different conditions to ensure the groups did not differ in the level of pre-intervention skills. Children were divided into six subgroups, according to the level of their reading and spelling skills (lowest 5% in reading, spelling, or both) and geographical region of their school (Central or South Finland). Within subgroups, each child was randomly assigned to one of three conditions: 1) intervention with GL Reading ($n = 23$), 2) intervention with GL Spelling ($n = 24$), or 3) Control, that is, typical support offered by school ($n = 23$). One child in the Control group did not complete the reading and spelling tests in the Post-assessment, reducing the final sample size to 69 children.

The intervention began in March. Each child in the two GL groups was provided with a 7-inch tablet computer, headphones, and a charger. Research assistants brought the equipment to the schools and teachers delivered them to the children. At the end of the intervention Posttest was administered and the equipment was collected from the teachers.

The game versions

GL Reading This game version consisted of tasks which had been used in the previous versions of the game. In these tasks the child hears a speech sound, syllable, or word and tries to find the corresponding letter, written syllable or written word from the alternatives shown on the screen (Figure 1). After an incorrect answer, corrective feedback is given and the trial is repeated. The difficulty level of the content adapts to the player's skill level. In the present study, based on the game log data, each player spent on average 38% of the playing time practicing letter-sound correspondences and 62% practicing with syllables and words.

GL Spelling This game version was like GL Reading in every other way except that besides letter-sound correspondence and word reading tasks, it also included tasks training phonological and spelling skills. The tasks were either word building or word completion tasks (Figure 2). The tasks started from easy ones (combining two letters to build a syllable) and progressed to more challenging word completion and word forming tasks. The phonological/spelling tasks appeared intermittently amongst the other tasks. According to the game log data, phonological/spelling practice eventually covered approximately 40% of the training time (varying from 24% to 58% between individuals). The rest of the playing time consisted of the letter-sound correspondence (33%) and syllable and word reading practice (27%). Therefore, the amount of letter-sound correspondence training was fairly equal between the two versions, but phonological/spelling practice replaced a large part of the “traditional” syllable/word-level reading tasks in the Spelling version.

To engage the player, both game versions presented the tasks in the form of mini games embedded in a fantasy world. The player could choose between two to four mini games that differed in visuals and fantasy, but the learning content of the trials was determined by the adaptation. Completing one mini game typically took between one and two minutes. After completing each mini game, the player received five virtual coins, which the player could exchange for new clothes, hair styles and other items to personalize his/her avatar.

Measures

Word reading fluency Reading skill was assessed with three tasks. In *Word list* reading test, the child was asked to read aloud a list of words of increasing difficulty. The time limit was 2 minutes. The maximum number of correctly read words was 90. The test was a subtest from a standardized Finnish reading test battery Lukilasse 2 (Häyrinen, Serenius-Sirve, & Korkman, 2013). In *Pseudoword list* reading test the child was asked to read aloud a list of pseudowords of increasing difficulty. The time limit was 45 seconds and maximum score was 90. The pseudoword list is published as a part of a standardized Finnish reading battery (Lerkkanen, Eklund, Löytynoja, Aro, & Poikkeus, 2018). If the child read at least 5 words correctly in the *Word list* test, a third test, *Text reading*, was administered. In this test a story of 124 words was read aloud for 1 minute. The number of correctly read words within 1 minute was the final score. If the task was skipped because of child’s low reading skill, a score of 0 was given. The task is not a published test but has been used in the large-scale studies of the University of Jyväskylä, such as the Jyväskylä Longitudinal Study of Dyslexia (see Puolakanaho et al., 2008), and the First Steps study (see Nurmi et al., 2013). The scores of these three tasks were standardized by using reference values for age-level performance obtained from the data collected in ReadAll and First Steps studies. A composite variable Word reading fluency was formed by calculating the mean of the standardized scores of the three tasks. Cronbach's alpha for Word reading fluency was .85 in the Pretest and .93 in the Posttest.

Spelling Spelling skill was measured by asking the child to spell 20 words of increasing difficulty on a sheet of paper. If the child spelled the first four words incorrectly, the task was discontinued. Each spelling of a word was given a score of 0, 1, or 2. If the spelling was flawless, 2 points were given. If there was a small error, such as missing a point above "i", 1 point was given. Words with more substantial errors (e.g. one extra or missing letter) and missing responses were scored as 0. The maximum score was 40. The test was a subtest from the Finnish reading test battery Lukilasse 2 (Häyrinen et al., 2013). The scores were standardized by using reference values obtained from the data of the ReadAll study.

Self-efficacy Self-efficacy was assessed by asking the children to rate their confidence in performing successfully in various reading and spelling -related tasks. The self-efficacy scale was adapted from the scale developed by Peura et al. (2019) and it included eight questions (e.g., How certain are you that you can say the names of all letters? How certain are you that you can read long words?). The questions were read aloud to the child, who responded on a scale from 1 (totally certain I cannot do this) to 5 (totally certain I can do this). Five squares of different sizes represented the response scale, and the child pointed to the square matching his or her level of confidence. Rehearsal items were presented before the actual items to ensure the child understood the response scale. Cronbach's alpha for the self-efficacy measure was .87 in the present data.

Intervention fidelity

In the beginning of the intervention, the research assistants introduced the functions of the tablet computers and GL to the teachers. The teachers and parents were also provided with written instructions for carrying out the intervention. There were no restrictions concerning the place or time of usage, to ensure children had unlimited access to the game. The exact start and end day of the intervention were determined by the teacher and the research assistant when the tablet was delivered to the school. It was recommended that each child should aim for the total playing time of 10 hours, and play two to three sessions of 5 to 10 minutes each day. Teachers and parents were asked to monitor the accumulation of playing time by checking the timer included in the game at least once a week and enter the number on a diary sheet. They were reminded that the tablet was intended only for the child's personal use of GL, and that the child should always log in the game using his/her personal user account. A guest user profile could be created for adults and other children who wanted to try the game.

Teachers and parents were given contact details for asking help and guidance during the intervention. Generally, there were very few technical issues with the tablet computers and none that would have restricted child's access to the game during the intervention.

Based on the game log data retrieved from the tablet computers after the intervention, children had used GL approximately 621 minutes (*SD* = 312 min), of which 344 minutes (*SD* = 181) had been spent completing the game tasks (= exposure time). This large gap between the total playing time and exposure time suggests that children tended to spend time with activities unrelated to learning, such as personalizing their avatar. The children used the game frequently, the average number of sessions being 45 (*SD* = 24) during the six-week intervention, suggesting the game was used more than once a day. Because in this research design the playing times were not controlled by the researchers, there was much variation in exposure. However, because the playing times did not differ between the two versions (see Table 1), and were not significantly associated with self-efficacy or learning gains (e.g., correlations with exposure time were $r = .106$, $r = .051$, and $r = .032$ for self-efficacy, word reading fluency gain, and spelling gain, respectively), we will not investigate the role of playing times further in the present study.

Additional information about the implementation of the intervention was collected by a short questionnaire delivered to teachers and parents after the intervention. Eventually, teachers' responses were obtained for 80.9% of the children and parents' responses for 55.3% of the children. The reports suggest that all children, for whom data was available, had used the game at school. Majority (89%) had used the game also at home. Some children had also played the game during recesses (24% of the children) and in after school clubs (16%). Other places (31.6%) were also reported, such as "in a car" or "at grandmother's". In short, 95% of the children had used to game in more than one location.

Nearly all parents and teachers reported being involved in children's GL usage during the intervention, typically by reminding the children about playing the game, and by providing encouragement and positive feedback. Provision of technical help was relatively common (62% of teachers and 50% of parents did this at least sometimes), but only a few reported that they sometimes helped the children with the game tasks (10.8% of the teachers and 26.9% of parents).

Most students were highly motivated to play GL according to both teachers' (64.7%) and parents' (73.1%) observations, but it was quite typical that children's motivation decreased during the intervention (this was reported for 40.9% and 73.7% of the highly motivated children based on teachers' and parents' observations, respectively). The rest of the children showed varying levels of motivation, and only very few expressed poor motivation.

Data analysis

All the analyses were performed with IBM SPSS Statistics 25. One-way analysis of variance (ANOVA) was used to compare group means. Mixed-Design ANOVAs including Time (pretest and posttest) as the within-subjects factor and Group (GL Reading, GL Spelling, and Control group) as the between-subjects factor was used to analyze differences in reading and spelling development during the intervention.

In the moderation and mediation analyses the PROCESS macro (Field, 2018; Hayes, 2018) integrated into SPSS was used. A moderator can be defined as a variable that affects the direction and/or strength of the relationship between an independent and dependent variable, specifying when the certain effect holds; whereas a mediator at least partially accounts for the relation between an independent and dependent variable, explaining how and why such an effect occurs (Baron & Kenny, 1986). We investigated if the level of pre-intervention self-efficacy would serve as moderator, affecting the relationship between Group and pretest-possttest reading/spelling development. We used Johnson-Neyman (1936) procedure (JN) included in PROCESS to produce so called significant regions. Moderator values from the significant regions significantly predict the difference between the groups in the outcome and at the same time show how the predictor and outcome change when moderator varies. In mediator analysis we tested the indirect effect of self-efficacy on gains in word reading fluency and spelling via a third variable (mediator), which in our case was mean response time/response accuracy within the game. That is, we tested whether self-efficacy affects gains in word reading fluency and spelling by boosting mean response time and response accuracy in the game. To get the evidence of the significant indirect effect we used the modern way instead of the classical Baron and Kenny method (1986). The modern method was presented by Hayes (2018). He recommends testing the coefficient of the indirect effect directly by using the bootstrapped confidence interval. If the zero point is included in the interval, H_0 hypothesis cannot be rejected and the indirect effect is not significant. All the predictors in the moderation and mediation models were centered.

Results

Descriptives and group comparisons

Distributions of all continuous measures were inspected, first, both within the total sample (n = 70) and within each of the three groups (GL Reading, n = 23, GL Spelling, n = 24, and Control, n = 23). All distributions were normal or close to normal demonstrated by skewness / standard error of skewness -absolute values close to two. Next, gender distributions were inspected. In all three groups, there were predominantly boys, the percentages being 70%, 75% and 61% in the GL Reading, GL Spelling, and Control group, respectively. No differences were found in gender distribution between the three groups, χ^2 (2, N = 70) = 1.10, p = .58.

Descriptive statistics for the three groups in all independent measures are presented in Table 1. No differences between the three groups (GL Reading, GL Spelling, and Control group) were found in self-efficacy. Also, no differences were found in playing times or response accuracy between the two intervention groups. However, the GL Reading group had shorter mean response time compared to the GL Spelling group, which is most likely caused by differences in task types (i.e., completing trials in phonological/spelling tasks required more time than completing trials in multiple-choice tasks).

Our first research question was whether a tablet-based intervention with GL Reading or GL Spelling would improve reading or spelling of children with severe difficulties (reading/spelling in the lowest 5th percentile) more than typical support offered by schools. Descriptive statistics before and after the intervention are presented in Table 1. No differences were found between the three groups in the pretest measures of word reading fluency and spelling, that is, all groups were at the same level in the beginning of the intervention. Moreover, effect sizes between the intervention groups and control children were small. Next, we examined the development of word reading fluency and spelling in the three groups. In the Mixed-Design ANOVA for word reading fluency the main effect of Time was significant, $F(1, 66) = 152.12$, $p < .001$, $\eta_p^2 = .70$, suggesting that all three groups improved their reading between the pretest and posttest. No Time x Group interaction was found, $F(2, 65) = 0.85$, $p = .43$, $\eta_p^2 = .03$, indicating similar increase in all groups. No differences in the overall level of word reading fluency between the three groups were found, either, $F(2, 66) = 0.91$, $p = .41$, $\eta_p^2 = .03$. For spelling, the main effect of Time was significant, $F(1, 66) = 123.70$, $p < .001$, $\eta_p^2 = .65$, suggesting that all three groups improved between the pretest and posttest. No Time x Group interaction was found, $F(2, 65) = 0.35$, $p = .71$, $\eta_p^2 = .01$, indicating similar increase in spelling in all groups. No differences in the overall level of spelling between the three groups were found, either, $F(2, 66) = 0.14$, $p = .87$, $\eta_p^2 = .00$.

Insert Table 1 here

Moderator analysis

Our second research question concerned whether self-efficacy moderates the effect of GL on the improvement of reading and spelling skills. To proceed with the moderation analysis, the data were divided into two datasets, one using GL Reading data and the other using GL Spelling data. In both datasets we used the pretest measure of word reading fluency or spelling as a covariate, group (intervention vs. control) and self-efficacy as predictors, and the posttest measure of word reading fluency or spelling as the outcome measure.

In the case of GL Reading, Group x Self-efficacy interaction effect was significant with both outcome measures (see Table 2). With word reading fluency, the model predicted the sample significantly, $F(4, 40) = 9.36$, $p < .001$, explaining 48% of the variance in posttest reading score. The Group x Self-efficacy interaction, $F(1, 40) = 4.58$, $p = .038$, explained 6% of the variance in posttest reading. Children in the GL Reading intervention group, who had a centered self-efficacy score higher than 0.57, scored significantly higher than controls in posttest reading. In the original self-efficacy scale (range 1–5) this corresponds to scores higher than 4.28. In our intervention group 8 cases (35%) were in the significant region (see Figure 3). Concerning posttest spelling (Table 2), the model predicted the sample significantly, $F(4, 40) = 3.36$, $p = .018$, explaining 25% of the posttest spelling variance. The Group x Self-efficacy interaction, $F(1, 40) = 5.61$, $p = .023$, explained 11% of the variance in posttest spelling. Low self-efficacy children (centered self-efficacy score less than -0.86, or 2.68 in the original scale) scored significantly lower than controls in posttest spelling when controlling for pretest spelling (see Figure 4). In our intervention group 6 cases (26%) were in the significant region.

In GL Spelling group, the Group x Self-efficacy interaction effect was not significant when either word reading fluency or spelling was used as the outcome measure (Table 3), that is, self-efficacy did not act as a moderator in the group training with GL Spelling.

Insert Table 2 here

Insert Figure 3 here

Insert Table 3 here

Insert Figure 4 here

Mediator analysis

Our third research question concerned the possible mediating roles of mean response time and response accuracy between self-efficacy and gains in word reading fluency and spelling during the intervention. Correlation matrices were inspected first to determine the possible mediator models (see Table 4).

Insert Table 4 here

Concerning the GL Reading group, there were marginal and significant correlations which could imply mediation effects: self-efficacy was significantly associated with reading gain, and marginally significantly with mean response time and spelling gain. Response accuracy was significantly associated with spelling gain and marginally significantly with reading gain. We therefore proceeded by testing two parallel multivariate mediation models (reading gain and spelling gain as the dependent measures, at a time), with two mediators (mean response time and response accuracy) in both models. Mediators were allowed to correlate. In the GL Spelling group, self-efficacy did not correlate with any of the variables, suggesting that self-efficacy had no significant association with children's in-game performance and learning gains. Therefore, we did not test any mediation models with the GL Spelling data.

Based on both component and index approaches (Table 5), no statistically significant indirect effects between self-efficacy and reading/spelling gains were found with the GL Reading data. Closer look at the results reveals that the total effect, when self-efficacy alone was regressed on reading gain, was statistically significant. In components from self-efficacy to mediators only the path to mean response time was marginally significant. Concerning the paths from mediators to outcome, only response time was marginally significant in the case of reading gain. In the case of spelling gain, the total effect of the predictors was marginally significant and the component path only from response accuracy to spelling gain was marginally significant (Figure 5).

Insert Table 5 here

Insert Figure 5 here

Discussion

Kommentti [A5]: The authors should mention some of the studies earlier in the paper, and tie back their findings with the relevant literature as well as explain, why in their considered view, do we not see a finding consistent with their original hypothesis etc.

The purpose of this study was twofold. First, we investigated if a mobile game-based intervention (with or without explicit training of phonological awareness) was effective at improving reading and spelling skills of first graders who had severe difficulties in reading or spelling. Second, we studied whether self-efficacy moderated the effect of the intervention, and whether game-based measures (mean response time and response accuracy) served as mediators between the possible effects of self-efficacy on reading and spelling skills. The results suggest that neither the standard game version focusing solely on letter-sound correspondence and reading, nor the new game version including additional phonological tasks, was effective at improving word-level reading or spelling skills more than typical support offered by school. At the group level, the skills of the intervention groups developed at the same rate with the control group, which did not use the game. However, the moderator analysis revealed that children with high self-efficacy developed more than the control group in word reading fluency, whereas children with low self-efficacy developed less than the control group in spelling.

The observed lack of training effect is consistent with the findings of the meta-analysis by McTigue et al. (2020), suggesting that GL is generally not effective at improving word-level reading skill. Because we were interested in transfer effects on reading outside the game context, t~~T~~he assessments of the present study included only words untrained by the game. ~~Also,~~ S~~s~~everal earlier studies investigating the effects of game-based training on children's word reading fluency suggest that the effects tend to be stronger for the words and sub-lexical items trained by the game in comparison to untrained items (Görgen et al., 2020; Heikkilä et al., 2013; Hintikka et al., 2008; Huemer et al., 2008). ~~The assessments of the present study included only untrained words.~~ Contrary to our hypothesis, GL Spelling, including additional phonological tasks, did not enhance transfer to word reading or spelling. There are at least two possible explanations for this. First, it is possible that the design of the phonological training was not optimal for this group of learners. We will discuss this matter in more detail later in this section. Second, the issue with transfer may be at least partially related to a more general difficulty in applying things learnt in game context to other learning situations. McTigue and Uppstad (2019) suggest that this type of transfer could be facilitated by more explicit adult guidance and by integration of game-based practice with other reading activities at school.

Our second research question concerned the role of self-efficacy in learning with GL. The results of the moderator analysis indicated that self-efficacy moderated the intervention outcomes in the group which used the standard version of the game, GL Reading. Children with high self-efficacy benefited from the intervention more than the control group in word reading fluency, whereas children with low self-efficacy benefited from the intervention less than the control group in spelling. The first finding is in accordance with the previous studies which suggest that high self-efficacy is related to better reading skills (Carroll & Fox, 2017; Lee & Zentall, 2012; Peura et al., 2019) and can positively predict children's response to a reading intervention (Cho et al., 2015). However, it was somewhat unexpected that GL Reading players with poor self-efficacy developed less than the control group in spelling. The reason for this could be related to GL Reading not including any phonological/spelling tasks but only multiple-choice association tasks. Although these children received spelling training as a part of their reading lessons at school, it is possible that the absence of phonological/spelling tasks in the game had some kind of negative effect on their ability to process the phonological structure of words, for example by over-emphasizing whole-word identification and fast responses. It is possible that some children who are not confident in their spelling skills may have been confused by this kind of training, which would slow down their development in spelling. However, this issue requires further studying. In the GL Spelling group, self-efficacy was not associated with learning, which may be because of the differences in the task types between the versions. Phonological tasks including word building and word completion require different kind of processing than the multiple-choice association tasks, and may require other characteristics from the player.

For our third research question we hypothesized that the effect of self-efficacy on learning is mediated by children's performance in game tasks. We tested models with mean response time and response accuracy as mediators, but our hypothesis was not supported. Among GL Reading players, the correlations suggested that self-efficacy was marginally associated with faster response times, and faster response times were related to improvement in word reading fluency, which is consistent with an earlier GL study, which suggested that game performance is a mediator between child's engagement in gameplay and learning gains (Ronimus et al., 2019). However, the indirect effects from self-efficacy via game performance to learning gains were not significant, which could be because of lack of statistical power caused by small sample size. Nevertheless, the correlations suggested that response times in GL Reading were related to development of word reading fluency. Fast response time can be considered as an indicator of fast speed of processing, which has been shown to positively affect reading achievement (Catts, Gillispie, Leonard, Kail, & Miller, 2002). Therefore, encouraging children for faster responses could potentially improve the effects of GL on reading fluency. The correlations also suggested that self-efficacious children tend to be faster responders. It is possible that children who are confident in their skills hesitate less when they select answers in GL. Children with low self-efficacy may be slower and more careful, perhaps because they are sensitized to negative feedback and try to avoid it. This suggests that special attention should be given to the design of the feedback systems of serious games aimed at children who may have poor self-efficacy, such as children with learning disabilities.

Among GL Spelling players, as indicated by the correlations, only response accuracy (not response time) was related to reading and spelling gains. Self-efficacy was unrelated to in-game performance, and therefore mediation analysis was not performed. The effectiveness of GL Spelling could potentially be improved by improving children's response accuracy, for example by provision of more explicit support for the phonological segmentation of the words, as was done in the study of Elimelech and Aram (2019), where the game spoke the word first normally, and then as segmented into different word parts, while highlighting the correct position of each letter on the screen. GL Spelling relied mostly on trial-and-error method (i.e., after incorrectly completed word, the learner was asked try other letters until the word was correctly formed) instead of giving cues for the correct solution. GL Spelling also included alternating task types, which required ability to constantly change the response strategy from fast identification of the correct alternative to the careful analysis of the phonological structure of the word. This may have been too demanding for many of the players, and only those with better attention regulation skills may have succeeded in this. We did not measure children's self-regulation and executive functions during the game play in this study, but these skills may be important for the intervention outcomes, as has been suggested by previous studies (Kegel & Bus, 2012; Kegel et al., 2009).

Limitations

The sample size was small, which limits the power of the statistical analyses in this study. The results of the moderator and mediator analyses should be interpreted as suggestions of the possible mechanisms explaining children's learning processes when GL is used as a tool of mobile learning. The role of self-efficacy and its relationship to in-game performance measures and learning gains should be addressed in future studies with larger samples.

The participants of the study had a severe delay in reading or spelling development, which may limit the generalizability of the results to other groups of learners. These children are likely to need intensive long-term support (Wanzek et al., 2018), and the GL training offered in this study may have been too light to cause observable changes in these children's skills. Also, children with severely delayed reading development may be sensitized to failure and negative feedback, and for this reason the role of self-efficacy may be more prominent in their learning in comparison to normally developing children (Multon et al., 1991).

Support from teachers and adults seems to have an important role in mobile game-based learning (see e.g. McTigue et al., 2020). Our data indicate that most teachers and parents were involved in children's GL usage, by keeping track of children's playing times and encouraging children to play, but more detailed data is needed before we can conclude how much and what kind of involvement is needed from teachers and parents for children to benefit from mobile serious games.

The data suggest that intensity of playing was generally high, although there was much individual variation in playing times. It is also possible that the timing of the intervention in the late first grade spring was not ideal. Children with severe risk for dyslexia might benefit more from interventions given in the beginning of first grade, when they have not yet had much exposure to reading-related negative feedback, which are likely to affect their reading self-efficacy. The roles of timing and intensity of the intervention should be investigated by future studies.

Finally, we wish to note that the children in our study were not new to GL, but most of them had used another version of the game before the study. We cannot rule out that a "novelty effect" might have produced different results, for example by increasing the enthusiasm of children, parents and teachers. An earlier GL study suggests that novelty may briefly increase children's interest, but it tends to wear out soon without causing significant learning gains (Ronimus, Kujala, Tolvanen, & Lyytinen, 2014). However, novelty could play some role in the manner the learner characteristics, such as self-efficacy, interact with in-game performance and learning gains.

Probably the most important limitation is that the training started after children had observed that they had not learned the reading skill at the same rate than their classmates. The main goal of the training game has from its beginning been that it is preventive in the sense that learners get the help before they face the experience of not being as successful than their classmates. Because about half of the children in Finland have the basic reading skills when they enter school it is too late to start prevention several months after the start of the school. This may explain also the results what was received from the role of the self-efficiency. There is still possible that children who face severe problems in learning the basic reading skill can be helped to overcome their problem. And this may be possible only when we are not compromising their self-image as learners too much and thus making the possibility to efficiently help them very difficult. Engagement associated with learning a thing is the necessary starting point of all learning.

Muotoilu: Sisennys:Ensimmäinen rivi: 1 cm

Muotoilu: Normaali, Sisennys:Vasen: 0 cm

Conclusions

Kommentti [A6]: At the end of the conclusion section, the authors should remind the reader of their contributions to the literature.

The study highlights the challenges in designing a successful mobile game for children who have severe difficulties in learning to read. The results are consistent with the studies which suggest that transfer effect to reading fluency of untrained words is difficult to achieve with a short intervention of a few weeks. The study extends previous GL research by showing that inclusion of phonological and spellings tasks in the game did not enhance transfer effect to word reading. The study is in accordance with earlier studies by showing that self-efficacy predicts achievement, and extends knowledge~~previous research~~ by showing that self-efficacy affects the learning also in ~~outcomes of a~~ mobile game-based learning~~intervention~~. It is important to continue studying how subgroups of learners can best be supported by serious games, especially in mobile learning, and how individual differences should be considered in game design. The study also demonstrates the value of moderator and mediator analyses which can help researchers and game developers gain a better understanding of the mechanisms affecting learning, and provide useful information for the design of more effective game versions in the future.

Muotoiltu: Fontti: englanti (Iso-Britannia)

References

Bandura, A. (1997). *Self-Efficacy. The Exercise of Control.* New York: W.H. Freeman.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Bergey, B.W., Ketelhut, D.J., Liang, S., Natarajan, U., & Karakus, M. (2015). Scientific inquiry self-efficacy and computer game self-efficacy as predictors and outcomes of middle school boys' and girls' performance in a science assessment in a virtual environment. *Journal of Science Education and Technology, 24*, 696–708. doi: 10.1007/s10956-015-9558-4

Carroll, J.M., & Fox, A.C. (2017). Reading self-efficacy predicts word reading but not comprehension in both girls and boys. *Frontiers in Psychology, 7*:2056. doi: 10.3389/fpsyg.2016.02056

Catts, H.W., Gillispie, M., Leonard, L.B., Kail, R.V., & Miller, C.A. (2002). The role of speed of processing, rapid naming, and phonological awareness in reading achievement. *Journal of Learning Disabilities, 35*, 510–525. doi: 10.1177/00222194020350060301

Cho, E, Roberts, G.J., Capin, P., & Roberts, G. (2015). Cognitive attributes, attention, and self-efficacy of adequate and inadequate responders in a fourth grade reading intervention. *Learning Disabilities Research & Practice, 30*, 159–170. doi: 10.1111/ldrp.12088

Ehri, L.C., Nunes, S.R., Willows, D.M., Schuster, B.V., Yaghoub-Zadeh, Z., & Shanahan T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the

National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*, 250–287. doi: 10.1598/RRQ.36.3.2

Ehri, L.C., & Wilce, L.S. (1987). Does learning to spell help beginners learn to read words? *Reading Research Quarterly, 22*, 47–65. doi: 10.2307/747720

Elimelech, A., & Aram, D. (2019). Using a digital spelling game for promoting alphabetic knowledge of preschoolers: The contribution of auditory and visual support. *Reading Research Quarterly*. doi:10.1002/rrq.264

Eloranta, A.-K., Närhi, V., Eklund, K., Ahonen, T., & Aro, T. (2019). Resolving reading disability – Childhood predictors and adult-age outcomes. *Dyslexia, 25*, 20–37. doi: 10.1002/dys.1605

Field, A.P. (2018). *Discovering statistics using SPSS* (5th ed.). Los Angeles: SAGE Publications.

Grant, M.M. (2019). Difficulties in defining mobile learning: analysis, design characteristics, and implications. *Educational Technology Research & Development, 67*, 361–388. doi: 10.1007/s11423-018-09641-4

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*. doi: 10.1016/j.compedu.2020.103834

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

Hautala, J., Heikkilä, R., Nieminen, L., Rantanen, V., Latvala, J.-M., & Richardson, U. (2020). Identification of reading difficulties by a digital game-based assessment technology. *Journal of Educational Computing Research*. doi: 10.1177/0735633120905309

Hayes, A.F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). New York: Guilford Publications.

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. doi: 10.1080/10888438.2012.753452

Muotoiltu: suomi

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. doi: 10.1007/s11881-008-0012-7

Huemer, S., Landerl, K., Aro, M., & Lyytinen, H. (2008). Training reading fluency among poor readers of German: many ways to the goal. *Annals of Dyslexia, 58*, 115–137. doi: 10.1007/s11881-008-0017-2

Häyrinen, T., Serenius-Sirve, S., Korkman, M. (2013). *Lukilasse 2.* Helsinki: Psykologien Kustannus Oy.

Jamshidifarsani, H., Garbaya, S., Lim, T., Blazevic, P., & Ritchie, J.M. (2019). Technology-based reading intervention programs for elementary grades: An analytical review. *Computers & Education, 128*, 427–451. doi: 10.1016/j.compedu.2018.10.003

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems*. Statistical Research Memoirs, 1,* 57–93.

Kegel, C., van der Kooy-Hofland, V., & Bus, A. (2009). Improving early phoneme skills with a computer program: Differential effects of regulatory skills. *Learning and Individual Differences, 19*, 549–554. doi: 10.1016/j.lindif.2009.07.002

Kegel, C.A.T., & Bus, A.G. (2012). Online tutoring as a pivotal quality of web-based early literacy programs. *Journal of Educational Psychology, 104*, 182–192. doi: 10.1037/a0025849

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology, 100*, 150–161. doi: 10.1037/0022-0663.100.1.150

Lee, J., & Zentall, S.S. (2012). Reading motivational differences among groups: Reading disability (RD), attention deficit hyperactive disorder (ADHD), RD + ADHD, and typical comparison. *Learning and Individual Differences, 22*, 778–785. doi: 10.1016/j.lindif.2012.05.010

Lerkkanen, M.-K., Eklund, K., Löytynoja, H., Aro, M., & Poikkeus, A.-M. (2018). YKÄ – Luku- ja kirjoitustaidon arviointimenetelmä yläkouluun [YKÄ – Assessment of reading and spelling skills for upper comprehensive school]. Jyväskylä, Finland: The Niilo Mäki Institute.

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. doi: 10.1016/j.brainres.2012.01.071

McGinty, A.S., Breit-Smith, A., Fan, X., Justice, L.M., & Kaderavek, J.N. (2011). Does intensity matter? Preschoolers' print knowledge development within a classroom-based intervention. *Early Childhood Research Quarterly, 26*, 255–267. doi:10.1016/j.ecresq.2011.02.002

McTigue, E.M., Solheim, O.J., Zimmer, W.K., & Uppstad, P.H. (2020). Critically reviewing GraphoGame across the world: Recommendations and cautions for research and implementation of computer-assisted instruction for word-reading acquisition. *Reading Research Quarterly, 55*, 45–73. doi: 10.1002/rrq.256

McTigue, E.M., & Uppstad, P.H. (2019). Getting serious about serious games: Best practices for computer games in reading classrooms. *The Reading Teacher, 72*, 453–461. doi: 10.1002/trtr.1737

Melby-Lervåg, M., Lyster, S.-A.H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138*, 322–352. doi: 10.1037/a0026744

Multon, K.D., Brown, S.D., & Lent, R.W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*, 30–38. doi: 10.1037/0022-0167.38.1.30

Nelson, B.C., & Ketelhut, D.J. (2008). Exploring embedded guidance and self-efficacy in educational multi-user virtual environments. *Computer-Supported Collaborative Learning, 3*, 413–427. doi: 10.1007/s11412-008-9049-1

Nurmi, J.-E., Kiuru, N., Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Ahonen, T., … Lyyra, A.-L. (2013). Teachers adapt their instruction in reading according to individual children's literacy skills. *Learning and Individual Differences, 23*, 72–79. doi: 10.1016/j.lindif.2012.07.012

Official Statistics of Finland (2018): *Educational structure of population*. Helsinki: Statistics Finland. Retrieved from: http://www.stat.fi/til/vkour/2018/vkour_2018_2019-11-05_tie_001_en.html

Patel, P., Torppa, M., Aro, M., Richardson, U., & Lyytinen, H. (2018). GraphoLearn India: The effectiveness of a computer-assisted reading intervention in supporting struggling readers of English. *Frontiers in Psychology, 9*:1045. doi: 10.3389/fpsyg.2018.01045

Peura, P., Aro, T., Viholainen, H., Räikkönen, E., Usher, E. L., Sorvo, R., & Aro, M. (2019). Reading self-efficacy and reading fluency development among primary school children: Does specificity of self-efficacy matter? *Learning and Individual Differences, 73*, 67–78. doi: 10.1016/j.lindif.2019.05.007

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P.H.T., Poikkeus, A.-M., … Lyytinen, H. (2008). Developmental links of very early phonological and language skills to second grade reading outcomes: Strong to accuracy but only minor to fluency. *Journal of Learning Disabilities, 41*, 353–370. doi: 10.1177/0022219407311747

Ramus, F., Rosen, S., Dakin, S.C., Day, B.L., Castellote, J.M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain, 126*, 841–865. doi: 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. doi: 10.17011/ht/urn.201405281859

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. doi: 10.1007/s11423-019-09658-3

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

Rosas, R., Escobar, J.-P., Ramírez, M.-P., Meneses, A., & Guajardo, A. (2017). Impact of a computer-based intervention in Chilean children at risk of manifesting reading difficulties. *Infancia y Aprendizaje / Journal for the Study of Education and Development, 40*, 158–188. doi: 10.1080/02103702.2016.1263451

Rutter, M., Caspi, A., Fergusson, D., Horwood, L. J., Goodman, R., Maughan, B., ... Carroll, J. (2004). Sex differences in developmental reading disability. *Journal of the American Medical Association, 291*, 2007–2012. doi:10.1001/jama.291.16.2007

Samur, Y. (2019). Kes Sesi: A mobile game designed to improve kindergarteners' recognition of letter sounds. *Journal of Computer Assisted Learning, 35*, 294–304. doi: 10.1111/jcal.12331

Santoro, L.E., Coyne, M.D., & Simmons, D.C. (2006). The reading-spelling connection: Developing and evaluating a beginning spelling intervention for children at risk of reading disability. *Learning Disabilities Research & Practice, 21*, 122–133. doi: 10.1111/j.1540-5826.2006.00212.x

Sha, L., Looi, C.-K., Chen, W., & Zhang, B.H. (2012). Understanding mobile learning from the perspective of self-regulated learning. *Journal of Computer Assisted Learning, 28*, 366–378. doi: 10.1111/j.1365-2729.2011.00461.x

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences, 61*, 136–150. doi: 10.1016/j.lindif.2017.11.015

Uhry, J.K., & Shepherd, M.J. (1993). Segmentation/spelling instruction as part of a first-grade reading program: Effects on several measures of reading. *Reading Research Quarterly, 28*, 218–233. doi: 10.2307/747995

Ukrainetz, T.A. (2009). Phonemic awareness: How much is enough within a changing picture of reading instruction? *Topics in Language Disorders, 29*, 344–359. doi: 10.1097/TLD.0b013e3181c29d9e

Vandervelden, M.C., & Siegel, L.S. (1997). Teaching phonological processing skills in early literacy: A developmental approach. *Learning Disability Quarterly, 20*, 63–81. doi: 10.2307/1511215

Van de Sande, E., Segers, E., & Verhoeven, L. (2016). Supporting executive functions during children's preliteracy learning with the computer. *Journal of Computer Assisted Learning, 32*, 468–480. doi: 10.1111/jcal.12147

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. doi: 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. doi: 10.1002/rrq.156

Wanzek, J., Stevens, E.A., Williams, K.J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of Learning Disabilities, 51*, 612–624. doi: 10.1177/00222194187751

Zimmerman, B.J. (2005). Attaining self-regulation. A social cognitive perspective. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Burlington, MA, USA: Elsevier.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Table 1

Descriptive Statistics and Group Comparisons in Self-efficacy, Game measures, and in Reading and Spelling Skills

		Group					
		GL Reading *n = 23*	GL Spelling *n = 24*	Control *n = 22 – 23* [a]		Cohen's *d* [c]	
	Range	*M (SD)*	*M (SD)*	*M (SD)*	*F* [b]	GL Reading vs. Control	GL Spelling vs. Control
Self-efficacy	1 – 5	3.65 (0.98)	3.42 (0.99)	3.66 (0.96)	0.46	0.01	0.25
Game measures							
Total playing time (min)	62 – 1721	607 (260)	634 (361)	-	0.09	-	-
Exposure time (min)	30 – 1038	335 (155)	353 (205)	-	0.12	-	-
Number of sessions	5 – 140	43 (18)	46 (29)	-	0.18	-	-
Response time (s)	2.00 – 5.26	2.96 (0.74)	3.50 (0.73)	-	6.29*	-	-
Response accuracy (%)	79.10 – 96.90	91.40 (3.14)	89.91 (4.24)	-	1.85	-	-
Reading							
Pretest	- 2.07 – -0.99	-1.71 (0.26)	-1.76 (0.27)	-1.68 (0.28)	0.63	0.11	0.29
Posttest	- 2.13 – 0.05	-1.09 (0.45)	-1.28 (0.58)	-1.13 (0.47)	0.91	0.09	0.30
Spelling							
Pretest	- 2.68 – -1.14	-2.16 (0.46)	-2.22 (0.46)	-2.27 (0.38)	0.36	0.27	0.12
Posttest	- 2.68 – 0.77	-1.10 (0.89)	-1.21 (1.03)	-1.04 (0.83)	0.20	0.07	0.19

Note.

[a] Number of participants in the Control group was 23 in pretest and 22 in posttest assessment. Consequently, degrees of freedom vary in group

[b] Degrees of freedom in group comparisons with One-way ANOVA were (2, 68) in Self-efficacy, (1, 45) in Game measures, and (2, 68) in pretest and (2, 67) in posttest Reading and Spelling measures.

[c] Cohen's d calculated using the pooled standard deviation of the two groups compared at a time.

* $p < .05$

Table 2

The Results of the Moderation Analysis with GL Reading Data with Pretest Scores as Covariates

	b	*SE*	*t*	*p*	95% LLCI	95% ULCI
Reading (posttest)						
Intercept	0.72	0.34	2.13	0.039	0.26	1.41
Group [a]	0.11	0.10	1.02	0.314	-0.10	0.32
Self-efficacy pretest [b]	0.06	0.06	1.02	0.312	-0.06	0.18
Group*Self-efficacy	0.25	0.12	2.14	0.038	0.01	0.48
Reading (pretest)	1.09	0.20	5.44	<.001	0.68	1.49
Spelling (posttest)						
Intercept	0.51	0.63	0.81	0.424	-0.76	1.78
Group [a]	-0.13	0.23	-0.54	0.594	-0.60	0.35
Self-efficacy pretest [b]	0.03	0.13	0.21	0.833	-0.23	0.29
Group*Self-efficacy	0.61	0.26	2.37	0.023	0.09	1.14
Spelling (pretest)	0.71	0.28	2.52	0.016	0.14	1.28

Note.

[a] GL Reading vs. control

[b] Self-efficacy score was mean centered

Table 3

The Results of the Moderation Analysis with GL Spelling Data with Pretest Scores as Covariates

	b	*SE*	*t*	*p*	95% LLCI	95% ULCI
Reading (posttest)						
Intercept	1.23	0.35	3.54	0.001	0.53	1.93
Group [a]	0.00	0.11	0.00	0.997	-0.22	0.22
Self-efficacy (pretest) [b]	0.00	0.06	0.01	0.990	-0.12	0.12
Group*Self-efficacy	0.13	0.12	1.11	0.275	-0.11	0.37
Reading (pretest)	1.42	0.20	7.05	<.001	1.01	1.82
Spelling (posttest)						
Intercept	1.69	0.66	2.57	0.014	0.36	3.02
Group [a]	-0.24	0.24	-1.03	0.307	-0.72	0.23
Self-efficacy (pretest) [b]	-0.12	0.13	-0.92	0.365	-0.38	0.14
Group*Self-efficacy	0.26	0.27	0.99	0.326	-0.27	0.80
Spelling (pretest)	1.25	0.29	4.28	<.001	0.66	1.84

Note.

[a] GL Spelling group vs. control

[b] Self-efficacy score was mean centered

Table 4.

Correlations between Game Measures, Reading and Spelling Gains, and Self-efficacy.

	Mean response time	Response accuracy	Reading gain	Spelling gain	Self-efficacy
Mean response time	1	-.32	-.55**	-.41[a]	-.41[a]
Response accuracy	-.14	1	.38[a]	.51*	.25
Reading gain	-.21	.68***	1	.60**	.50**
Spelling gain	-.00	.62**	.68***	1	.35[a]
Self-efficacy	.04	.16	.25	-.06	1

[a] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note. The upper triangle presents the correlations within the GL Reading group and the lower triangle presents the correlation within the GL Spelling group.

Table 5

The Results of the Parallel Multivariate Mediator Model Analysis with GL Reading Data

Antecedent	Path	Coeff.	*SE*	*t*	*p*	95% Bootstrap LLCI	95% Bootstrap ULCI
			Mean Response Time				
Constant		4.09	0.57	7.20	< .001	2.62	5.70
Self-efficacy	a_1	-0.31	0.15	-2.05	.053	-0.69	0.05
			$R^2 = .17$				
			$F(1, 21) = 4.20, p = .053$				
			Response accuracy				
Constant		88.45	2.54	34.82	< .001	83.92	93.75
Self-efficacy	a_2	0.81	0.67	1.20	.244	-0.63	2.03
			$R^2 = .06$				
			$F(1, 21) = 1.44, p = .244$				
			Reading gain				
Constant		-1.20	2.02	-0.56	.560	-5.45	3.90
Self-efficacy	c'	.011	0.07	1.55	.137	-0.02	0.26
Mean Response Time	b_1	-0.18	0.10	-1.89	.074	-0.39	0.02
Response Accuracy	b_2	0.02	0.02	1.01	.325	-0.03	0.07
			$R^2 = .43$				
			$F(3, 19) = 4.71, p = .013$				
			Spelling gain				
Constant		-8.94	5.35	-1.67	.111	-21.81	2.13
Self-efficacy	c'	0.15	0.19	0.81	.426	-0.14	0.43
Mean Response Time	b_1	-0.26	0.25	-1.02	.320	-0.79	0.18
Response Accuracy	b_2	0.11	0.06	1.99	.061	-0.01	0.25
			$R^2 = .35$				
			$F(3, 19) = 3.35, p = .041$				

Figure legends

Figure 1. In GL, the player hears a phoneme and sees a set of letters on the screen. The task of the player is to choose the letter that corresponds to the spoken speech sound. Immediate feedback tells whether the choice was correct or incorrect. In the case of an incorrect response, the correct response is shown, and the trial is repeated to strengthen the correct association.

Figure 2. In GL Spelling, the target syllable or word is first spoken aloud and the child completes the word by moving the letter tiles in their appropriate positions. For example, in this screenshot, the player is expected the spell the first syllable (ka) of the word “kala” (fish). In the case of an incorrect response, the misplaced letter tiles return to the bottom on the screen, the word is spoken again, and the player gets another try. The process continues until the word is correctly spelled.

Figure 3. The effect of GL Reading intervention and self-efficacy on posttest reading score. Group regression lines are estimated by PROCESS controlling for Word reading fluency pretest score. Solid line represents the control group, cut line the intervention group. Vertical line shows the border for the Johnson-Neyman significant region.

Figure 4. The effect of GL Reading intervention and self-efficacy on posttest spelling scores. Group regression lines are estimated by PROCESS controlling for spelling pretest score. Solid line represents the control group, cut line the intervention group. Vertical line shows the border for the Johnson-Neyman significant region.

Figure 5. Mean response time and response accuracy as possible mediators between self-efficacy and reading/spelling gain. In b_1, b_2 and c' the first number expresses the coefficient of the reading gain model and the second number the coefficient of the spelling gain model.