Author(s): Puolakanaho, Anne; Latvala, Juha-Matti

Title: Embedding Preschool Assessment Methods into Digital Learning Games to Predict Early Reading Skills

Year: 2017

Version:

Please cite the original version:

Puolakanaho, A., & Latvala, J.-M. (2017). Embedding Preschool Assessment Methods into Digital Learning Games to Predict Early Reading Skills. Human Technology, 13(2), 216-236. https://doi.org/10.17011/ht/urn.201711104212

HUMAN TECHNOLOGY

ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 13(2), November 2017, 216–236

EMBEDDING PRESCHOOL ASSESSMENT METHODS INTO DIGITAL LEARNING GAMES TO PREDICT EARLY READING SKILLS

Anne Puolakanaho
Department of Psychology
University of Jyväskylä
Finland

Juha-Matti Latvala
Niilo Mäki Institute
Finland

Abstract: *The aim of this pilot study was to explore the predictive accuracy of computer-based assessment tasks (embedded within the GraphoLearn digital learning game platform) in identifying slow and normal readers. The results were compared to those obtained from the traditional paper-and-pencil tasks currently used to assess school readiness in Finland. The data were derived from a cohort of preschool-age children (mean age 6.7 years,* N *= 57) from a town in central Finland. A year later, at the end of first grade, participants were categorized as either slow (*n *= 11) or normal readers (*n *= 46) based on their reading scores. Logistic regression analyses indicated that computer tasks were as efficient as traditional methods in predicting reading outcomes, and that a single computer-based task—the letter–sound knowledge task,—provided an easy method of accurately predicting reading achievement (sensitivity 95.7%; specificity 81.8%). The study has practical implications in classrooms.*

Keywords: *computer-based assessment, preschool, early reading skills, slow readers, prediction, letter knowledge.*

DOI: http://dx.doi.org/10.17011/ht/urn.201711104212

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

INTRODUCTION

First-grade teachers typically consider how to identify those children who need special help in learning the basic principles of alphabetic decoding and fluent reading. Furthermore, teachers must determine what kinds of screening methods are most efficient and easy to use in a school environment.

An abundance of freeware and commercial computer games are available nowadays for assessing reading-related skills (Carson, Gillon, & Boustead, 2011; Forster & Souvignier, 2011; Sainsbury & Benton, 2011) and improving those skills (Karemaker, Pitchford, & O'Malley, 2010; Price et al., 2009). An important benefit of digital learning games is that they seem to attract children, thus increasing the opportunity for engagement and motivation (Gros, 2007; Hall, Hughes, & Filbert, 2000; Mioduser, Tur-Kaspa, & Leitner, 2000). Digital learning games and assessment methods offer the possibility to present and repeat instructions and tasks in the same format for all players, and these games typically work with minimal guidance from adults. However, only a few computer games aiming to improve reading skills include a reliable assessment tool for evaluating skills in the initial phase and for predicting reading outcomes. The aim of this study was to explore the predictive accuracy of the computer-based screening measures implemented in the GraphoLearn digital learning environment in identifying normal and slow reading learning. If this study were able to demonstrate easy assessment and prediction of reading skills, such results could be utilized in planning classroom teaching and individual reading training.

The GraphoLearn Learning Environment

GraphoLearn[1] (referred henceforth with acronym GL) is an internationally implemented and studied (e.g., Brem et al., 2010; Hintikka, Aro, & Lyytinen, 2005; Kyle, Kujala, Richardson, Lyytinen, & Goswami, 2013; Ojanen et al., 2015; Ronimus & Richardson, 2014; Saine, Lerkkanen, Ahonen, Tolvanen, & Lyytinen, 2011) digital learning environment for supporting children in learning to read. Versions of GL have been piloted in about 20 languages, such as English, Swiss-German, Spanish, Chinese, and a number of Bantu languages in Africa. The aim of GL is to improve children's basic reading and writing skills, especially for those who have difficulty learning these skills. The various game versions have been designed for both orthographically transparent and opaque languages, because writing systems vary across languages and have a significant impact on the learning processes of reading (Aro, 2006; Richardson & Lyytinen, 2014; Ziegler et al., 2010). In orthographically transparent languages (e.g., Finnish, Italian, and many Bantu languages), the correspondence between the written letter and speech sound is consistent. For opaque languages (e.g., English, Portuguese, and French), the game works with larger units (like rhymes and words) that behave consistently (Kyle et al., 2013). For detailed descriptions of how the game works, see Richardson and Lyytinen (2014) and Lyytinen, Erskine, Kujala, Ojanen, and Richardson (2009).

The Finnish version of GL, Ekapeli,[2] is distributed via the Internet and is freely available to teachers, special educators, and parents. Like all versions of GL, Ekapeli focuses on learning the connections between spoken and written language by using a synthetic phonics approach that systematically introduces speech sounds, then syllables and words, and, later, connecting them to the written counterparts.

Learning to Read in a Transparent Orthography

Nowadays, it is widely recognized that orthographic transparency has a significant effect on reading. Studies have shown that children learn to read more quickly in a transparent orthography (Seymour, Aro, & Erskine, 2003). In addition, the consistency of the orthography influences the initial adoption of strategies for word recognition. Finnish is one of the most transparent alphabetic languages, with nearly 100% consistent letter–sound correspondence (Aro, 2006; Ziegler et al., 2010). However, as Aro (2006) explained, Finnish words change and their length increases due to their inflections, which increases the decoding burden. The process of transferring written words into spoken words demands rapidly matching a letter or combination of letters to their sounds and recognizing the patterns that make syllables and words (Aro, 2006).

Nearly all Finnish children enter preschool during the year they turn 6 years of age, one year before they start school. Many Finnish children can already read when they enter school (Holopainen, Ahonen, Tolvanen, & Lyytinen, 2000), while the rest normally learn basic decoding within a few months. At the end of first grade, nearly all children can read any words (and pronounceable pseudowords) and sentences in Finnish. From this point on, the most critical requirement is to learn to read words and sentences fluently, which is a phase of reading faced many years later by natives of languages with non-transparent orthography (Aro, 2006; Richardson & Lyytinen, 2014).

The connection between preschool skills and first-grade reading skills is especially visible in transparent languages, and this connection forms the basis for the later phases of fluent reading (Aro, 2006; Seymour et al., 2003; Ziegler et al., 2010). In addition, the connection is likely very sensitive to the timing of the assessment because the letter–sound knowledge skills and decoding skills of regular letter–sound correspondences can develop very rapidly during the first few months of given reading instruction in school, even among the slowest learners (Holopainen, Ahonen, & Lyytinen, 2001).

Scope of the Current Study

Recent results (e.g., Thompson et al., 2015), and earlier studies of orthographically opaque languages (e.g., Carroll & Snowling, 2004; Compton, Fuchs, Fuchs, & Bryant, 2006; de Jong & van der Leij, 2003; Pennington & Lefly, 2001), have shown that, during the kindergarten and preschool phases, two of the most powerful precursors of early reading skills are letter knowledge and phonological awareness. Similar results have been found in studies of Finland's orthographically highly transparent language (Holopainen et al., 2001; Puolakanaho et al., 2008; Silvén, Poskiparta, & Niemi, 2004). In the present study, tasks for assessing letter knowledge and phonological awareness skills were modified and embedded within the GL environment. The aim was to explore whether these measures for preschool-age children offered a reliable tool for predicting their reading skills at the end of first grade. These computerized tasks were compared to a set of standardized, traditional paper-and-pencil letter knowledge and phonological awareness tasks that are commonly used in Finland, along with other assessment instruments, to predict readiness for school entry (Elomäki, Huolila, Poskiparta, & Saranpää, 1999).

The current study was based on data from 57 preschool children (mean age 6.7 years) whose prereading skills were evaluated using both computer-based and traditional instruments. In the current study the children were categorized by the researchers into groups of slow and normal readers based on their reading skill a year later, at the end of first grade. The main research questions addressed in the present study were

1. What is the predictive accuracy of the traditional paper-and-pencil method and of the computer-based GL assessment method in identifying slow readers?
2. What is the most economical (i.e., the smallest combination of subtests) GL screening method required to predict reading ability after one year at school?

Because earlier studies have shown that computer-based screening methods can be used to predict reading outcomes (Carson et al., 2011; Forster & Souvignier, 2011; Sainsbury & Benton, 2011), we proposed the following hypothesis:

> H1: Computer-based screening methods will be as accurate as paper-and-pencil screening methods in predicting reading outcomes.

For screening purposes, it is important to determine the most cost- and time-effective and easy-to-use procedure to predict reading outcomes. Previous studies (Catts, Fey, Zhang, & Tomblin, 2001; Elbro, Bostrom, & Petersen, 1998; Pennington & Lefly, 2001; Puolakanaho et al., 2007; Thompson et al., 2015) have shown that only a few tasks typically are needed to predict reading outcomes. Based on these findings, we also proposed the following hypothesis:

> H2: Only a few screening measures of the GL assessment are needed to predict reading outcomes reliably and sensitively.

In the current study, we were also interested in how to make use of the results in everyday life and school day practice. Therefore we also want to show how to determine the pupil´s individual risk of reading difficulties. This, in turn, allows a teacher to make plans for classroom teaching and individual reading training for pupils who need special help.

METHOD

Participants

The present data were drawn from the LukiMat project[2] (Latvala, Koponen, Salmi, & Heikkilä, 2012). The aim of the LukiMat project is to offer professional knowledge and information about various aspects of reading and mathematics to Finnish educators and parents, as well as evaluation and training tools for reading and mathematics skills.

Finnish legislation (Government of Finland, 1998/628) defines that every child must participate in a pre-education program one year before starting school. More than 98% (Kinos & Palonen, 2013) of Finnish children attend formal preschool. Toward the end of the preschool autumn semester (i.e., November and December, 2012), the entire cohort (*N* = 83) of preschool-age children living in a small town in the province of Central Finland was assessed via a standardized paper-and-pencil instrument (Elomäki et al., 1999) to evaluate the children's readiness for school. All seven subtasks from the test were used for screening purposes. The computer-based GL screening tests were administered four months later (in March). The reading outcomes were measured more than a year later, in May, at the end of their first year of schooling.

The data presented in this paper were drawn from 57 participants whose parents gave permission for them to take part in the study and whose full data set was available. The teachers in the participants' elementary school were not encouraged to use the GL learning game with the pupils, although usage was not forbidden. The typical situation in Finnish schools is that some teachers use computer-based means (e.g., GL) in addition to traditional training methods that support children with reading delays (see, e.g., Ise et al., 2011). All participating children were ethnically Finnish, spoke Finnish as their native language, and had no reported mental, physical, or sensory deficiencies. However, three of the participants were children whose school entry had been postponed for one year, five had a history of speech and language delay (specific language impairment), and two were reported to have attention problems. In Finland and internationally, studies show that 5–10% of children have some learning difficulty (Holopainen, 2002; Lyytinen, Ahonen, Korhonen, Korkman, & Riita, 2002), which indicates that the research sample closely resembled the typical Finnish school population. Of course, the incidence of learning difficulties depends on the definition and measures used to assess it, which means an exact percentage of the population is difficult to determine.

Assessment Methods

The predictive measures (assessed in preschool) consisted of the traditional assessment measures presented by Elomäki et al. (1999). This group assessment of school readiness (developed in Turku, Finland) is referred to here as the "Turku battery." For the digital assessment component of this study (conducted in preschool), we used the Finnish-language version of GL (i.e., Ekapeli). Our main aim was to compare the sensitivity of the reading related tasks (phonological and letter knowledge) using two different methods (Turku Battery and GL), but we also wanted to find out if the other tasks (mathematical, memory and visuo-motoric) in the Turku battery would increase the sensitivity in prediction of reading outcomes. Therefore the sensitivity of the whole Turku battery was analyzed.

We used three outcome (reading) measures to assess reading skills at the end of first grade, and we utilized these measures to categorize the children into groups of slow readers (SR) and normal readers (NR). We also assessed background measures (i.e., vocabulary, performance level, and familial history of reading problems) to investigate whether the categorized groups differed from each other. In the following section, we describe the measures in detail.

Predictive Measures Before School

Traditional Paper-and-Pencil Assessment Tasks

The paper-and-pencil tasks consisted of seven subtasks (Elomäki et al., 1999) and they are prefixed in this study with the initials Tu. Cronbach's alpha for the entire test battery (which we later refer to as *TuPre*) was .80. The subtests of the TuPre are presented in following.

Copying from a Model (TuCOP). The children drew a cross, a square, a circle, and a diamond three times according to a given model. The children's performance was estimated based on the standard guidelines outlined in the assessment manual. One point was awarded for each correct shape (maximum score = 12 points).

Memory (TuMEM). Three separate sets of pictures were presented to the children. The teacher named the pictures in each set out loud consecutively. Here is an example of a one set, in which teacher read aloud:[3] "Look at the set where there is a *kuppi* [mug], *sieni* [mushroom], *lasi* [glass], *omena* [apple], *kärpänen* [fly], and *kenkä* [shoe])." The teacher then asked the students to draw a line across some of the pictures: "Draw a line through a fly, a glass, an apple, and a shoe. Start now!" The first set included six objects and targeted four to cross out. The second set of pictures presented seven objects and targeted five to cross out and the third and final set presented nine objects, of which five had to be crossed out. One point was given for each correct object crossed out, a point subtracted if an unnamed object had a line through it, and zero points given if a named object was not marked (maximum score = 14).

Mathematical Readiness (TuMAT). This task evaluated mathematical concepts (e.g., as many as, one more, one less, first, fourth, seventh). The tasks were fulfilled using a picture on which children had to draw their answers (e.g., a box of balls was presented with some numbers below, and the child was asked to draw a line across the number that indicated how many balls there were in the box). One point was awarded for each correct answer (maximum score = 18).

Copying Through Dots (TuDRA). The children were asked to draw, with the help of dots, a similar design as in the model picture. One point was given for each correct pair of line-connected dots (maximum score = 10). The task included one item for practice.

Initial Phoneme Matching (TuPHM). Each item included a target picture and comparison pictures that were presented side by side. A line of pictures was first identified aloud by the teacher, and the children were given the test instructions beforehand, in the form of, "Words begin with a sound that can be heard. At the beginning of the Finnish word *risu* [twigs], the [r] can be heard. There are also four other pictures on your paper (teacher points with finger). Here is *auto* [car], *linna* [castle], *ruusu* [rose], and *tyyny* [pillow]. The name of one picture among the four starts with the same sound as *risu*. Is it *auto, linna, ruusu, or tyyny*?" (In this case, the correct answer is *ruusu*.) Participants were asked to draw a line through the picture they selected (maximum score = 10). Three practice items were presented before the test items.

Identifying Rhyming Words (TuRHY). This procedure was similar to the initial phoneme-matching task. However, the children were required to show which two (rhyming) words sounded alike. The teacher would say, "Look at the following pictures in this row; the first picture is *nappi* [button] and next to that are *pappi* [priest], *sika* [pig], and *kaappi* [closet]. Which one sounds like *nappi*? Is it *pappi, sika, or kaappi*?" (In this case, the correct answer is *pappi*.[4]; maximum score = 10). Two practice items were presented before the test items.

Writing Letters (TuLEW). The children were asked to write on a paper, next to a series of pictures of objects, the requested 19 letters (e.g., "Write *R* beside the picture of the cat"). The letters were presented orally one at a time by the instructor. Correct forms of uppercase and lowercase letters were accepted. Mirror images, rotated letters, or problems in fine-motor skills were not considered errors (maximum score = 19).

TuPHM, TuRHY, and TuLEW measure prereading skills. The combination of these three measures is called TuRead in the following analysis and Cronbach's alpha for the TuRead portion was .73.

Computer-based Assessment Tasks

The computer-based assessment tasks (*GLRead*), embedded in the GL, consisted of two subtasks and they are prefixed here with the initials GL. The internal consistency of GLRead, measured using Cronbach's alpha, was .64. The subtests of GLRead are presented here.

Phonological Awareness (GLPhon). GLPhon tasks were age-specific modifications of the tasks used earlier in the Jyväskylä Longitudinal Study of Dyslexia project in a computer-animated program called Heps-Kups Land (for a detailed description, see Puolakanaho, Poikkeus, Ahonen, Tolvanen, & Lyytinen, 2003). However, in this study, all of the task instructions and individual test items were presented using the GL program, and the child gave his/her answer by clicking it via the mouse. After giving the answer, the child could see the pictures from that series vanish from the screen, and a new series of pictures would appear. No other feedback on the answer was given.

The GLPhon score (maximum score = 20) was based on the following two subtasks:

a) The first was syllable-level segment identification (10 items), where three pictures of objects were presented on the screen, immediately followed by the name written form of each object (e.g., *kissa* [cat], *koira* [dog], *kukko* [rooster]). The child was asked to identify the sub-word-level units (syllables) within the target (e.g., "In which picture can you hear the sound /*koi*/?"). If the child was successful in five or more of the items, the program continued to the single- sound-level targets.

b) Sound identification (10 items). The procedure was the same as for the first task, but the targets were single sounds (e.g., Child hears: *auto* [car], *juna* [train], *vene* [boat] and sees the corresponding pictures; "In which picture you can hear the sound /o/?" The child chooses the picture to which the target sound refers. (In the example, the correct answer is the picture of the car, "*auto.*")

Letter–Sound Knowledge (GLLeSo). Here, the child was asked to match the letter sounds to the letter-appropriate characters (of the 23 presented), written in capital letters and displayed simultaneously on the screen. The sounds of the letters (via recorded natural speech produced by the computer program) were played aloud one at a time, and the child gave an answer by clicking the letter via the mouse. The child could replay the sound by right-clicking the mouse (or by asking the instructor). The placement of the letters on the screen was presented in a fixed order throughout the subtask, and feedback on the answer was not given. (maximum score = 23).

Background Measures

We wanted also to know if the samples in the current study were equal and if the pupils´ abilities were among the normal age range. Therefore we evaluated samples using three different background measures; a receptive vocabulary task, a performance-level task, and an evaluation of familial risk of reading related problems.

Vocabulary. The Peabody Picture Vocabulary Test–Revised Form L (PPVT–R; Dunn & Dunn, 1981) was used to obtain a measure of receptive vocabulary. In this test, the child heard a word, and he or she was asked to point out via the mouse which, of the four pictures presented, the word referred to. The pictures and vocabulary items were presented to the children via the computer. The raw sum score of the correct items (for a description of the Finnish shortened version; see Lyytinen & Lyytinen, 2004) was used as a score (maximum score = 75).

Performance Level. Raven's Colored Matrices (Raven & Court, 1998) were used to assess nonverbal IQ, that is, the performance level at preschool age. The matrices were presented to the children via computer. Because the task has not yet been standardized in Finland, the raw sum score of correct items was employed (maximum score = 36).

Familial Risk. Parents were asked to fill out a paper questionnaire designed to evaluate a history of the reading and spelling problems in the children's close relatives (siblings, parents, aunts, uncles, and grandparents). The criterion for the familial risk of readin*g* problems was fulfilled if one of the child's parents or siblings was reported to have a history of persistent reading or spelling problems. For the sake of interest, the reported reading problems of other relatives are also presented in this paper.

According to the nonparametric Shapiro-Wilk test (see Field, 2013), the sample in the Vocabulary and Performance Level tasks was normally distributed. We used nonparametric tests due the small sample size.

Outcome Reading Measures at the End of First Grade

The reading outcomes were evaluated using three different reading related measures described below. These measures were used to classify children into normal and slow readers in the current study. There were no missing data concerning the outcome measures.

Lukilasse. A timed word-list reading test from the Lukilasse test (Häyrinen, Serenius-Sirve, & Korkman, 1999) was used, in which the children read aloud a list of words that gradually became longer and more difficult. The number of correctly read words within the 2-minute time limit was transformed into a standard score according to the guidelines in the test manual (maximum score = 90).

Reading Text. The child's task was to read a short story (*Jännittävät matkat*; 24 words/901 characters) as quickly and accurately as he or she could. The story was presented on paper using lowercase letters. (For clarification, the beginning of the sentences also started with lowercase letters.) The score was calculated by dividing the number of words read by the time spent reading and converted to a final score of correctly read words per minute. This task was used previously in the Jyväskylä Longitudinal Study of Dyslexia (see Puolakanaho et al., 2007, 2008).

Luksu. The Luksu is a Finnish-modified version of the Woodcock-Johnson reading fluency task (Woodcock, McGrew, & Mather, 2001). The child's task was to read sentences presented in the GL environment and, after each sentence, to indicate if the given statement was true or false by clicking with the mouse. The final scores were calculated by determining how many appropriately true and false statements were correctly identified within 3 minutes (maximum score = 70).

Cronbach's alpha for the three reading measures was .92. The nonparametric Shapiro-Wilk test (see Field, 2013) showed that there was a normally distributed sample in Lukilasse and Luksu.

Categorizing Children into Slow Reader (SR) and Normal Reader (NR) Groups

The three measures of reading (described in the previous subsection) were conducted at the end of the first grade (mean age = 7.9 years, *SD* = 0.36) and used in classifying the participating children as those with and without reading problems. A similar procedure was used earlier in the Jyväskylä Longitudinal Study of Dyslexia (Puolakanaho et al., 2007). In this study, the following procedure led to the classification:

1. A cutoff point was calculated using the score at one standard deviation below the whole group's mean performance for each of the three first-grade outcome measures (i.e., Lukilasse, Reading Text, and Luksu).
2. A child was considered to have deficient skills in each respective task if his or her score fell at or below the cutoff point of the task.
3. To classify a child as a slow reader, the child's skills had to fall at or below the cutoff point for at least two of the three measures.

Using these criteria, 11 children (19.2%) in the study group were classified into the slow reader group. In the following sections, we use the acronym SR for children classified with slow reading skills and the acronym NR (46 children) for those classified with normal reading skills. The groups were compared using nonparametric methods because the sample size was relatively small and some measures were skewed.

Participants' Background Information

The background information relating to the sample is presented in Table 1. The study included 29 girls and 28 boys; however, there were more boys in the SR group than the NR group, proportionately. Gender differences in the predictive and reading-related measures were explored using the Mann-Whitney U-test (see Metsämuuronen, 2017). Girls outperformed boys slightly on the TuRead test battery's TuPHM task (Girls: M = 7.5, Boys: M = 5.9, Mann-Whitney U = 278.0, p = .038), the TuLEW task (Girls: M = 14.7, Boys: M = 11.7, Mann-Whitney U = 283.0, p = .047) and the TuMAT task (Girls: M = 16.2, Boys: M = 15.2, Mann-Whitney U = 282.0, p = .038). The results for both girls and boys are combined in the subsequent analyses.

Table 1. Background of Children in the Sample (*N* = 57).

	NR *n* = 46 (26 girls, 20 boys)		SR *n* = 11 (3 girls, 8 boys)	
	No	**Yes**	**No**	**Yes**
Familial Risk (immediate relatives)*	42	3	8	3
Reported RD Issues in close relatives*	30	15	7	4
* missing information in one case				
	NR		**SR**	
	M	***SD***	***M***	***SD***
Vocabulary* (*n* = 27 / *n* = 10)	18.9	3.0	17.8	3.1
Performance Level* (*n* = 25 / *n* = 11)	23.0	4.6	20.1	5.1

* missing information in some cases.
Note. NR = normal readers, SR = slow readers, RD = reading disability
immediate relatives = parents & siblings; close relatives = immediate relatives plus aunts, uncles, & grandparents

In this sample, 33% (15 of 46) of the parents whose children were in the NR group, and nearly 36% (4 of 11) of the parents who had children with SR, reported having close relatives with reading or spelling problems. However, the percentage of immediate relatives (at least one among parents or siblings) with reported reading problems was about 7% (3 of 46) in the NR group and 27% (3 of 11) in the SR group. It seems that children whose immediate relatives have reading or spelling problems are more prone to have reading difficulties than children from background with no familial risk. The finding is in line with previous studies of reading problems (Elbro et al., 1998; Pennington & Lefly, 2001; Puolakanaho et al., 2007).

No statistically significant difference was found in vocabulary and performance levels between the SR and NR groups. This result indicates that the groups were equal and further suggests the children's backgrounds did not affect the following results. However, it is notable, that we did not have data of all participants and the interpretation should be taken with caution.

RESULTS

The Connection Between Computer-based and Traditional Measures

In general, the nonparametric correlations (Spearman's rho; see Field, 2013) indicated that the children's various skills were interconnected (Table 2). The traditional and digitally measured letter–sound knowledge tasks' correlation indices showed that they capture a substantial share of the common variance. The various phonological awareness tasks (i.e., TuPHM and TuRHY from the TuRead task battery and GLPhon from the GLRead task battery) have significant connections.

Outcome reading measures had high correlations with each other and with predictive measures in general. However, the Luksu fluency task only correlated with a few (the TuPHM, TuLEW and GLLeSo predictive measures, and the Lukilasse outcome reading measure).

Table 2. Correlations (Spearman's rho) Between the Measures in the Whole Sample.

	TuCOP	TuMEM	TuPHM	TuMAT	TuDRA	TuRHY	TuLEW	GLLeSo	GLPhon	Lukilas	Read	Luksu
TuCOP	1											
TuMEM	.310*	1										
TuPHM	.270*	.376**	1									
TuMAT	.423**	.392**	.501**	1								
TuDRA	.454**	.392**	.292*	.652**	1							
TuRHY	.368*	.350**	.268*	.364**	.271*	1						
TuLEW	.396**	.440**	.665**	.629**	.364**	.387**	1					
GLLeSo	.505**	.443**	.564**	.618**	.414**	.384**	.739**	1				
GLPhon	.447**	.370**	.551**	.377**	.271*	.374**	.488**	.578**	1			
Lukilas	.370**	.442**	.529**	.540**	.417**	.427**	.592**	.615**	.375**	1		
Read	.294*	.341*	.530**	.413**	.334*	.362**	.545**	.596**	.373**	.948**	1	
Luksu	.061	.166	.307*	.239	.223	.176	.355**	.397**	.125	.649**	.787**	1

Note. TuCOP = Copying from Model (Turku Battery), TuMEM = Memory (Turku Battery), TuPHM = Initial Phoneme Matching (Turku Battery), TuMAT = Mathematical Readiness (Turku Battery), TuDRA = Copying through Dots (Turku Battery), TuRHY = Identifying Rhyming Words (Turku Battery), TuLEW = Writing Letters (Turku Battery), GLLeSo = Letter–Sound Knowledge (GraphoLearn), GLPhon = Phonological awareness (GraphoLearn), Lukilas = Lukilasse reading measure, Read = Reading Text "Jännittävät matkat", Luksu = Luksu fluency task (computer version),

*** = $p < .001$, ** = $p < .01$, * = $p < .05$

Differences Between SR and NR groups

The descriptive statistics of the predictive and outcome measures in the SR and NR groups are presented in Table 3. No ceiling or floor effects were discovered, although the measures were slightly skewed, indicating that the tasks were well mastered, especially in the NR group. However, because the analysis method used (logistic regression) is not sensitive to the deviation from normality, the measures were accepted in the analyses.

Statistically significant differences were found in the nonparametric Mann-Whitney U-test (between the SR and NR groups in the mean scores of all predictive and outcome measures. The nonparametric effect size values (*r*; see Fritz, Morris, & Richler, 2012) indicate that most of the measures have discriminative power. The most powerful predictors are the tasks that measure letter–sound knowledge.

Predictive Accuracy of the Battery of Traditional Tasks

All of the measures from the school readiness battery, that is, the TuPre (i.e., TuCOP, TuMAT, TuDRA, TuPHM, TuRHY, and TuLEW), were included in the logistic regression analyses. Because Performance Level and Vocabulary, which were used as background measures, failed to show differences between the groups, they were omitted from the analyses.

Table 3. Descriptive Statistics of the Preschool Measures and the First-grade Reading Outcomes in the NR and SR Groups.

		NR n = 46				SR n = 11					
Measures	Min-Max	*M*	*SD*	*Skewness*	*Kurtosis*	*M*	*SD*	*Skewness*	*Kurtosis*	*Mann-Whitney U*	*Effect size (r)*
TuCogn											
TuCOP	4-12	11.1	1	-0.I	0.1	8.9	2.3	-0.8	0.3	96.0**	0.44
TuMEM	6-14	11.3	1.5	-0.2	-0.1	9.0	2.1	0.1	-0.3	90.5**	0.44
TuMAT	7-18	16.5	2.5	-1.9	3.2	12.6	3.5	-0.4	-0.7	79.5***	0.49
TuDRA	0-8	5.8	2.7	-1,0	-0.3	3.3	3.0	0.4	0.7	131.0**	0.34
TuRead											
TuPHM	1-10	7.2	2.7	-0.5	-1.0	4.4	1.6	0	-0.8	104.0**	0.41
TuRHY	3-10	9.2	1.2	-1.7	2.3	8.1	2.0	-1.7	3.7	152.0*	0.29
TuLEW	1-19	15.0	4.7	-0.9	-0.7	5.9	2.3	-0.4	1.4	33.5***	0.59
GLRead											
GLPhon	8-20	16.9	3.0	-1.0	0.4	13.6	3.1	-0.1	-0.6	110.0**	0.39
GLLeSo	0-23	18.2	4.1	-0.9	0.0	7.2	3.8	-0.3	-0.2	16.5***	0.64
Reading outcomes											
Lukilasse	0-90	58.0	17.7	0.4	-1.0	20.1	13.7	-0.8	-1.0	7.5***	0.66
Reading Text	9-103	46.7	21.8	0.6	-0.5	11.5	1.8	-0.2	-1.7	0.0***	0.59
Luksu	6-61	33.5	11.7	0.5	-0.7	20.9	14.0	0.2	-2.3	149.0*	0.28

Note. TuCOP = Copying from Model (Turku Battery), TuMEM = Memory (Turku Battery), TuPHM = Initial Phoneme Matching (Turku Battery), TuMAT = Mathematical Readiness (Turku Battery), TuDRA = Copying through Dots (Turku Battery), TuRHY = Identifying Rhyming Words (Turku Battery), TuLEW = Writing Letters (Turku Battery), GLLeSo = Letter–Sound Knowledge (GraphoLearn), GLPhon = Phonological awareness (GraphoLearn),

*** = $p < .001$, ** = $p < .01$, * = $p < .05$, $r = \frac{z}{\sqrt{N}}$ (Fritz, Morris, & Richler, 2012)

We first explored how well the TuPre could predict reading outcomes (i.e., predict that a child would be in either the NR or SR group). The aim of this phase was to answer the first part of the first research question. Table 4 presents the results of the logistic regression analyses.

The regression models for the Turku battery were carried out using the Forced Choice procedure (i.e., enter method; see Field, 2013), and the analyses indicated that all of the predictors contributed to the outcome. The coefficient of determination of the whole TuPre battery was .72 (adjusted R^2 value), indicating the model's good predictive ability. The coefficient of reading-related measurements (TuRead) on the TuPre battery was $R^2 = .64$.

Table 4. The Logistic Regression Analyses of the Turku Battery and GraphoLearn Battery.

Model	Variables in Equation	R^2	Sensitivity %	False pos.	Specificity %	False neg.	Overall %
TuPre measures							
TuPre (enter) cut-off .50	TuCOP TuMEM TuPHM TuMAT TuDRA TuRHY TuLEW	.72	95.7	2/46	81.8	2/11	93.0
TuRead (enter) cut-off .50	TuPHM TuRHY TuLEW	.64	91.3	4/46	72.7	3/11	87.7
GraphoLearn							
GLRead (enter) cut-off .50	GLLeSo GLPhon	.75	95.7	2/46	81.8	2/11	93.0
Economical Model							
GLRead (fcond) cut-off .50	GLLeSo	.75	97.8	1/46	72.7	3/11	93.0
GLRead (fcond) cut-off .35	GLLeSo	.75	95.7	2/46	81.8	2/11	93.0

Note. TuCOP = Copying from Model (Turku Battery), TuMEM = Memory (Turku Battery), TuPHM = Initial Phoneme Matching (Turku Battery), TuMAT = Mathematical Readiness (Turku Battery), TuDRA = Copying through Dots (Turku Battery), TuRHY = Identifying Rhyming Words (Turku Battery), TuLEW = Writing Letters (Turku Battery), GLLeSo = Letter–Sound Knowledge (GraphoLearn), GLPhon = Phonological awareness (GraphoLearn). Two different regression analysis procedures were used i.e., enter = Forced Choice procedure and fcond = Forward Conditional procedure.

Predictive Accuracy of Computer-based Methods

In the second phase of the analysis, to answer the second part of the first research question, a logistic regression procedure was used to examine how the GLRead test battery (i.e., the combined GLPhon and GLLeSo tests) predicted reading outcomes. The results showed that the coefficient of determination (the adjusted R^2 value) for the GLRead test battery was .75, for TuPre was .72, and for TuRead was .64 (see Table 4). It is notable that the predictive accuracy of both the traditional tasks and computer-based tasks were similar.

The Most Economical Predictive GL Reading Readiness Model

The second research question addressed identifying the most economical—that is, the simplest possible—computer-based procedure for predicting reading ability after one year at school. Therefore, the logistic regression analyses were conducted using the Forward Conditional procedure (i.e., fcond; see Field, 2013) for the GLRead (see Table 4). The analyses indicated that just a single measure of GL—letter–sound knowledge (GLLeSo)—was needed (adjusted $R^2 = .75$) to predict whether children would be in the SR or NR group.
In logistic regression, it is possible to include more children in the SR group by setting the cutoff criterion lower than .50 (see Elbro et al., 1998). This procedure will recognize more true SR-cases, but at the same time it also increases the number of NR-children falsely recognized as SR-children. From practical point of view, it is more important to try to identify all SR-children, than leave them unrecognized and therefore without support. Decreasing the cutoff criterion of SR from .50 to .35 produced a 93% classification accuracy for the overall model (with a sensitivity score of 95.7% and a specificity score of 81.8%). In our sample, this result meant that 9 of the 11 SR cases were recognized, and two SR cases (i.e., 18% of the total SR cases) could not be predicted using the GLLeSo task. In addition, 2 of the 46 NR cases (i.e., 4.3% of the total SR cases) were misleadingly predicted as having problems with reading at the end of first grade.

Implementation into Practice: Determining the Individual Risk of SR

In the current study, we also were interested in implementing the results into practice to determine the individual risk of reading difficulties. We see that the classification rate of the above-mentioned prediction model (i.e., GLLeSo) is high when compared with other non-Finnish prediction models, particularly those from abroad (i.e., Catts et al., 2001; Elbro et al., 1998; Pennington & Lefly, 2001) but even some from Finland (i.e., Holopainen et al., 2001; Lerkkanen, Rasku-Puttonen, Aunola, & Nurmi, 2004). This implies that the model may be clinically useful, and the results could be utilized in school practice, in that the logistic regression model (and the mathematical equations behind it) can be used to predict the individual risk of SR at preschool age (see Elbro et al., 1998). The individual scores for those at risk of SR (based on the most economical model presented above, i.e., GLLeSo) can be calculated from the assessed letter–sound knowledge scores using the following mathematical equation:

$$\text{Individual SR risk score} = \frac{1}{1 + \mathrm{e}^{-(6{,}059 - 0{,}607 \times [\text{GL Letter}-\text{Sound Knowledge score}])}}$$

The index of risk varies between 0 and 1, where values close to zero indicate a minimal risk of SR. The closer the value is to 1, the greater the risk of SR (see Elbro et al., 1998; Puolakanaho et al., 2007; Thompson et al., 2015). We applied the formerly presented results of the logistic regression model and the equation to illustrate the distribution of the individual risk scores, GLLeSo scores, and SR cases in the current study sample (see Table 5).

In Table 5, we present the distribution of SR cases in the present study sample. The table also gives the corresponding GLLeSo scores and individual SR risk ranges of different level-of-risk areas. Of the participants, 66% had a minimal, 18% had a moderate, and 16% had a high risk of SR. Thus, for example, if the assessed preschool child correctly identified 12 letters (based on

Table 5. Distribution of Risk of Being a Slow Reader in the Study Sample.

Level of risk	GLLeSo score at the start of the first grade	Individual SR risk score	Percentage (%) in the current sample	Number of children identified to this group	Number of true positive SR cases	Percentage of the children identified falsely (%): false negative SR/false positive SR
High Risk of SR	0-9	0.644-0.988	16	9	8	27.3 / 11
Moderate Risk of S	10-13	0.137-0.497	18	10	3	0 / 70
Minimal Risk of SR	14-23	0.08 -0.0	66	38	0	0 / 0

Note. SR=Slow Reader; GLLeSo= Letter–Sound Knowledge

their sounds) in GL, he or she would be considered to have a moderate risk of SR, as only 3 out of 10 study participants who scored between 10 and 13 letters correctly were classified accurately as having slow reading learning development one year later in school. We want to emphasize, that this implementation is suggestive and generalization of the results is cautioned due to the small sample size in this study.

DISCUSSION

The first aim of the present study was to compare the predictive accuracy of identifying children at risk of reading problems by a computer-based screening method (GL) versus traditional paper-and-pencil screening methods. The second aim was to determine the most economical model of a GL screening method to predict reading ability after one year of school. This study was conducted with a group of preschool-age children, whose reading performance was measured one year later when they were at the end of first grade. The two most promising preschool predictors of early reading outcomes in languages with a transparent orthography—letter knowledge and phonological awareness—were embedded in the GL learning environment. These GL assessment methods were compared with a standardized battery of paper-and-pencil tasks to predict reading outcomes in young Finnish children.

The current study confirmed the first hypotheses and showed that GL-based screening methods are as accurate as paper-and-pencil screening methods in predicting reading outcomes. The study indicated that the GL-based screening methods offer a reliable and easy-to-use procedure for predicting reading skills. The finding is in accord with other computer-based reading assessment studies (e.g., Carson et al., 2011; Forster & Souvignier, 2011; Sainsbury & Benton, 2011). In addition, we determined that, in the simplest form, only a single computer-based task was needed to predict reading skills in our study, namely letter knowledge assessed using a letter–sound matching activity (the GLLeSo task). Therefore, the study also confirmed the second hypotheses and showed that only a few screening measures are needed to predict reading outcomes reliably and sensitively.

These findings are in line with those of other studies that have indicated that, specifically in writing systems relatively consistent at the letter–sound level, knowledge of letter–sound correspondence is an important factor, or a preliterate skill, that will further promote reading acquisition skills (Aro, 2006; Georgiou, Torppa, Manolitsis, Lyytinen, & Parrila, 2012; Thompson et al., 2015). Puolakanaho et al. (2008) showed that, starting at 5 years of age, the development of phonological awareness and letter knowledge begin to intertwine with each other in a child's development. From this point onwards, letter knowledge starts to play a more important role than phonological awareness in reading achievement prediction in most orthographic languages (see also the cross-language investigation by Ziegler et al., 2010).

However, these results do not imply that skills such as phonological awareness, memory, or vocabulary are not important during the preschool years. On the contrary, the most accurate prediction model was obtained from the school readiness test battery—if the combination of measures included GL-based reading methods (GLRead and GLPhon) and cognitive measures from the school readiness tests (Turku battery). This combination of screening measures would allow for the evaluation of cognitive, phonological, and other preliterate skills, thus forming the foundation for the development of necessary interventions.

The results of multiple studies indicate that early language, phonological, and cognitive skills form the basis for literacy development. This is shown in both orthographically opaque languages (e.g., Carroll & Snowling, 2004; Compton et al., 2006; de Jong & van der Leij, 2003; Pennington & Lefly, 2001; Snowling, Gallagher, & Frith, 2003) and more transparent writing systems (e.g., Elbro et al., 1998), such as the Finnish language (e.g., Holopainen et al., 2001; Silvén et al., 2004; Puolakanaho et al., 2008). These early measures are also highly predictive of dyslexia, as shown by Torppa et al. (2007), based on the Jyväskylä Longitudinal Study of Dyslexia.

In the present study, preschool letter knowledge predicted reading outcomes in first grade. However, we do not assume that preschool letter knowledge is the only measure that predicts the development of reading skills, particularly reading fluency. Because the Finnish language has one of the most transparent orthographies, children usually acquire reading skills very rapidly whereas reading acquisition takes more time in less transparent orthographies (Aro, 2006). Due to this, the timing of the assessment is crucial when the results are interpreted. Therefore, these results might not be transferable, as such, to orthographically opaque languages.

We recognize several limitations in the study. The present investigation is a pilot study where the sample is considered as representative of the wider population. However, the sample size was small, and therefore the sizes of the groups (NR and SR) were correspondingly small. In addition, the data were drawn from a single town in Central Finland, which might cause some bias. Moreover, we could not control the possible use of additional training methods, such as the GL training game or other methods. Therefore, the present study should be considered a pilot study exploring the reading-level screening possibilities of a computer game-based assessment. Notably, there was a 4-month gap between the Turku battery tests and the GL assessments, and therefore the maturation effects cannot be fully excluded. However, the aim of the study was not to replace the Turku battery for evaluation purposes, but rather to determine if there was an effective computer-based alternative for predicting reading achievement.

The results must also be interpreted with caution because the study was conducted in the context of the Finnish language, which has transparent alphabetic orthography with symmetrically consistent letter–sound (grapheme–phoneme) relationships. In opaque orthographies, our approach might not work as well, or it might work similarly but at a different age and school phase, as Seymour et al. (2003) and Ziegler et al. (2010) have pointed out.

The challenge of future studies is to determine whether GL assessment methods can be used to predict longer-term reading achievements and to explore how the GL method can be used for languages with less consistent orthographic grapheme–phoneme correspondences. Future studies should use larger sample sizes and wider populations (e.g., several towns in Finland, participants with different developmental backgrounds, or other countries with different linguistic orthographies). In this case, it would be important to confirm our results regarding the computer-based assessment task by using a larger variety of GL linguistic environments. Additional studies could follow the reading learning paths from the basic reading acquisition to more advanced fluent reading skills, as well as to reading comprehension skills.

IMPLICATIONS FOR RESEARCH AND APPLICATION

The current study showed that the GraphoLearn (GL) computer-based screening methods, implemented at preschool age, are as accurate as traditional paper-and-pencil screening methods in predicting reading outcomes one year later. Given the transparent orthography of the Finnish-language, only one assessment task, that is, the letter–sound knowledge identified by the GL LeSo task, was required to identify reliably the children with a high risk of SR. The scores from this task will allow Finnish classroom teachers to identify children with a high risk of reading difficulties and therefore facilitate their planning and executing effective teaching activities during the first school year.

The question that arises after identifying individuals at risk of SR is how to support them in the learning process of reading. Many training studies conducted abroad (see the reviews by Bus & van Ijzendoorn, 1999 and Tornéus, Hedström, & Lundberg, 1991) and in Finland (Poskiparta, 2002) have indicated that training in phonological awareness skills with speech processing and letter knowledge has a positive effect on the development of reading skills, especially in the early phases of reading education (e.g., Lundberg, Frost, & Petersen, 1988; Schneider, Roth, & Ennemoser, 2000). In addition, efficient training in letter–sound associations would be useful. These skills are the focus of digital learning games such as GL (see Lyytinen et al., 2009; Saine et al., 2011).

Reading skills are essential for learning other topics in school. Therefore, identification of SR children at an early phase of their school entry and supported start will help SR children not to fall behind their peers in other school subjects. This, in turn, will minimize other school related problems including motivational and behavioral shortcomings that may have notable effects on the pupil´s forthcoming learning paths.

ENDNOTES

1. More information about the GraphoLearn learning game can be found at http://info.GraphoLearn.com. The GraphoLearn learning game was earlier called the GraphoGame (Ekapeli in Finnish), and these names can be found in the earlier published papers.
2. The LukiMat project aimed at developing knowledge and material for assisting children in learning to read and for mastering mathematics skills. The project was funded by the Finnish Ministry of Education and Culture. More information is available from http://www.lukimat.fi

3. For clarification, all the instructions to the children were provided in Finnish. But the teacher's instructions have been translated to English for this report for readers' understanding. The words associated with the tasks have been presented in Finnish (with translations) because the task is based on the Finnish spelling and pronunciation.
4. In the Finnish language, every letter is pronounced, and therefore a double consonant or vowel will affect the rhyming (and therefore *kaappi* is not the correct answer in the given example). Accordingly, the spelling of a work with the same letters, of which some are doubled, changes the meaning of the word. For instance, *tuli* means fire, whereas *tuuli* means wind.

REFERENCES

Aro, M. (2006). Learning to read: The effect of orthography. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 531–550). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Brem, S., Bach, S., Kucian, K., Guttorm, T. K., Martin, E., Lyytinen, H., Brandeis, D., & Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. In *Proceedings of the National Academy of Sciences of the United States of America* (PNAS), *107*(17), 7939–7944. doi: 10.1073/pnas.0904402107

Bus, A. G., & van Ijzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, *91*(3), 403–412. doi: 10.1037//0022-0663.91.3.403

Carroll, J. M., & Snowling, M. J. (2004). Language and phonological skills in children at high risk of reading difficulties. *Journal of Child Psychology and Psychiatry*, *45*(3), 631–640. doi: 10.1111/j.1469-7610.2004.00252.x

Carson, K., Gillon, G., & Boustead, T. (2011). Computer-administrated versus paper-based assessment of school-entry phonological awareness ability. *Asia Pacific Journal of Speech, Language and Hearing*, *14*(2), 85–101. doi:10.1179/136132811805334876

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools*, *32*(1), 38–50. doi: 10.1044/0161-1461(2001/004)

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*, *98*(2), 394–409. doi: 10.1037/0022-0663.98.2.394

de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, *95*, 22–40. doi: 10.1037/0022-0663.95.1.22

Dunn, L. M., & Dunn, D. M. (1981). *Peabody picture vocabulary test–revised*. Circle Pines, MN, USA: American Guidance Service.

Elbro, C., Borstrom, I., & Petersen, D. K. (1998). Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly*, *33*, 36–60. doi: 10.1598/RRQ.33.1.3

Elomäki, T., Huolila, R., Poskiparta, E., & Saranpää, P (1999). *Kouluvalmiuden arviointi ryhmässä* [A group-based assessment of school readiness]. Turku, Finland: Publications of the Social Service Centre for the City of Turku.

Field, A. (2013). *Discovering statistics using IBM SPSS statistic*s (4th ed.). London, UK: Sage Publications.

Forster, N., & Souvignier, E. (2011). Curriculum-based measurement: Developing a computer-based assessment instrument for monitoring student reading progress on multiple indicators. *Learning Disabilities: A Contemporary Journal*, *9*(2), 65–88.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation—Correction to Fritz et al. (2011). *Journal of Experimental Psychology: General*, *141*(1), 1–30.

Georgiou, G. K, Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. Reading and Writing, 25, 321–346. doi: 10.1007/s11145-010-9271-x

Government of Finland, (1998/628). Perusopetuslaki. Annettu 21.8.1998. Viimeisin muutos 1.1.2017. [Basic Education Act. Passed August 21, 1998; Last amended January 1, 2017]. Retrieved November 1, 2017 from http://www.finlex.fi/fi/laki/ajantasa/1998/19980628

Gros, B. (2007). Digital games in education: The design of games-based learning environments. *Journal of Research on Technology in Education*, *40*(1), 23–38. doi: 10.1080/15391523.2007.10782494

Hall, T. E., Hughes, C. A., & Filbert, M. (2000). Computer assisted instruction in reading for students with learning disabilities: A research synthesis. *Education and Treatment of Children*, *23*(2), 173–193.

Häyrinen, T., Serenius-Sirve, S., & Korkman, M. (1999). *Lukilasse. Lukemisen, kirjoittamisen ja laskemisen seulontatestistö peruskoulun ala-asteen luokille 1–6* [Lukilasse. Screening test for reading, writing, and counting in grades 1–6]. Helsinki, Finland: Psykologien kustannus Oy.

Hintikka, S., Aro, M., & Lyytinen, H. (2005). Computerized training of the correspondences between phonological and orthographic units. *Written Language & Literacy, 8*, 155–178. doi: 10.1075/wll.8.2.07hin

Holopainen, L. (2002). *Development in reading and reading related skills: A follow-up study from pre-school to the fourth grade* (Jyväskylä Studies in Education, Psychology and Social Research, Report No. 200). Jyväskylä, Finland: University of Jyväskylä. Retrieved from http://urn.fi/URN:ISBN:978-951-39-5285-3.

Holopainen, L., Ahonen, T., & Lyytinen, H. (2001). Predicting delay in reading achievement in a highly transparent language. *Journal of Learning Disabilities, 34*, 402–413. doi: 10.1177/002221940103400502

Holopainen, L., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2000). Two alternative ways to model the relation between reading accuracy and phonological awareness at preschool age. *Scientific Studies of Reading, 4,* 77–100. doi: 10.1207/S1532799XSSR0402_01

Ise, E., Blomert, L., Bertrand, D., Faísca, L., Puolakanaho, A., Saine, Nl., Suranyi, Z., Vassen, A., Faisca, L., Csepe, V., Lyytinen, H., Reis, A., Ziegler, J. C., & Schulte-Körne, G. (2011). Support systems for poor readers: Empirical data from six EU member states. *Journal of Learning Disabilities*, *44*(3), 228–245. doi: 10.1177/0022219410374235

Karemaker, A., Pitchford, N. J., & O'Malley, C. (2010). Enhanced recognition of written words and enjoyment of reading in struggling beginner readers through whole-word multimedia software. *Computers & Education*, *54*(1), 199–208. doi: 10.1016/j.compedu.2009.07.018

Kinos, J., & Palonen, T. (2013). *Selvitys esiopetuksen velvoittavuudesta* [A report clarifying the possibility of common mandatory preschool] Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä, 2013:5 [Memos and Reports by Ministry of Education and Culture, Report No. 2013:5]. Helsinki, Finland: Opetus- ja kulttuuriministeriö.

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

Latvala, J.-M., Koponen, T., Salmi, P., & Heikkilä, R. (2012). *LukiMat-palvelu tukemassa lukemisen ja matematiikan taitojen oppimista ja oppimisen arviointia* [LukiMat-service supporting learning and assessment of reading and mathematics skills]. *NMI Bulletin*, *22*(2), 36–53.

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*(1), 67–92, doi: 10.1080/01411920310001629974

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly*, *23*, 263–284. doi: 10.1598/RRQ.23.3.1

Lyytinen, H., Ahonen, T., Korhonen, T., Korkman, M., & Riita, T. (2002). *Oppimisvaikeudet. Neuropsykologinen näkökulma* [Learning difficulties: Neuropsychological view]. Helsinki, Finland: WSOY.

Lyytinen, H., Erskine, J., Kujala, J., Ojanen, E., & Richardson, U. (2009). In search of a science-based application: A learning tool for reading acquisition. *Scandinavian Journal of Psychology*, *50*(6), 668–675. doi: 10.1111/j.1467-9450.2009.00791.x

Lyytinen, P., & Lyytinen, H. (2004). Growth and predictive relations of vocabulary and inflectional morphology in children with and without familial risk for dyslexia. *Applied Psycholinguistics*, *25*(3), 397–411. doi:10.1017/S0142716404001183

Metsämuuronen, J. (2017). *Essentials of research methods in human sciences: Vol 3. Advanced analysis*. New Delhi, India: Sage Publications India.

Mioduser, D., Tur-Kaspa, H., & Leitner, I. (2000). The learning value of computer-based instruction of early reading skills. *Journal of Computer Assisted Learning*, *16*(1), 54–63. doi: 10.1046/j.1365-2729.2000.00115.x

Ojanen, E., Ronimus, M., Ahonen, T., Chansa-Kabali, T., February, P., Jere-Folotiya, J., Kauppinen, K,-P., Ketonen, R., Ngorosho, D., Pitkänen, M., Puhakka, S., Sampa, F., Walubita, G., Yalukanda, C., Pugh, K., Richardson, U., Serpell, R., & Lyytinen, H. (2015). GraphoGame: A catalyst for multi-level promotion of literacy in diverse contexts. *Frontiers in Psychology*, *6*, Article 671. doi: 10.3389/fpsyg.2015.00671

Pennington, B. F., & Lefly, D. L. (2001). Early reading development in children at family risk for dyslexia. *Child Development*, *72*, 816–833. doi: 10.1111/1467-8624.00317

Poskiparta, E (2002). *Remediation of reading difficulties in grades 1 and 2: Are cognitive deficits only part of the story?* [Studies of University of Turku. Series B. Humaniora] Turku, Finland: Universtiy of Turku.

Price, P., Tepperman, J., Iseli, M., Duong, T., Black, M., Wang, S., Boscardin, C. K., Heritage, M., Pearson, D. P., Narayanan, S., & Alwan, A. (2009). Assessment of emerging reading skills in young native speakers and language learners. *Speech Communication*, *51*, 968–984. doi: 10.1016/j.specom.2009.05.001

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P., Poikkeus, A., Tolvanen, A., Torppa, M., & Lyytinen, H. (2007). Very early phonological and language skills: Estimating individual risk of reading disability. *Journal of Child Psychology and Psychiatry*, *9*, 923–931. doi: 10.1111/j.1469-7610.2007.01763.x

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P. H. T., Poikkeus, A., Tolvanen, A., Torppa, A., & 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*(4), 353–370. doi: 10.1177/0022219407311747

Puolakanaho, A., Poikkeus, A-M., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2003). Assessment of three-and-a-half-year-old children’s emerging phonological awareness in a computer animation context. *Journal of Learning Disabilities*, *5*, 416–423. doi: 10.1177/00222194030360050301

Raven, J. C., & Court, J. H. (1998). *Raven’s progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologists Press.

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: An Interdisciplinary Journal on Humans in ICT Environment*s, *10*(1), 39–60. doi: 10.17011/ht/urn.201405281859

Ronimus, M., & Richardson, U. (2014). Digital game-based training of early reading skills: Overview of the GraphoGame method in a highly transparent orthography. *Estudios de Psicologia*, *35*(3), 648–661. doi: 10.1080/02109395.2014.974424

Saine, N. L., Lerkkanen, M. K., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2011). Computer-assisted remedial reading intervention for school beginners at risk for reading disability. *Child Development*, *82*(3), 1013–1028. doi: 10.1111/j.1467-8624.2011.01580.x

Sainsbury, M., & Benton, T. (2011). Designing a formative e-assessment: Latent class analysis of early reading skills. *British Journal of Educational Technology*, *42*(3), 500–514. doi: 10.1111/j.1467-8535.2009.01044.x

Schneider, W., Roth, E., & Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: A comparison of three kindergarten intervention programs. *Journal of Educational Psychology*, *92*(2), 284–295. doi: 10.1037/0022-0663.92.2.284

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174. doi: 10.1348/000712603321661859

Silvén, M., Poskiparta, E., & Niemi, P. (2004). The odds of becoming a precocious reader of Finnish. *Journal of Educational Psychology*, *96*, 152–164. doi: 10.1037/0022-0663.96.1.152

Snowling, M. J., Gallagher, A. and Frith, U. (2003), Family risk of dyslexia is continuous: Individual differences in the precursors of reading skill. *Child Development, 74*, 358–373.

Thompson, P. A., Hulme, C., Nash, H. M., Gooch, D., Hayiou-Thomas, E., and Snowling, M. J. (2015). Developmental dyslexia: Predicting individual risk. *Journal of Child Psychology and Psychiatry*, *56*, 976–987. doi: 10.1111/jcpp.12412

Tornéus, M., Hedström, G.-B., & Lundberg, I. (1991). *Löytöretki Kieleen—Leikkejä ja Harjoituksia* [Games and exercises to explore language]. Helsinki, Finland: VAPK-kustannus.

Torppa, M., Tolvanen, A., Poikkeus, A. M., Eklund, K., Lerkkanen, M. K., Leskinen, E., & Lyytinen, H. (2007). Reading development subtypes and their early characteristics. *Annals of Dyslexia*, *57*(1), 3–32. doi: 10.1007/S11881-007-0003-0

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson tests of achievement*. Itasca, IL, USA: Riverside Publishing.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Lyytinen, H., Vaessen, A., & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, *21*(4), 551–559. doi: 10.1177/0956797610363406

Authors' Note

We are grateful to all the families and children, as well as to their preschool and school teachers, for their cooperation in this study Also, we thank LukiMat- and GraphoGame –project teams, especially Professor Heikki Lyytinen, for their collaboration. We would like to offer special thanks to Anne Mönkkönen for the organization of the data collection in the small town in the Central Finland and Kenneth Eklund for his valuable comments on this paper. We also thank the Ministry of Education and Culture for funding the LukiMat-project, in which the data were collected.

All correspondence should be addressed to
Anne Puolakanaho
Department of Psychology
University of Jyväskylä
P.O. Box 35 (Agora)
40014 University of Jyväskylä
Finland
anne.puolakanaho@jyu.fi

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi