COMPUTERIZED READING FLUENCY ASSESSMENT

- Task validity and the strongest discriminators of fluency skills among second-graders

Maija Koikkalainen

Master's Thesis

Department of Psychology

University of Jyväskylä

September 2015

UNIVERSITY OF JYVÄSKYLÄ

Department of Psychology

KOIKKALAINEN, MAIJA: Computerized reading fluency assessment – task validity and the strongest discriminators of fluency skills among second-graders

Master's thesis, 24 pages

Instructor: Riikka Heikkilä

Psychology

September 2015

Abstract

This study examines how well computerized reading fluency assessment measures correspond to the more traditional pen-and-paper methods; and which computerized measures best discriminate between children poor reading fluency skills and the control group. The study was conducted in Finland among 203 second-graders, i.e. 7-8 year olds. The children were divided into two groups: the at-risk for reading difficulties group and control group, based on their standardized score on word-level fluency test. 36 of the participants placed in the 'at-risk' group, and 167 in the control group. On the course of the study, the participants conducted various tasks in a computerized, interactive GraphoGame learning environment, which recorded their results on the following measures: grapheme-phoneme correspondence, phonological awareness, phonological memory, phonemic length discrimination, short-term memory, rapid automatized naming, receptive vocabulary, word recognition, pseudo-word recognition, spelling, pseudoword spelling and sentence reading fluency (both large and normal letter spacing). They also took part in traditional pen-and-paper tests with the corresponding measures. Pearson correlations and logistic regression analysis were used to study the research questions. The results showed that majority of the computerized measures compared to the traditional measures well. Sentence reading with large letter spacing, word recognition accuracy and rapid automatized naming were found as the strongest discriminators of reading fluency skills between the two groups. The results shed positive light on using computerized methods in reading assessment.

Keywords: Reading fluency, computerized assessment, GraphoGame

JYVÄSKYLÄN YLIOPISTO
Psykologian laitos

KOIKKALAINEN, MAIJA: Lukemisen tietokoneistetut arviointimenetelmät – mittareiden validiteetti ja lukusujuvuuden vahvimmat erottelijat 2. luokalla
Pro gradu –tutkielma, 24 s.
Ohjaaja: Riikka Heikkilä
Psykologia
Syyskuu 2015

Abstrakti

Tämän tutkimuksen tarkoituksena oli selvittää, kuinka hyvin lukemisen tietokoneistetut arviointimenetelmät vastaavat perinteisiä kynä-paperi-menetelmiä, ja mitkä tietokoneistetuista mittareista parhaiten erottelevat normaalisti ja heikosti lukevia lapsia toisistaan. Tutkimukseen osallistui 203 oppilasta suomalaisten peruskoulujen 2. luokilta. Lapset jaettiin heikosti lukeviin ja kontrolliryhmään sanasujuvuustestin perusteella. 36 osallistujaa sijoittui heikosti lukevien ryhmään ja 167 osallistujaa kontrolliryhmään. Tutkimuksen aikana osallistujat suorittivat useita tehtäviä interaktiivisessa Ekapeli-oppimisympäristössä, joka tallensi osallistujien tulokset seuraavien muuttujien osalta: kirjain-äänne vastaavuus, fonologinen tietoisuus, fonologinen muisti, foneemisen keston erottelu, lyhytkestoinen muisti, nopea automaattinen nimeäminen, reseptiivinen sanasto, sanantunnistus, epäsanojen tunnistus, kirjoittaminen, epäsanojen kirjoittaminen sekä lausetason lukusujuvuus (normaaleilla ja suurilla kirjainväleillä). Osallistujat myös suorittivat samoja muuttujia vastaavat tehtävät perinteisillä kynä-paperi-testeillä. Analyysit tehtiin käyttämällä Pearsonin korrelaatiokerrointa ja logistista regressioanalyysia. Tulokset osoittivat, että suurin osa tietokoneistetuista tehtävistä vastasi perinteisiä menetelmiä hyvin. Lausetason lukemisen sujuvuus suurilla kirjainväleillä, sanantunnistuksen tarkkuus sekä nopea automaattinen nimeäminen yhdessä erottelivat parhaiten heikosti ja normaalisti lukevia lapsia toisistaan. Tulokset luovat uskoa tietokoneistettujen arviointimenetelmien käyttöönottoon lukemisen arvioinnissa.

Avainsanat: Lukusujuvuus, tietokoneistettu arviointi, Ekapeli

Table of Contents

Index of tables

INTRODUCTION

The importance of reading has been well recognized in research during recent years. Majority of the research has, however, been focusing on children speaking English – leading to exaggeration and overgeneralization of certain literacy precursors (Share, 2008). English is an opaque orthography in which the letter-sound correspondence is not consistent; the same letter or sequence of letters may be pronounced in various ways, depending on their position within a word (f.ex. the digraph 'gh' is pronounced /f/ in tough /'tʌf/ and /ɡ/ in ghost /ˈɡoʊst/). In such opaque orthographies, the development of accurate reading usually takes a long time – hence, reading difficulties are mainly reflected in reading *accuracy* and slow reading (Seymour, Aro & Erskine, 2003; Ziegler, Perry, Ma-Wyatt, Ladner & Schulte-Körne, 2003). The same does not hold true for transparent orthographies, which have a consistent letter-sound correspondence. In transparent orthographies, such as Finnish, decoding skills and reading accuracy can be achieved fairly soon in reading development, and reading difficulties are predominantly reflected in reading *fluency* and slow reading (Aro, Huemer, Heikkilä & Mönkkönen, 2011; Escribano, 2007; Holopainen, Ahonen, & Lyytinen, 2001). In order to avoid overgeneralizing conclusions on reading precursors and interventions, it is crucial to investigate reading fluency and its development in transparent orthographies more thoroughly. Fluency studies may also unveil essential information on promoting reading motivation (Leinonen et al., 2001) and comprehension (Bretznitz, 2006), thus benefiting those who struggle with reading.

Reading fluency is often defined as accurate reading with sufficient speed and prosody (Kuhn & Stahl, 2003). Occasionally, the aspect of reading comprehension is also included (Lai, Benjamin, Schwanenflugel & Kuhn, 2014; Norton & Wolf, 2012). While all these features are undoubtedly essential to fluent text reading, the present study refers to *fluency* merely as the rate and accuracy of reading – a definition proposed by Philips and Torgesen (2006). In the shallow orthography of Finnish language, children master reading accuracy fairly quickly (Aro & Wimmer, 2003; Aro, 2006). Therefore, reading rate becomes a more definite indicator of reading fluency. Fluent reading is a complicated act that presupposes several reading processes and subskills (Wolf & Katzir-Cohen, 2001); including, inter alia, the following: letter knowledge, naming skills, automatic decoding, processing speed, orthographic knowledge, and vocabulary (Hudson, Pullen, Lane & Torgesen, 2008; van den Boer & de Jong, 2015). The most crucial part in reading development is learning how to decode written words, i.e. to connect sounds with appropriate letters and letter combinations (Wimmer & Schurtz, 2010). Once a person becomes proficient in decoding, cognitive resources are released for higher level processing – such as reading comprehension and fluency (LaBerge & Samuels, 1974; Perfetti, 1985).

Transparent orthographies enable children to acquire decoding accuracy early in their reading development. In Finland, approximately 85% – 95% of children reach word-level reading accuracy by the end of their first grade at school (Aro & Wimmer, 2003; Aro, 2006; Seymour, Aro & Erskine, 2003; Torppa, Lyytinen, Erskine, Eklund & Lyytinen, 2010). Finnish is a very shallow orthography with only 24 phonemes (8 vowel and 16 consonant phonemes), and three additional "foreign phonemes" (/b/, /g/, /f/) that are used in more recent loan words. Phonemes are consistently marked with 24 graphemes (23 single letter graphemes and one digraph: ng). All these graphemes are bidirectionally regular, with merely one exception; the letter combination: ng, representing /ŋ/ sound. The near-perfect transparency between letters and phonemes in both directions enables readers to attend to very small units and follow a serial, letter-by-letter strategy in reading (Pagliuca & Monaghan, 2010; Ziegler & Goswami, 2005).

In addition to decoding acquisition, orthographic complexity affects the rate certain literacy precursors predict later reading skills (Aro, Huemer, Heikkilä & Mönkkönen, 2011; Landerl et al., 2013). Although a consensus prevails on several precursors and reading-related skills that predict and differentiate reading accuracy, the connection with reading fluency seems more complex (Aro, 2006; Puolakanaho et al., 2008; Seymour, 2005; Seymour, Aro, & Erskine, 2003). Further studies are needed to reach an agreement. Below, previous studies on common reading predictors and subskills are presented.

Certain cognitive background skills are able to predict children's later reading abilities throughout childhood. Torppa *et al.* (2010) found that letter knowledge at the age of five was the strongest predictor of second-grade reading accuracy and fluency among Finnish children. Similarly, other studies have found a significant connection with kindergarten letter-knowledge and second-grade reading fluency (Puolakanaho et al., 2008), as well as later spelling and writing skills (Torppa et al., 2013). The predictive impact of letter-knowledge diminishes, however, as children enter formal education. According to a Finnish study, poor readers reach the control group in letter-naming skills by the end of first grade (Ketonen, 2010).

Rapid automatized naming (RAN) is a known and well-researched precursor of reading fluency. In transparent orthographies, it seems to be one of the strongest predictors of literacy among children with reading difficulties (Holopainen, Ahonen, & Lyytinen, 2001; Landerl & Wimmer, 2008; Puolakanaho et al., 2007; Salmi & Torppa, 2011; Torppa et al., 2010). For example in Finland, a longitudinal study with over 1500 school children revealed a single RAN deficit a predictor of poorer reading fluency (Torppa et al., 2013). An Italian study found that pseudo-word reading accuracy and RAN alone were able to explain up to 69% of the variance in reading fluency among children with dyslexia, and 52% in the control group among 11-13 year-old Italian children (Zoccolotti et al., 2014). Furthermore, RAN seems to remain a stable predictor of reading skills across age groups (Kairaluoma, Torppa, Westerholm, Ahonen & Aro, 2013; Salmi, 2008; Scarborough, 1998; Torppa et al., 2010; Van den Bos, Zijlstra & Spelberg, 2002).

Early phonological skills also predict later literacy skills. Phonological awareness (PA), i.e. an individual's ability to understand the sound structure of a word, is among the most influential predictors of literacy, especially reading accuracy (Landerl et al., 2013; Melby-Lervåg, Lyster & Hulme, 2012; Puolakanaho et al., 2008). PA also predicts spelling (Babayigit & Stainthorp, 2011; Landerl & Wimmer, 2008), pseudo-word spelling accuracy, and fluency (Tobia & Marzocchi, 2014; Torppa et al., 2013). Previous research suggests, however, that the association between PA and reading is limited to early stages of reading development (e.g., Landerl & Wimmer, 2008; Leppänen et al., 2006; Torppa et al., 2013), and as children reach ceiling in their ability to decode words accurately, they start relying less on phonological skills (e.g. phonological awareness) and more on fluency-related skills. The orthographic depth of a language dictates when this shift occurs; children reading more transparent languages shift away from phonology earlier in schooling (Vaessen et al. 2010). In the context of Finnish language, children may progress from the first steps of decoding words to reasonably accurate albeit slow reading within weeks or months. According to Torppa *et al.* (2013), the shifting pattern of contributing roles of PA and RAN to reading fluency was evident on grade two, and they proposed it would be even clearer among older children. As for the current study, a great majority of second-graders have likely shifted from relying on phonological skills to fluency skills. Phonological skills might still, however, differentiate those with reading difficulties to a small degree, as dysfluent readers in shallow orthographies are known to exhibit lower phonological awareness scores than those of both age-matched and younger reading-matched peers (Landerl, Wimmer & Frith, 1997; Ziegler et al. 2003).

Phonemic length discrimination (PLM) is another feature of phonological skills connected to reading ability. Finnish studies have shown that children with reading difficulties tend to make more mistakes in phonological discrimination than typically reading children on grade 2 (Pennala et al., 2010). PLM is also critical to children's spelling accuracy (Pennala et al., 2013); it explains a unique variance of grade 2 spelling accuracy even after other factors such as verbal short-term memory, phonological memory and naming speed are taken into account (Pennala et al., 2010).

Phonological memory or verbal short-term memory, a system responsible for storing, retrieving and processing of stable associations between spoken and written material (Mann & Liberman, 1984; Kibby, 2009; Swanson, Kehler, & Jerman, 2010) has been found to contribute to reading acquisition in studies conducted in opaque orthographies (Gathercole, Service, Hitch, Adams & Martin, 1999; Nithart et al., 2011). A longitudinal study among Dutch children also showed that dyslexic children score significantly lower on non-word-repetition test from kindergarten to third-grade (Dandache, Wouters & Ghesquière, 2014).

Short term memory (STM) may also be connected to reading acquisition. Short-term memory is a concept used to describe the storage component of working memory (Wilson et al., in press), and it is often measured with simple span tests (Baddeley, 2012). Children with reading disabilities have been found to achieve significantly lower scores in short-term-memory tasks requiring the recall of phonemes compared to average-reading children (Swanson, Zheng & Jerman, 2008). However, the relationship with reading and other kinds of short-term memory tasks, such as the digit span, seems more complex. Studies conducted in opaque orthographies have found children with reading difficulties to score significantly less in STM tasks (Wang & Gathercole, 2013), whereas recent studies in transparent orthographies have found no significant effect (Dandache, Wouters & Ghesquière, 2014).

In addition to memory capacity, vocabulary plays a role in reading development. Receptive vocabulary has been found to predict decoding performance, word recognition, letter-naming and phonemic abilities in opaque orthographies (Ouellette, 2006; Wise et al., 2008). In transparent orthographies, research shows receptive vocabulary as moderately connected with difficulties in reading, spelling and writing (Torppa et al. 2013). Receptive vocabulary may also partly predict pseudoword recognition (Nation & Cocksey, 2009) and reading comprehension (Hemphill & Tivnan, 2008).

When children begin to read, reciprocal interactions between their development of reading and their development of related skills may bewilder our perception of causal pathways (Korkman, Barron-Linnankoski & Lahti-Nuuttila, 1999; Perfetti, 1985; Perfetti, Beck, Bell, & Hughes, 1987). It would be tempting to conclude that reading-related cognitive background factors, mentioned afore, would similarly improve prediction of reading across the later grades. Often, however, initial reading level becomes an autoregressor that increasingly explains variation in reading performance throughout formal instruction. Thus, the predictive role of some of the cognitive background factors diminishes (Parrila, Kirby & McQuarrie, 2004; Sprugevica & Hoein, 2003; Vaessen & Blomert, 2010).

Certain subskills of literacy may discriminate fluency abilities in the course of formal education. Spelling, for example, is in a close relationship with reading (Lyon, Shaywitz & Shaywitz, 2003; Ise & Schulte-Körne, 2010). Children with reading difficulties are often poor in spelling (Pennala et al., 2010; Puolakanaho et al., 2008; van Bergen, de Jong, Maassen & van der Leij, 2014); especially pseudoword spelling has been found weak among second-grade children with dyslexia (Eklund, Torppa, Aro, Leppänen & Lyytinen, 2015; Torppa et al. 2013).

Visual word recognition is also a crucial element for reading fluency (van den Boer & de Jong, 2015), and a highly stable indicator of reading skills in transparent orthographies (Landerl & Wimmer, 2008). Fluent reading is characterized by large lexicality effect and by a smaller number-of-letter effect for words than for pseudowords (Weekes, 1997; Martens & de Jong, 2006; 2008; Zoccolotti, De Luca, Judica & Spinelli, 2008). Dysfluent readers in transparent orthographies tend to process words in a slower, more fragmented manner; using a serial letter-to-letter decoding strategy (De Luca, Borrelli, Judica, Spinelli & Zoccolotti, 2002; Eklund et al., 2015; Hutzler & Wimmer, 2004). Pseudoword recognition is characterized by serial decoding strategy among all readers (Hautala, Aro, Eklund, Lerkkanen & Lyytinen, 2013), but dysfluent readers tend to exhibit delayed responses and slow identification of pseudowords (Hautala & Parviainen, 2014; Moll, Hutzler & Wimmer, 2005) – thus likely experiencing the process more laborious compared to fluent readers.

Rapid and automatic word identification is a pre-requisite for growth in reading fluency (Perfetti, 1992; van den Boer & de Jong, 2015) which, in turn, improves reading comprehension (O'Connor, White & Swanson, 2007). Text or passage reading tasks are common measures in fluency studies, but need to be grounded on the notion that text reading is depended on processes beyond single word reading – like prosody, and syntactic skills (Huemer, 2009). This study considers fluency merely as the rate and accuracy of word-level reading, and thus it is justified to leave passage-reading tasks out. Sentence reading fluency tests, similar to the Woodcock-Johnson timed reading fluency test that measures technical reading fluency with simple to moderate level sentences (Woodcock, Grew & Mather, 2001), are used in the study. Sentences with both normal and large letter spacing are used, as earlier studies have indicated that large letter spacing might improve reading of those with dyslexia (Zorzi et al., 2012) by diminishing crowding, the interference of flanking letters on the recognition of target letters (Whitney & Levi, 2011).

To date, computerized fluency assessment methods are rather few, despite the fact that computerized interventions have been discovered effective (Bhide, Power & Goswami, 2013; Saine, Lerkkanen, Ahonen, Tolvanen & Lyytinen, 2010). Similarly to computer-assisted interventions, computerized assessment is an appealing alternative when considering cost-effective evaluation methods. Valid computerized assessment methods would diminish the need of human resources – and might prove to be an engaging method, as digital environments are found motivating by students (Gee, 2003; Yang, 2012). Computerized assessment has aroused critique on its possible 'test-mode effect', referring to the influence of the condition under which one administers the test, e.g. pen-and-paper vs. computerized (Goldberg & Pedulla, 2002). Clariana and Wallace (2002), however, found no test-mode effect among undergraduates who were randomly assigned to either computer-based or identical paper-based test. Gender, competitiveness or computer-familiarity seemed to have no implications on the students' performance – content familiarity was the only variable with a significant effect. Moreover, Protopapas and Skaloumbakas (2007) used computer-based assessment in screening for reading disability and discovered that computer-based screening matched expert evaluations both in validity and reliability. These results encourage in seeking and developing accurate computerized assessment methods. Valid and reliable automatic measures could have a various implications benefiting researchers, pupils and education professionals in the field. Considering the rapid evolution of technology, advancing to computerized assessment methods is also not only effective, but rather expected, in the years to come.

The goal of this study is to evaluate validity of computerized reading assessment measures compared to traditional pen-and-paper measures, and to find the strongest discriminators of word-level reading fluency skills in the transparent orthography of Finnish language among second-graders (7–8 year-olds). A computerized learning environment is used to measure and assess pupils' performance in 13 literacy tasks.

METHOD

Participants and procedure

The present study is a part of an on-going research project called ReadAll (2013–2017), which seeks to create valid computerized assessment measures for various literacy precursors and reading-related skills in GraphoGame®[1](introduced below), as well as improve the GraphoGame technology. 203 Finnish second graders (105 girls/98 boys) from regular Finnish public schools participated in the study. Each child and their parent were asked a written consent for using the child's test scores in the study. The participants were tested in two phases: 1) in a whole-class situation where the children played a revised version of GraphoGame on their tablets, covering various measures (see below) 2) individually with a researcher, when each child took part in traditional pen-and-paper tests with the corresponding measures. On average, it took 120-150 minutes for each child to complete the required tests.

GraphoGame learning environment

GraphoGame (Ekapeli in Finnish) is a computer-based, game-like literacy learning program applied in over 20 countries (Lyytinen & Richardson, 2014). The program essentially trains two key alphabetic principles: phoneme awareness and letter knowledge. It provides the learner with both visual and audio feedback on their correct and incorrect choices in grapheme-phoneme

[1] GraphoGame is the registered trademark of the University of Jyväskylä and Niilo Mäki Foundation for a noncommercial computerized game aimed at learning to read.

correspondences, gradually progressing onto phonological recoding and decoding. Thus, the program conditions crucial skills required for fluent and accurate reading. (Lyytinen, Erskine, Kujala, Ojanen & Richardson, 2009; Saine et al., 2010). Furthermore, GraphoGame keeps players in their zone of proximal development by adapting individually to their skill levels (Lyytinen et al., 2009). GraphoGame has been found a highly beneficial tool in remedial reading interventions for children who are at risk for reading difficulties (Jere-Folotiya et al., 2014; Lovio, Halttunen, Lyytinen, Näätänen & Kujala, 2012); even more efficient than traditional-style teaching interventions (Saine et al., 2010; Saine, Lerkkanen, Ahonen, Tolvanen & Lyytinen, 2011).

The revised adaptation of GraphoGame used in the study contained 13 different tasks on common literacy precursors and reading-related skills, presented in an interactive, game-like environment. As the aim of the study was to develop valid assessment measures for the GraphoGame technology, the aspect of adapting to a child's zone of proximal development was left out. Two game versions were created by mixing the order of the tasks to avoid systematic interference in the task results (Table 1). Half the participants played version 1, whereas the other half played version 2. The GraphoGame program carefully recorded each participant's results, play time seconds and other necessary aspects in the tasks.

Table 1. Task order in game versions 1 and 2

Version 1	**Version 2**
Grapheme-phoneme correspondence 1	Receptive vocabulary
Word recognition	Pseudo-word recognition
Short-term memory	Sentence reading fluency 2
Phonological awareness	Phonological memory
Spelling task 1	Grapheme-phoneme correspondence 2
Sentence reading fluency 1	Short-term memory
Receptive vocabulary	Spelling task 2
Phonological memory	Phonological awareness
Pseudo-word recognition	Grapheme-phoneme correspondence 1
Grapheme-phoneme correspondence 2	Word recognition
Spelling task 2	Spelling task 1
Sentence reading fluency 2	Sentence reading fluency 1
Phonemic length discrimination	Phonemic length discrimination

Computerized measures

Grapheme-phoneme correspondence. The GraphoGame test pattern used in the study encompassed two letter-knowledge tasks. In the grapheme-phoneme correspondence task 1, the participant hears a sound of a letter, i.e. phoneme, from their headphones. They then choose the corresponding grapheme from the computer screen, where all Finnish letters (n=23) are presented at the same time in a randomized order. In the grapheme-phoneme correspondence task 2, the participant similarly hears a phoneme on their headphones and then chooses a corresponding grapheme on the screen, where eight letters are presented simultaneously. The order in which the phonemes are presented is prescribed and the grapheme-stimuli in relation to a certain phoneme are the same for each player. However, the graphemes on the screen are presented in a randomized order, and thus vary from player to player. Both grapheme-phoneme correspondence tasks contain 23 trials, one for each Finnish letter. Numbers of correct answers in each task are used as the outcome scores.

Rapid automatized naming. An adaptation of a Finnish standardization of rapid serial naming test (Ahonen, Tuovinen & Leppäsaari, 1999) was used in the study. Lists of letters (U, I, K, S, T) in the first part of the task and objects (star, fence, hand, worm, and button) in the second part of the task, are presented in 10-item rows on the computer screen, each replicated 4 or 8 times so that none of the items are repeated successively. The participant is instructed to name each letter or object in sequence as accurately and fluently as they can. Their performance is recorded with an open-source speech perception program (Bolaños, 2012), and latencies for each section are used as the outcome scores.

Phonological awareness. In the phonological awareness task, the participant detects differences between words in the word beginnings. The participant hears three different words and is then asked to identify, which of the three words had a different beginning compared the other two. The participant chooses the deviating word by pressing one of three buttons on the screen. The task has a total of 20 trials. Number of correct answers is used as the outcome score.

Phonemic length discrimination. Phonemic length discrimination task requires a child to choose the longer of the two non-words they hear: */ata/* or */atta/*. The task has been developed by Richardson, Thomson, Scott & Goswami (2004) by constructing a sequence of 100 stimuli of the naturally produced non-word /*ata/,* in which the duration of the silent closure of the word medial stop (varying from 65 to 265 ms) was augmented in stepwise fashion with increments of 2 ms. The stimulus with a 65 ms closure duration is used as a standard stimulus. Auditory detection threshold, i.e. the stimulus shorter of which the child is not able to distinguish, is used as the outcome score.

Phonological memory. Phonological short-term memory is measured by presenting sets of three pseudowords or pseudoword series to the participant through their headphones. The participant is asked to detect which of the three pseudoword series differed from the other two. Pseudowords gradually increase in length, ranging from three-syllable words up to 14-syllable pseudoword combinations. Differences between pseudoword series are created by replacing one of the pseudowords in a string with another pseudoword of the same length and syllable structure, but a distinct phonological structure. E.g. /***laa****ppari/* vs. /***kii****ppari/.* The participant chooses the deviating word by pressing one of three buttons on the screen. Number of correct answers is used as the outcome score, and the number of total trials is 18.

Short-term memory. Short-term memory is measured by displaying a set of colors (red, yellow, blue, green black, white) on a computer screen. The participant is asked to select colors in the same order they heard them in their headphones. The length of the series may increase up to 6 colors, but the task is discontinued after three consecutive errors. Number of correct answers is used as the outcome score, and the maximum score is 10.

Receptive vocabulary. The receptive vocabulary test resembles Peabody Picture Vocabulary test by Dunn (1959), although the stimuli used in the current study is different. Four alternative pictures are presented to the participant on the screen, and the participant is instructed to pick the alternative that matches the word they hear from their headphones. The task has a total of 30 trials. Number of correct answers is used as the outcome score.

Spelling. The spelling task presents a set of sublexical items (phonemes and/or syllables) on the screen. Participant is instructed to construct the word they hear from their headphones by choosing the right items in the right order on the screen. Length of the target word and size of sublexical items vary, and options also include distractor items. Real words (spelling task 1) and nonwords (spelling task 2) are used in the tests. Numbers of correct answers are used as the outcome scores. The total trial amount is 20 for word spelling and 24 for pseudoword spelling.

Word recognition. In the word recognition task, the participant chooses a word they hear from four alternative spellings on the screen. Length of the words increases gradually from short, two-syllable words up to 6-syllable words (max. 16 letters). The words, and the four word alternatives, are presented in the same order for each participant. The total trial amount is 40, but a cut-off of 4 consecutive errors is used. Reaction accuracy and latency are used as the outcome scores.

Pseudoword recognition. The pseudoword recognition task begins with one-syllable pseudowords and develops gradually to multisyllabic pseudowords (maximum of five syllables and 15 letters). Four alternative pseudoword spellings are presented on the screen and the child is instructed to choose the one that matches to the pseudoword they heard. The words, and the four word alternatives, are presented in the same order for each participant. The total trial amount is 40, but a cut-off of 4 consecutive errors is used. Reaction accuracy and latency are used as the outcome scores.

Sentence reading fluency. Luksu sentence reading task is a Finnish adaptation of the Woodcock-Johnson reading fluency task (Woodcock, McGrew, & Mather, 2001). In the task, participant is told to read a sentence as rapidly as they can, and to decide whether the given statement (e.g. strawberries are blue) is true or false. The outcome score is the amount of correct answers given within a two minutes' time limit. The tasks used in the study include 1) sentence reading with normal letter spacing and 2) sentence reading with large letter spacing. In both tasks, the maximum score is 70.

Traditional pen-and-paper measures

Letter naming. The traditional letter naming task origined from a previous version of GraphoGame, but was used in this study as a pen-and-paper measure. In the task, the participant is instructed to name the letters they recognize in a list displaying all Finnish letters (n=23) in a randomized order. Both the name of the letter and its phoneme are considered suitable answers. The number of correct answers is used as the outcome score.

Rapid automatized naming. A Finnish standardization of rapid serial naming test (Ahonen, Tuovinen & Leppäsaari, 1999) was used in the study. Lists of letters (O, A, S, T, P) in the first part of the task and objects (car, house, fish, pen, ball) in the second part of the task, are presented in 10-item rows on a paper, each replicated 4 or 8 times so that none of the items are repeated successively. The participant is instructed to name each letter or object in sequence as accurately and fluently as they can. Naming time is recorded with a stopwatch and used as the outcome measure.

Phonological awareness. The phonological awareness task used in the study belonged to the NEPSY II –test pattern (Korkman, Kirk & Kemp, 2007). The task has two parts: word segment recognition and phonological segmentation. The first part requires a participant to identify a correct picture from three alternatives. The pictures represent different words that are orally presented to the participant, and the participant is instructed to detect a certain word segment that belongs to one of the words, e.g. ***–ah*** in *r**ah**a* (*money* in Finnish). In the second part, the participant is asked to repeat a word and then to create a new word by omitting a certain syllable or a phoneme, or by substituting one phoneme in a word for another, e.g. *takka* without *–t* becomes *akka*. The maximum score is 53, but the children start according to their age-norm from the item number 23, thus receiving 22 points automatically (unless if the child gives an incorrect answer to either of the first two questions in the word segment part, items are gone backwards until the child reaches two consecutive correct answers, and in this case, the starting score might become less than 22). The task is discontinued after six consecutive errors.

Phonemic length discrimination. The phonemic length discrimination task was identical to the computerized measure (presented afore under computerized measures), merely played on a laptop instead of a tablet.

Phonological memory. Phonological memory was measured with a NEPSY II pseudoword repetition test (Korkman, Kirk & Kemp, 2007), in which a participant first hears a pseudoword from their headphones and is then asked to repeat it. The task includes 16 pseudowords, but the task is discontinued after four consecutive errors. The number of correct answers is used as the outcome score.

Short-term memory. Digit span forwards and Digit span backwards from the WISC-IV test pattern (Wechsler, 2003) were used to measure short-term memory. In the forwards recall section, the participant is to repeat 3–9 numbers correctly in the same order they heard them, and in the backwards recall, 2–9 digits correctly in a backwards order. In both sections, the child has two trials with each stimulus amount. The task is discontinued when the child makes a mistake in both of the trials. The maximum score in each section is 16.

Receptive vocabulary. The Peabody Picture Vocabulary test by Dunn (1959) was used to measure receptive vocabulary. In the task, four alternative pictures are presented to the participant on paper, and the participant is instructed to pick the alternative that matches the word they heard from their instructor. The maximum trial amount is 30, but the task is discontinued after four consecutive errors. The number of correct answers is used as the outcome score.

Word spelling. The word spelling task belonged to the Lukilasse 2 –test pattern (Häyrinen, Serenius-Sirve & Korkman, 2013). In the task, the participant is to write down 20 words that are presented to them from an audio track. Each word is repeated two times. The maximum score is 40; 2 for each word. A wrong or missing letter in a word brings the word score to 0, and a correct letter, but a deviation from its writing form gives one point.

Pseudoword spelling. The pseudoword spelling task was created for the ReadAll –research project. The stimuli used in this task matched to the stimuli in the computerized version with their syllable structure and also had slightly similar phonological structures, e.g. *rimmes* vs. *vennas*. The task contained 10 pseudowords and the maximum score was 10.

Word recognition. Lukilasse 2 graded word-level reading fluency test (Häyrinen, Serenius-Sirve & Korkman, 2013) was used to measure word recognition skills. The test consists of 90 Finnish words ranging from one-syllable words to multisyllabic words. The participant is instructed to read the words out loud as rapidly and accurately as they can within a two-minute time limit. Number of correctly read words is used as the outcome score. This test was used as the selection criterion in forming the at-risk and control groups (see *Strongest predictors of word-level reading fluency* in the Results-section).

Pseudoword reading. The pseudoword reading task was created for the research project based on LukiMat reading assessment materials for second grade (Salmi, Eklund, Järvisalo & Aro, 2011). 30 stimuli were selected from the materials based on the following selection criteria: 1) all one-syllable words were excluded, 2) the selected words had to be in accordance with the Finnish phonotax, and 3) the syllable structures had to match to those in the computerized version, and they were required to contain similar phonemes. The maximum score was 30, and reading latency was recorded with a stopwatch.

Sentence reading fluency. Luksu sentence reading task is a Finnish adaptation of the Woodcock-Johnson reading fluency task (Woodcock, McGrew, & Mather, 2001). In the task, participant is told to read a sentence as rapidly as they can, and to decide whether the given statement (e.g.

strawberries are blue) is true or false. The outcome score is the amount of correct answers given within a two minutes' time limit. The tasks used in the study include 1) sentence reading with normal letter spacing and 2) sentence reading with large letter spacing. In both tasks, the maximum score is 70.

RESULTS

Missing data and descriptive analysis

Preliminary analyses included missing data analyses and examination of variable distributions. Data was found missing on two variables: thirty percent (n=61) of the data in the computerized letter-naming RAN-task and 30.5 per cent (n=62) of the computerized object-naming RAN-data were missing. In the letter-naming task, 30 (49.2%) of the missing cases were boys and 7 (11.5%) of the missing cases belonged to the at-risk group. In the object-naming task half (n=31) the missing cases were boys and 8 (12.9%) cases belonged to the at-risk group. In order to ensure that the missing data was not systematically related to any of the measures used in the study, several t-tests with missingness (1 = group with data missing, 2 = group with no missing data) as an independent variable and all the other measures as dependent variables were performed. In both the letter and object naming RAN tasks, t-tests revealed a significant difference between the groups on merely one variable; the computerized phonological memory task (Table 2). In both occasions, equal variances were not assumed (Levene's test), as the variance in the missing data -group appeared greater.

Table 2. T-test results of the *Rapid Automatized Naming (RAN) missing data* –variables and phonological memory (PM).

Variables	Group	n	M	t	df	p
RAN letter missing data – PM	Data missing	61	9.84	-2.723	99.266	.008
	No data missing	142	11.64			
RAN object missing data – PM	Data missing	62	9.69	-3.079	101.11	.003
	No data missing	141	11.72			

Examinations of distributions indicated normal or close to normal distributions for the majority of the variables (see Table 3). However, ceiling effects were found in the letter-naming tasks, both in computerized and traditional measures. In traditional tasks, distributions of word spelling and pseudoword reading accuracy tasks were skewed toward lower values, and distribution of phonological memory variable indicated a skew toward higher values. Distributions of computerized RAN measures also showed a skew toward higher values, similarly to the traditional RAN object naming task. Distributions of both the phonemic length discrimination variables were slightly skewed toward higher values. Transformations were not made, as logistic regression analysis, used in the study, does not make strict assumptions on normal distributions (Nummenmaa, 2010).

Few outliers were found, and they were excluded from the data when an experimental error could be tracked; e.g. a score above the measurement scale. In cases where outliers were considered to occur due variability, they were retained. Data was administered in a pairwise manner; the maximum amount of data available was utilized in the analyses. The computerized and traditional measures were somewhat correlated within themselves and with each other (see Table 4).

Correlations between traditional and pen-and-paper tests

Pearson correlations were used to study the first research question: the overall validity of the computerized measures in comparison to the traditional pen-and-paper measures. Correlations between computerized measures and corresponding pen-and-paper tests mostly indicated a moderate to high (0.44–0.79) correlation (Table 4). The highest correlations were found between sentence reading fluency tasks, both with large and normal spacing in the text. Phonemic length discrimination tasks also demonstrated a remarkable correlation, but the test used was identical to what pupils conducted as part of the tablet tests; merely played on a laptop. Low (<0.3) correlations were observed between latency measures of pseudoword recognition and the traditional pseudoword reading task, grapheme-phoneme correspondence tasks, and phonological memory tasks. Digit span forward correlated moderately to the computerized short-term memory task, but digit span backwards only showed a weak correlation. This was likely due to the fact that the computerized short-term memory task did not contain a section with a backwards recall.

Owing to the main focus of the present study; which computerized measures best discriminate between children with a risk for reading difficulties and the control group, no further investigations on the correspondence of the traditional and computerized measures were conducted. Being a crucial topic, however, deeper insights into the matter will be shared under the ReadAll–project in the near future.

Table 3

Descriptive statistics of reading measures in the at-risk and control group

	At-risk (n= 36/ *n= 27/**n=28/***n=34)						Control (n=167/*n=114/**n=114/***n=166)					
Measures	Maximum	Minimum	M	SD	Skewness	Kurtosis	Maximum	Minimum	M	SD	Skewness	Kurtosis
A) Computerized measures												
RAN object naming*	70	36	47.8	9.1	1.26	1.11	89	29	45.6	9.8	1.46	3.80
RAN letter naming**	46	23	30.5	6.1	1.03	0.40	49	17	26.7	6.0	1.28	2.62
Receptive vocabulary	27	15	19.9	3.0	0.35	-0.38	28	10	20.4	3.5	-0.23	-0.22
Word recognition, accuracy	36	2	15.0	8.5	0.77	0.05	38	2	25.3	8.1	-0.46	-0.72
Word recognition, latency	225	18	104.5	60.8	0.70	-0.59	254	21	124.4	45.2	0.17	-0.45
Pseudoword recogn., acc.	34	3	16.3	8.0	0.58	-0.41	39	2	25.3	8.1	-0.58	-0.34
Pseudoword recogn., laten.	273	25	121.2	65.6	0.79	0.13	312	23	139.0	48.2	0.15	0.47
Short-term memory***	9	4	6.2	1.3	-0.05	-0.55	10	0	6.5	1.9	-1.08	2.11
PA	20	5	13.1	4.3	-0.30	-0.86	20	3	15.3	4.0	-1.14	0.74
Word spelling	19	4	13.3	3.9	-1.01	0.36	20	5	15.1	3.3	-0.93	0.50
Sentence reading fluency	35	9	20.8	7.2	0.20	-1.00	64	14	34.3	9.3	0.33	0.22
Phonological memory	17	0	9.8	5.1	-0.50	-1.13	17	1	11.4	3.9	-0.99	0.43
Pseudoword spelling	19	2	12.6	4.0	-0.84	0.42	24	5	16.3	4.1	-0.73	-0.29
Sentence reading, sparce text	39	8	19.1	5.9	1.01	2.48	56	15	31.9	8.0	0.39	-0.07
Phonemic length discrim.	100	16.8	57.4	31.6	0.30	-1.68	100	8.9	35.7	23.5	1.46	1.15
Grapheme-Phoneme corr. 1	23	3	20.3	3.6	-3.46	15.23	23	15	21.1	1.7	-1.03	1.35
Grapheme-Phoneme corr. 2	23	14	20.6	2.0	-1.06	1.72	23	16	21.1	1.5	-0.85	0.59
B) Pen-and-paper measures												
RAN object naming	115	40	60.9	19.1	1.56	1.71	125	34	57.1	14.6	1.70	4.33
RAN letter naming	52	27	38.1	6.9	0.54	-0.46	58	20	31.4	6.6	0.87	1.31
Receptive vocabulary; PPVT	24	2	11.2	5.4	0.35	-0.34	27	1	11.7	6.1	0.24	-0.97
Word recognition score/2min[a]	55	27	45.4	7.3	-0.91	0.24	90	51	76.0	12.1	-0.36	-1.22
Pseudoword reading score	29	7	24.0	4.4	-2.00	5.65	30	15	27.3	2.8	-1.91	4.49
Pseudoword reading time	238	53	104.5	40.8	1.80	3.96	107	27	54.8	17.2	0.84	0.52
STM: digit strings onwards	8	3	5.6	1.2	-0.58	0.16	12	3	6.2	1.2	1.27	3.34
STM digit string backwards	8	3	5.2	1.2	0.46	0.29	9	3	5.8	1.1	0.14	-0.07
PA; Nepsy	47	23	38.2	5.0	-0.68	1.08	53	21	43.6	5.1	-0.94	1.47
Word spelling	40	14	31.9	6.9	-1.09	0.41	40	14	37.0	4.5	-2.38	6.46
Sentence reading fluency	32	5	15.3	5.1	0.57	2.33	50	1	25.7	8.0	0.47	0.67
Phonological memory	13	4	9.7	2.1	-1.06	1.43	16	2	10.1	2.2	2.04	20.7
Pseudoword spelling	10	2	6.5	2.0	-0.04	-0.70	10	3	8.2	1.9	-0.92	-0.27
Sentence reading, sparce text	31	5	14.6	4.5	0.97	4.16	44	9	25.1	7.1	0.39	-0.09
Phonemic length discr.	100.0	4.7	42.3	31.3	0.10	-0.51	99.9	3.5	24.8	21.4	2.05	3.81
Letter naming	23	20	22.2	0.9	-0.61	-0.73	23	20	22.7	0.7	-2.02	3.71

Notes. [a]The measure is the Lukilasse II graded word-level reading fluency test used as a criterion variable in forming the at-risk and control groups.

Table 4
Pearson correlations between computerized measures and pen-and-paper measures

	A) 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	B) 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A) Computerized measures																																	
1 RAN object naming	1																																
2 RAN letter naming	0.55**	1																															
3 Receptive vocabulary	-0.27**	-0.02	1																														
4 Word recognition, acc.	-0.36**	-0.15	0.22**	1																													
5 Word recognition, lat.	-0.31**	-0.01	0.17*	0.79**	1																												
6 Pseudoword recog., acc.	-0.29**	-0.10	0.26**	0.62**	0.39**	1																											
7 Pseudoword recog., lat.	-0.11	0.11	0.16*	0.40**	0.46**	0.79**	1																										
8 Short-term memory	-0.15	0.06	0.22**	0.29**	0.20**	0.32**	0.24**	1																									
9 Phonological awareness	-0.34**	-0.18*	0.32**	0.41**	0.19**	0.42**	0.22**	0.36**	1																								
10 Word spelling	-0.24**	0.00	0.29**	0.49**	0.33**	0.43**	0.28**	0.27**	0.44**	1																							
11 Sentence reading fluen.	-0.22*	-0.19*	0.26**	0.45**	0.05	0.42**	0.03	0.20**	0.41**	0.28**	1																						
12 Phonological memory	-0.31**	-0.15	0.24**	0.28**	0.16*	0.34**	0.22**	0.43**	0.46**	0.33**	0.17*	1																					
13 Pseudoword spelling	-0.32**	-0.16	0.31**	0.52**	0.32**	0.56**	0.36**	0.28**	0.48**	0.61**	0.33**	0.43**	1																				
14 Sparce sentence r.fluen.	-0.29**	-0.28**	0.32**	0.47**	0.07	0.50**	0.08	0.21**	0.46**	0.36**	0.86**	0.27**	0.40**	1																			
15 PLD	0.35**	0.18*	-0.12	-0.44**	-0.25**	-0.49**	-0.28**	-0.35**	-0.45**	-0.38**	-0.36**	-0.38**	-0.46**	-0.43**	1																		
16 Grapheme-Phoneme 1	-0.23**	-0.23**	0.13	0.23**	0.18*	0.24**	0.20**	0.16*	0.32**	0.29**	0.09	0.26**	0.33**	0.18**	-0.12	1																	
17 Grapheme-Phoneme 2	-0.19*	-0.13	0.18*	0.21**	0.13	0.39**	0.26**	0.23**	0.29**	0.35**	0.12	0.39**	0.41**	0.21**	-0.22**	0.51**	1																
B) Pen-and-paper measures																																	
1 RAN object naming	0.57**	0.28**	-0.07	-0.25**	-0.18*	-0.23**	-0.16*	-0.24**	-0.23**	-0.15*	-0.12	-0.26**	-0.24**	-0.14*	0.24**	-0.21**	-0.17*	1															
2 RAN letter naming	0.56**	0.54**	-0.05	-0.32**	-0.11	-0.25**	0.03	-0.18*	-0.28**	-0.14*	-0.34**	-0.17*	-0.20**	-0.35**	0.32**	-0.14*	-0.19**	0.39**	1														
3 Receptive vocabulary	-0.17*	0.01	0.46**	0.15*	0.10	0.12	0.02	0.19**	0.23**	0.17*	0.24**	0.16*	0.27**	0.25**	-0.13	0.07	0.03	-0.13	-0.04	1													
4 Word recognition	-0.26**	-0.29**	0.23**	0.50**	0.07	0.54**	0.09	0.19**	0.36**	0.32**	0.76**	0.20**	0.41**	0.80**	-0.44**	0.14*	0.21**	-0.25**	-0.44**	0.16*	1												
5 Pseudoword read. acc.	-0.26**	-0.18*	0.20**	0.48**	0.29**	0.53**	0.29**	0.17*	0.33**	0.36**	0.40**	0.27**	0.41**	0.43**	-0.48**	0.14*	0.21**	-0.14	-0.25**	0.05	0.53**	1											
6 Pseudoword read. time	0.32**	0.40**	-0.16*	-0.40**	-0.07	-0.43**	-0.08	-0.11	-0.23**	-0.23**	-0.54**	-0.17*	-0.33**	-0.64**	0.32**	-0.10	-0.17*	0.34**	0.53**	-0.10	-0.79**	-0.45**	1										
7 STM: onward recall.	-0.24**	-0.04	0.23**	0.26**	0.08	0.27**	0.05	0.44**	0.37**	0.30**	0.28**	0.36**	0.37**	0.25**	-0.32**	0.16*	0.22**	-0.21**	-0.22**	0.23**	0.25**	0.21**	-0.11	1									
8 STM backward recall	-0.14	-0.08	0.29**	0.34**	0.18*	0.32**	0.21**	0.27**	0.37**	0.29**	0.32**	0.31**	0.35**	0.30**	-0.22**	0.12	0.18*	-0.13	-0.17*	0.18*	0.28**	0.32**	-0.19**	0.36**	1								
9 Phonological awareness	-0.06	-0.08	0.32**	0.41**	0.12	0.53**	0.26**	0.30**	0.56**	0.43**	0.54**	0.48**	0.51**	0.62**	-0.51**	0.22**	0.27**	-0.08	-0.19**	0.31**	0.55**	0.48**	-0.32**	0.33**	0.45**	1							
10 Word writing	-0.29**	-0.14	0.13	0.48**	0.28**	0.52**	0.30**	0.25**	0.46**	0.55**	0.39**	0.36**	0.50**	0.48**	-0.54**	0.22**	0.34**	-0.18*	-0.18**	0.08	0.48**	0.62**	-0.35**	0.29**	0.28**	0.56**	1						
11 Sentence reading fluen.	-0.23**	-0.23**	0.27**	0.40**	0.01	0.39**	0.02	0.26**	0.38**	0.31**	0.74**	0.21**	0.30**	0.75**	-0.33**	0.11	0.13	-0.27**	-0.34**	0.22**	0.75**	0.34**	-0.55**	0.32**	0.31**	0.49**	0.33**	1					
12 Phonological memory	-0.18*	-0.14	0.26**	0.07	-0.04	0.22**	0.13	0.18*	0.32**	0.13	0.20**	0.29**	0.19**	0.21**	-0.23**	0.15*	0.16*	-0.11	0.02	0.13	0.18**	0.27**	-0.08	0.24**	0.21**	0.28**	0.25**	0.21**	1				
13 Pseudoword spelling	-0.21*	-0.11	0.23**	0.51**	0.29**	0.56**	0.33**	0.30**	0.52**	0.53**	0.43**	0.32**	0.48**	0.45**	-0.44**	0.24**	0.32**	-0.10	-0.22**	0.17*	0.45**	0.55**	-0.29**	0.36**	0.30**	0.52**	0.72**	0.38**	0.30**	1			
14 Sparce sentence r. fluen.	-0.30**	-0.31**	0.28**	0.42**	0.01	0.43**	0.04	0.26**	0.40**	0.32**	0.77**	0.26**	0.37**	0.79**	-0.39**	0.15*	0.19**	-0.28**	-0.36**	0.25**	0.79**	0.38**	-0.60**	0.33**	0.36**	0.53**	0.41**	0.89**	0.19**	0.42**	1		
15 PLD	0.28**	0.09	-0.12	-0.40**	-0.23**	-0.44**	-0.29**	-0.30**	-0.39**	-0.40**	-0.31**	-0.44**	-0.43**	-0.38**	0.71**	-0.26**	-0.30**	0.28**	0.18*	-0.15*	-0.37**	-0.42**	0.28**	-0.30**	-0.19**	-0.47**	-0.48**	-0.29**	-0.21**	-0.40**	-0.38**	1	
16 Letter knowledge	-0.08	0.01	0.08	0.27**	0.09	0.34**	0.17*	0.21**	0.34**	0.25**	0.30**	0.20**	0.21**	0.29**	-0.29**	0.15*	0.19**	-0.10	-0.12	0.04	0.32**	0.29**	-0.10	0.18*	0.18*	0.34**	0.38**	0.28**	0.18*	0.43**	0.29**	-0.29**	1

Notes. *p < .05 **p <.01. Abbreviations: *acc.*=accuracy, *lat.*=latency, *fluen.*=fluency, PLD = Phonemic length discrimination.

Strongest predictors of word-level reading fluency

Logistic regression analyses were conducted to answer the second research question, i.e. to determine the strongest predictors of word-level reading fluency in the data. The participants were divided into two groups: the ‘at-risk for reading difficulties’ group and the control group. The groups were formed based on children’s test scores, which were compared to relevant age-norms (Häyrinen, Serenius-Sirve & Korkman, 2013) to form standardized scores for each child. Children with a z-score of -1,33 or below (n=36 / 17,7%), belonged to the ‘at-risk’ group, and children with a z-score above -1,33 belonged to the control group (n=167). A dummy variable ‘reading level’ was formed as a result.

In the logistic regression analysis, the variable ‘reading level’ was placed as the dependent variable, while computerized measures were introduced to the model one or two at a time, using the enter-method. The following variables indicated a significant effect ($p < .001$ with $df = 1$) individually: word recognition accuracy, pseudoword recognition accuracy, sentence reading fluency, sentence reading fluency with large letter spacing, phonemic length discrimination and pseudoword spelling. Among these, sentence reading fluency with large letter spacing was able to predict and explain cases most effectively ($\chi^2 = 80.072$, $p < .001$ with $df = 1$), identifying 20 of the 36 cases (55.6%) in the at-risk group, and 159 of the 167 (95.2%) of the cases in the control group correctly (overall sensitivity = 88.2 %). Nagelkerke’s R^2 of .54 also indicated a strong relationship between the task scores and grouping. Thus, sentence reading fluency task with large letter spacing was chosen as a predictor for the model.

As many of the computerized task variables also correlated with each other (Table 4), not all variables that initially showed a significant effect could be added to the model. Ultimately, the highest coefficient of determination (Nagelkerke’s $R^2 = .59$) could be found when sentence reading fluency with large spacing, word recognition accuracy and rapid automatized object naming task were included in the model ($\chi^2 = 64.803$, $p < .001$ with $df = 3$) (see Table 5). This model could correctly identify 19 of the 27 (70.4%) cases in the at-risk group, and, respectively, 111 of the 114 (97.4%) in the control group (overall sensitivity = 92.2%). One must note that when RAN object naming was added to the model, the group sizes dropped from 36 to 27 in the at-risk group and from 167 to 114 in the control group, due to the missing data.

Table 5. Strongest discriminators of word-level reading fluency skills using logistic regression analysis.

Predictor	*β*	*SE β*	*Wald*	*df*	*p*	*Odds ratio*
Constant	-9.401	2.443	14.805	1	.000	.000
Sentence reading fluency, sparce	.263	.060	19.197	1	.000	1.301
Word recognition accuracy	.089	.038	5.556	1	.018	1.093
RAN Object naming	.052	.030	2.898	1	.089	1.053

DISCUSSION

The present study examined discriminators of word-level reading fluency skills among Finnish second-graders using computerized assessment. The two research questions included: 1) How well do the computerized reading assessment measures correspond to the more traditional pen-and-paper measures and 2) Which of the computerized reading assessment measures best discriminate children with poor word-level reading fluency skills and the control group.

To answer the first research question, correlations between the computerized and traditional measures were observed. The correlations demonstrated a moderate to high correlation for a great majority of the measures; indicating a rather good overall correspondence between the computerized and traditional tasks. Similarly to earlier findings of Suokas (2009), high correlations could be witnessed between the traditional and computerized sentence reading fluency tasks, both with large and normal letter spacing. Low correlations were found between letter naming tasks, phonological memory variables and pseudoword recognition/pseudoword reading latencies. As perceived in earlier studies (Ketonen, 2010), children on their second-grade had a good knowledge of the Finnish alphabet – and thus ceiling effects could be observed in the task results. This might have inflated the correlation sizes between the letter naming tasks. When it comes to the phonological memory variables, the computerized measure differed from the traditional pseudoword repetition task in the sense that it only required to detect an auditory difference between words; not to repeat and produce the entire word correctly. The distribution of the traditional phonological memory task was skewed toward higher values, as opposed to the computerized task, indicating that the task was experienced more difficult than the computerized equivalent. The latency measures of pseudoword recognition/reading tasks, then, might differ for two reasons. Firstly, the computerized tasks used a cut-off of four consecutive errors. Thus, those experiencing difficulties in pseudoword recognition were likely to obtain lower values in the latency measure, whereas in the traditional pseudoword reading task, difficulties might have led to a higher latency score due to slow or laborious pseudoword recognition (Hautala & Parviainen, 2014; Moll, Hutzler & Wimmer, 2005). Second, as the computerized tasks required children to detect the right word from four different alternatives, as opposed to merely reading one word in the traditional measures, the reading times might partly reflect strategies used in detecting differences between the words. Overall, it can be concluded that based on the observed correlations, a great majority of the computerized tasks can be considered as good or moderately valid assessment measures. A few measures showed low correlations, but this might have occurred due to practical differences in the tasks.

As for the second research question, which computerized measures best discriminate children who are at risk for reading difficulties and the control group, the following variables were found: word recognition accuracy, sentence reading fluency with sparce letter spacing, and rapid automatized naming object task. Previous studies have shown visual word recognition a highly stable indicator of reading skills in transparent orthographies (Lander & Wimmer, 2008), and a crucial element for reading fluency (Perfetti, 1992; van den Boer & de Jong, 2015). Results of the present study support these notions, but it is worth noticing that the correlation of the traditional and computerized word recognition measures only reached a moderate level (r = 0.5). The computerized and traditional word recognition tasks differed in the way that in the traditional task, oral reading of words from a wordlist was expected – whereas in the computerized task, the participant was to pick a matching written alternative to a word they heard from four different options on the screen. This difference might have lowered the observed correlation, despite them both measuring word recognition skills. However, the validity of the measure requires further investigation, and thus, conclusions should be treated with caution.

The sentence reading fluency tasks used in study measured fluency with simple sentences, thus relying mostly on automatic word recognition and mechanical reading skills – and to some degree, comprehension. Mechanical reading and word-level reading fluency were central to the fluency measure used as a criterion variable, and thus, the tasks likely required slightly overlapping skills, as can be concluded from their high correlation (r = 0.8). The fact that the sentence reading fluency task with sparce letter spacing turned out a better discriminator of fluency skills than the equivalent with normal spacing, is contrary to earlier studies indicating extra-large letter spacing improves reading speed of dysfluent readers (Zorzi et al., 2012). One viable explanation to this is that the large letter-spacing also improved reading outcomes in the control group, as school-aged children in general are known to be affected by crowding (Jeon, Hamid, Maurer & Lewis, 2010). Standard deviations were slightly smaller in the task with large letter-spacing, which might support this notion, but the maximum score in the control group was also considerably lower compared to the task with normal letter spacing, and thus, absolute interpretations cannot be made. Another, contrary explanation, is that if the spacing seems excessively broad on a tablet screen, it might appear as too wide and thus lower reading speed, as indicated by some studies (Pelli et al., 2007; Yu, Cheung, Legge & Chung, 2007). The topic still requires further investigation.

Rapid automatized naming has been considered among the strongest and most stable predictors of reading skills in transparent orthographies (Holopainen, Ahonen, & Lyytinen, 2001; Kairaluoma, et al., 2013; Puolakanaho et al., 2007; Torppa et al., 2010; Salmi, 2008; Scarborough, 1998). Results of the current study support the strong role of RAN in discriminating reading fluency skills. In the present study, the RAN task only included the naming latency, not the accuracy, and the data contained some tampering effects, such as additional time in a number of recordings – and of course, the data loss. In spite of this, the RAN-task was able to correctly identify cases between the at-risk and control groups; thus, it potentially has an even greater effect on word-level fluency skills than what was observed in this study.

All in all, the three discriminators that best differentiated between the at-risk and control groups were mostly in line with earlier reading fluency research. Belonging to a certain group could be predicted well based on the three measures; they were highly specific in terms of the control group. The measures were also sensitive in identifying participants in the at-risk group, but could only reach 70% of the cases, thus leaving room for more definite estimators. This might reflect variance in children's reading development; while some of the children have shifted from relying on phonological skills to fluency-related skills on grade two, dysfluent readers might still depend on phonological skills – and many of the children would place somewhere in between. In the current study, none of the phonological measures (PA, PM, PLD) showed ceiling effects, but rather, variation in phonological skills. However, the only phonological variable that significantly explained variation between the at-risk and control groups was phonemic length discrimination. This could be interpreted as lending support to previous studies indicating that the shift towards fluency-related skills is evident already on second-grade in Finland (Torppa et al., 2013). It is crucial to note, however, that although the criterion variable took reading accuracy into account and mainly required mechanical reading skills, it was a fluency measure, and thus was likely to somewhat affect the eventually selected strongest discriminators. Considering the slight bias caused by the criterion variable and the fact that ceiling effects could not be observed in phonological skills, the results also pose the question of whether fluency measures are sensitive enough to discriminate children with reading difficulties on grade two, and whether aspects of reading accuracy should also be included in reading skills assessment among young children in transparent orthographies.

The study contains some further limitations that need to be considered when interpreting the results. Firstly, the study was conducted in Finland, and transparency of the Finnish language as well as Finnish educational settings might drastically differ from those of other countries. Second, despite the observed correlations, the overall validity of all the computerized measures should be inspected more thoroughly. Computerized reading assessment is in its early stages, and developing valid assessment methods requires time and sufficient research. In addition, some of the variables that initially explained variation between the two groups were excluded after sentence reading fluency was selected. Although some of the excluded variables evidently measured similar aspects of reading skills (for ex. sentence reading fluency with normal spacing); seemingly different variables had also to be overlooked (f.ex. phonemic length discrimination). In future research, it would be beneficial to examine whether the measured skills are truly overlapping, an autoregressor effect can be found, or whether the computerized measures still need some modifications.

Additional limitations regard regression analysis. In the present study, the standard method (enter) of logistic regression analysis was used. Stepwise-methods would have led to slightly more variables in the model, but might also have caused problems to reliable interpretation, and likely biased results due to the small at-risk group in the study (Steyerberg, Eijkemans & Habbema, 1999). Logistic regression analysis always requires rather large sample sizes for both groups, as it uses maximum likelihood estimation techniques. Even using the enter-method, the great difference in the size of the at-risk and control groups might have hindered the effect of some of the variables used in the study. Also, one-third of the data in two variables had gone missing on the course of the data collection, due to errors in saving the data. Although the loss was detected random in this study; the amount of lost data was considerable, and lead to a smaller at-risk group (n=27) in the analyses.

This study examined computerized reading fluency assessment among Finnish second-graders. Overall, the computerized assessment measures compared to the traditional pen-and-paper measures well. Word recognition, sentence reading fluency and rapid automatized naming were identified as the strongest discriminators of word-level reading fluency – which can be considered to stand in line with previous studies conducted in transparent orthographies. The study raises further topics of investigation, including: a thorough inspection of validity of the computerized measures, examination of possible test-mode effects e.g. in sentence reading fluency tasks, and whether fluency measures are sensitive enough to differentiate dysfluent readers on second-grade in transparent orthographies. Research should also be carried out in different age groups, countries and various orthographies. The results of the study are encouraging in terms of computerized reading assessment; developing valid computerized reading assessment methods is not only attainable, but could soon be accessible to psychologists, teachers, and other educators. Computerized reading assessment measures have numerous applications ranging from mapping individual problem areas for intervention purposes, to scanning reading skills of certain age groups. Computerized measures also enable collecting data on test variables in great detail; thus opening new dimensions for literacy research.

REFERENCES

Ahonen, T., Tuovinen, S., & Leppäsaari, T. (1999). *Nopean sarjallisen nimeämisen testi*. Jyväskylä: School of Haukkaranta.

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). Hillsdale, NJ: Lawrence Erlbaum.

Aro, M., Huemer, S., Heikkilä, R. & Mönkkönen, V. (2011). Sujuva lukutaito suomalaislapsen haasteena. *Psykologia 46(02-03), 153-155.*

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*(04), 621-635. DOI: 10.1017.S0142716403000316

Babayigit, S. & Stainthorp, R. (2011) Modeling the relationships between cognitive–linguistic skills and literacy skills: New insights from a transparent orthography. *Journal of Educational Psychology, Vol 103(1), 169-189*. DOI: 10.1037/a0021671

Baddeley, A. (2012). Working memory: theories, models, and controversies.*Annual review of psychology*, *63*, 1-29. DOI: 10.1146/annurev-psych-120710-100422

Bhide, A., Power, A., & Goswami, U. (2013). A Rhythmic Musical Intervention for Poor Readers: A Comparison of Efficacy With a Letter-Based Intervention. *Mind, Brain, and Education*, *7*(2), 113-123. DOI: 10.1111/mbe.12016

Breznitz, Z. (2006). *Fluency in reading: Synchronization of processes*. Mahwah, NJ: Erlbaum.

Bolaños, D. (2012). The Bavieca Open-Source Speech Recognition Toolkit. In *Proceedings of IEEE Workshop on Spoken Language Technology (SLT).* Miami, FL.

Clariana, R. & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test-mode effect. *British Journal of Educational Technology,* 33 (5), 593-602. DOI: 10.1111/1467-8535.00294

De Luca, M., Borrelli, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2002). Reading Words and Pseudowords: An Eye Movement Study of Developmental Dyslexia. *Brain and language, 80,* 617-626. DOI: 10.1006/brln.2001.2637

Dandache, S., Wouters, J., Ghesquière, P. (2014). Development of Reading and Phonological Skills of Children at Family Risk Dyslexia: A Longitudinal Analysis Kindergarten to Sixth Grade. *Dyslexia, 20, 305-329.* DOI: 10.1002/dys.1482

Dunn, L. M. (1959). Peabody Picture Vocabulary Test. Minneapolis: American Guidance Service. *Inc., l959*.

Eklund, K., Torppa, M., Aro, M., Leppänen, P. H., & Lyytinen, H. (2015). Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and 8. *Journal of Educational Psychology*, *107*(1), 126–140. DOI: 10.1037/a0037121

Escribano, C. (2007). Evaluation of the double-deficit hypothesis subtype classification of readers in Spanish. *Journal of Learning Disabilities*, *40*, 319–330. DOI:10.1177/00222194070400040301

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A. M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship. *Applied cognitive psychology*, *13*(1), 65-77. DOI: 10.1002/(SICI)1099-0720(199902)13:1<65::AID-ACP548>3.0.CO;2-O

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

Goldberg, A. L., & Pedulla, J. J. (2002). Performance differences according to test mode and computer familiarity on a practice graduate record exam. *Educational and Psychological Measurement*, *62*(6), 1053-1067. DOI: 10.1177/0013164402238092

Hammill, D. D. (2004). What we know about correlates of reading. *Exceptional Children*, *70*(4), 453-469.

Hautala, J., Aro, M., Eklund, K., Lerkkanen, M. K., & Lyytinen, H. (2013). The role of letters and syllables in typical and dysfluent reading in a transparent orthography. *Reading and Writing*, *26*(6), 845-864. DOI: 10.1007/s11145-012-9394-3

Hautala, J., & Parviainen, T. (2014). Gaze Position Reveals Impaired Attentional Shift during Visual Word Recognition in Dysfluent Readers. DOI: 10.1371/journal.pone.0108937

Hemphill, L., & Tivnan, T. (2008). The importance of early vocabulary for literacy achievement in high-poverty schools. *Journal of Education for Students Placed at Risk*, *13*(4), 426-451. DOI: 10.1080/10824660802427710

Holopainen, L., Ahonen, T., & Lyytinen, H. (2001). Predicting delay in reading achievement in a highly transparent language. *Journal of Learning Disabilities*, *34*, 401–413. DOI: 10.1177/002221940103400502

Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2008). The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly*, *25*(1), 4-32. DOI: 10.1080/10573560802491208

Huemer, S. (2009). *Training reading skills: Towards fluency*. Jyväskylä: Jyväskylä University.

Hutzler, F., & Wimmer, H. (2004). Eye movements of dyslexic children when reading in a regular orthography. Brain and Language, 89(1), 235-242. DOI: 10.1016/S0093-934X(03)00401-2

Häyrinen, T., Serenius-Sirve S., & Korkman, M. (2013). Lukilasse 2. Lukemisen, kirjoittamisen ja laskemisen seulontatesti 1.–6. vuosiluokille. Helsinki: Hogrefe Psykologien Kustannus Oy.

Häyrinen, T., Serenius-Sirve, S., Korkman, M., Ekebom, U. M., Kivelä, A., Lassander, P., & Repokari, L. (1999). Lukilasse. Lukemisen, kirjoittamisen ja laskemisen seulontatestistö peruskoulun ala-asteen luokille 1-6. Helsinki: Psykologien kustannus.

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

Jeon, S. T., Hamid, J., Maurer, D., & Lewis, T. L. (2010). Developmental changes during childhood in single-letter acuity and its crowding by surrounding contours. *Journal of experimental child psychology*, *107*(4), 423-437. DOI: 10.1016/j.jecp.2010.05.009

Jere-Folotiya, J., Chansa-Kabali, T., Munachaka, J. C., Sampa, F., Yalukanda, C., Westerholm, J., ... & Lyytinen, H. (2014). The effect of using a mobile literacy game to improve literacy levels of grade one students in Zambian schools. *Educational Technology Research and Development*, *62*(4), 417-436. DOI: 10.1007/s11423-014-9342-9

Kairaluoma, L., Torppa, M., Westerholm, J., Ahonen, T., & Aro, M. (2013). The nature of and factors related to reading difficulties among adolescents in a transparent orthography. *Scientific Studies of Reading*, *17*(5), 315-332. DOI: 10.1080/10888438.2012.701257

Ketonen, R. (2010). Dysleksiariski oppimisen haasteena: fonologisen tietoisuuden interventio ja lukemaan oppiminen. Jyväskylä: Jyväskylä University.

Kibby, M. Y. (2009). There are multiple contributors to the verbal short-term memory deficit in children with developmental reading disabilities. *Child Neuropsychology, 15, 485–506.* DOI: 10.1080/09297040902748218

Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology, 100*, 310–321. doi:10.1037/0022-0663.100.2.310

Korkman, M., Barron-Linnankoski, S., & Lahti-Nuuttila, P. (1999). Effects of age and duration of reading instruction on the development of phonological awareness, rapid naming, and verbal memory span. *Developmental Neuropsychology*, *16*(3), 415-431. DOI: 10.1207/S15326942DN1603_24

Korkman, M., Kirk, U., & Kemp, S. (2007). NEPSY-II: Clinical and interpretive manual. *San Antonio, TX: The Psychological Corporation*.

Kuhn, M. R., & Stahl, S. (2003). Fluency: A review of developmental and remedial practices. *The Journal of Educational Psychology, 95,* 3–21. DOI:10.1037/0022-0663.95.1.3

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive psychology*, *6*(2), 293-323.

Lai, S. A., Benjamin, R. G. Schwanenflugel, P. J. & Kuhn, M. R. (2014). The Longitudinal Relationship Between Reading Fluency and Reading Comprehension Skills in Second-Grade Children. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 30:2, 116-138.* DOI: 10.1080/10573569.2013.789785

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

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H., Lohvansuu, K., ... & Schulte-Körne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry*, *54*(6), 686-694. DOI: 10.1111/jcpp.12029

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, Vol 100(1), Feb 2008, 150-161. DOI: 10.1037/0022-0663.100.1.150

Leinonen, S., Müller, K., Leppänen, P. H. T., Aro, M., Ahonen, T., & Lyytinen, H. (2001). Heterogeneity in adult dyslexic readers: Relating processing skills to the speed and accuracy of oral text reading. *Reading and Writing: An Interdisciplinary Journal*, *14*, 265–296. DOI:10.1023/A:1011117620895

Leppänen, U., Niemi, P., Aunola, K., & Nurmi, J.-E. (2006). Development of reading and spelling Finnish from preschool to Grade 1 and Grade 2. Scientific Studies of Reading, 10, 3–30. DOI: 10.1207/s1532799xssr1001_2

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

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53,* 1–14. DOI:10.1007/s11881-003-0001-9

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*, 668–675. DOI: 10.1111/j.1467-9450.2009.00791.x

Lyytinen, H., & Richardson, U. (2014). Guest Editors' Introduction: Supporting Urgent Basic Reading Skills in Children in Africa and Around the World. *Human Technology*, 10 (1), 1-4. DOI: 10.17011/ht/urn.201405281856

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities, 17, 592–599.* DOI: 10.1177/002221948401701005

Martens, V. E., & de Jong, P. F. (2006). The effect of word length on lexical decision in dyslexic and normal reading children. *Brain and Language*, *98*(2), 140-149. DOI: 10.1016/j.bandl.2006.04.003

Martens, V. E., & De Jong, P. F. (2008). Effects of repeated reading on the length effect in word and pseudoword reading. *Journal of Research in reading*,*31*(1), 40-54. DOI: 10.1111/j.1467-9817.2007.00360.x

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*(2), 322. DOI: 10.1037/a0026744

Miller, J. & Schwanenflugel, P.J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly,* 43(4), 336-354. DOI: dx.doi.org/10.1598/RRQ.43.4.2

Moll, K., Hutzler, F., & Wimmer, H. (2005). Developmental dyslexia in a regular orthography: A single case study. *Neurocase*, *11*(6), 433-440. DOI: 10.1080/13554790500263537

Nation, K., & Cocksey, J. (2009). The relationship between knowing a word and reading it aloud in children's word reading development. *Journal of experimental child psychology, 103*(3), 296-308. DOI: 10.1016/j.cognition.2008.11.004

Nithart, C., Demont, E., Metz-Lutz, M. N., Majerus, S., Poncelet, M., & Leybaert, J. (2011). Early contribution of phonological awareness and later influence of phonological memory throughout reading acquisition. *Journal of research in reading, 34*(3), 346-363. DOI: 10.1111/j.1467-9817.2009.01427.x

Noell, G. H., Connell, J. E., & Duhon, G. J. (2006). Spontaneous response generalization during whole word instruction: Reading to spell and spelling to read. *Journal of Behavioral Education, 15, 121–130.*

Norton, E. & Wolf, M. (2012). Rapid Automatized Naming (RAN) and Reading Fluency: Implications for Understanding and Treatment of Reading Disabilities. *Annual Reviews of Psychology, 63,* 427–52. DOI: 10.1146/annurev-psych-120710-100431

Nummenmaa, L. (2010). Käyttäytymistieteiden tilastolliset menetelmät. Helsinki: Kariston Kirjapaino Oy.

O'Connor, R. E., White, A., & Swanson, H. L. (2007). Repeated reading versus continuous reading: Influences on reading fluency and comprehension.*Exceptional Children*, *74*(1), 31-46. DOI: 10.1177/001440290707400102

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, Vol 98(3), 554-566.* DOI: 10.1037/0022-0663.98.3.554

Pagliuca, G., & Monaghan, P. (2010). Discovering large grain sizes in a transparent orthography: Insights from a connectionist model of Italian word naming. *European Journal of Cognitive Psychology*, *22*(5), 813-835. DOI:10.1080/09541440903172158

Papadopoulos, T. C., Georgiou, G., & Kendeou, P. (2009). Investigating the double-deficit hypothesis in Greek: Findings from a longitudinal study. Journal of Learning Disabilities, 42, 528–547. DOI: 10.1177/0022219409338745

Parrila, R., Kirby, J.R., & McQuarrie, L. (2004). Articulation Rate, Naming Speed, Verbal Short-Term Memory, and Phonological Awareness: Longitudinal Predictors of Early Reading Development? *Scientific Studies of Reading, 8:1, 3-26.* DOI: 10.1207/s1532799xssr0801_2

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of vision,7*(2), 20. DOI: 10.1167/7.2.20

Pennala, R., Eklund, K., Hämäläinen, J., Martin, M., Richardson, U., Leppänen, P. H., & Lyytinen, H. (2013). Precursors and Consequences of Phonemic Length Discrimination Ability Problems in Children With Reading Disabilities and Familial Risk for Dyslexia. *Journal of Speech, Language, and Hearing Research, 56*(5), 1462-1475. DOI: 10.1044/1092-4388(2013/12-0072)

Pennala, R., Eklund, K., Hämäläinen, J., Richardson, U., Martin, M., Leiwo, M., ... & Lyytinen, H. (2010). Perception of phonemic length and its relation to reading and spelling skills in children with family risk for dyslexia in the first three grades of school. *Journal of Speech, Language, and Hearing Research,53*(3), 710-724. DOI: 10.1044/1092-4388(2009/08-0133)

Perfetti, C. A. (1985). *Reading ability*. New York, US: Oxford University Press.

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly (1982-)*, 283-319.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In *P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), Reading acquisition (pp. 145–174).* Hillsdale, NJ: Lawrence Erlbaum.

Phillips, B. M., & Torgesen, J. K. (2006). Phonemic awareness and reading: Beyond the growth of initial reading accuracy. In *Dickinson, D.K. & Neuman, S.B. (edit.) Handbook of early literacy research*, *2*. pp. 101-112. Guilford Publications.

Protopapas, A. & Skaloumbakas, C. (2007). Traditional and Computer-Based Screening and Diagnosis of Reading Disabilities in Greek. *Journal of Learning Disabilities 40 (2007).* DOI: 10.1177/00222194070400010201

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppanen, P. H., Poikkeus, A. M., et al. (2007). Very early phonological and language skills: estimating individual risk of reading disability. Journal of Child Psychology and Psychiatry, 48, 923e931. http://dx.doi.org/10.1111/j.1469-7610.2007.01763.x.

Puolakanaho, A., Ahonen, T., Aro, M. Eklund, K. Leppänen, P.H.T., Poikkeus, A-M., Tolvanen, A., Torppa, 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.* DOI: 10.1177/0022219407311747

Richardson, U., Thomson, J. M., Scott, S. K., & Goswami, U. (2004). Auditory processing skills and phonological representation in dyslexic children. *Dyslexia*,*10*(3), 215-233. DOI: 10.1002/dys.276

Saine, N.L., Lerkkanen, M-K., Ahonen, T., Tolvanen, A. Lyytinen, H. (2010) Predicting word-level reading fluency outcomes in three contrastive groups: Remedial and computer-assisted remedial reading intervention, and mainstream instruction. *Learning and individual differences 20 (2010) 402-414.* DOI: 10.1016/j.lindif.2010.06.004

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

Salmi, P. (2008). Nimeäminen ja lukemisvaikeus: kehityksen ja kuntoutuksen näkökulma. Jyväskylä: Jyväskylä University.

Salmi, P., Eklund, K., Järvisalo, E. & Aro, M. LukiMat - Oppimisen arviointi: Lukemisen ja kirjoittamisen tuen tarpeen tunnistamisen välineet 2. luokalle. Käyttäjän opas. Retrieved from: http://www.lukimat.fi/lukimat-oppimisen-arviointi/materiaalit/ tuen-tarpeen-tunnistaminen/2lk/lukeminen/kayttajan-opas (*User's guide)*
Materials from: http://www.lukimat.fi/lukimat-oppimisen-arviointi/materiaalit/tuen-tarpeen-tunnistaminen/2lk/lukeminen/lomakkeet/FI_Lukeminen_2012Syksy_2lk_Tehtvmateriaali_Yksil.pdf (*Last accessed Sept 26, 2015)*

Salmi, P., & Torppa, M. (2011). Miten nopea nimeäminen ennakoi lukutaitoa. *Psykologia 46 (2011): 2-3.*

Scarborough H. (1998). Predicting the future achievement of second graders with reading disabilities: contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia* 48, *115–36*. DOI: 10.1007/s11881-998-0006-5

Seymour, P.H.K., Aro M., Erskine J.M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology. 94:143–74.* DOI*:* 10.1348/000712603321661859

Seymour, P.H.K. (2005). Early reading development in European orthographies. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 296–315). Malden, MA: Blackwell.

Share, D. 2008. On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological bulletin, 134, 584-615.* DOI: 10.1037/0033-2909.134.4.584

Sprugevica, I., & Hoien, T. (2003). Early phonological skills as a predictor of reading acquisition: A follow-up study from kindergarten to the middle of grade 2. *Scandinavian Journal of Psychology*, *44*(2), 119-124. DOI: 10.1111/1467-9450.00329

Steyerberg, E. W., Eijkemans, M. J., & Habbema, J. D. F. (1999). Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *Journal of clinical epidemiology*, *52*(10), 935-942. DOI: 10.1016/S0895-4356(99)00103-1

Swanson, H. L., Kehler, P., & Jerman, O. (2010). Working memory, strategy knowledge, and strategy instruction in children with reading disabilities. *Journal of Learning Disabilities, 43, 24–47*. DOI: 10.1177/0022219409338743

Swanson, H. L., Jerman, O., & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, *100*(2), 343–379. DOI: 10.1037/0022-0663.100.2.343

Tobia, V. & Marzocchi, G. M. (2014). Predictors of reading fluency in Italian orthography: Evidence from a cross-sectional study of primary school students. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence, 20:4, 449-469.* DOI: 10.1080/09297049.2013.814768

Torppa, M., Lyytinen, P., Erskine, J., Eklund, K., & Lyytinen, H. (2010). Language development, literacy skills, and predictive connections to reading in Finnish children with and without familial risk for dyslexia. *Journal of learning disabilities*. DOI: 10.1177/0022219410369096

Torppa, M., Parrila, R., Niemi, P., Lerkkanen, M. K., Poikkeus, A. M., & Nurmi, J. E. (2013). The double deficit hypothesis in the transparent Finnish orthography: a longitudinal study from kindergarten to Grade 2. *Reading and Writing*, *26*(8), 1353-1380. DOI 10.1007/s11145-012-9423-2

Vaessen, A., & Blomert, L. (2010). Long-term cognitive dynamics of fluent reading development. *Journal of experimental child psychology*, *105*(3), 213-231. DOI: 10.1016/j.jecp.2009.11.005

Vaessen A., Bertrand D., Tóth, D., Csépe V, Faísca L, Reis, Reis, A. & Blomert, L. (2010). Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. *Journal of Educational Psychology* 102 (4), 827–42. DOI: 10.1037/a0019465

Van Bergen, E., de Jong, P. F., Maassen, B., & van der Leij, A. (2014). The Effect of Parents' Literacy Skills and Children's Preliteracy Skills on the Risk of Dyslexia. *Journal of abnormal child psychology*, 1-14. DOI: 10.1007/s10802-014-9858-9

Van den Boer, M., & de Jong, P. F. (2015). Parallel and serial reading processes in children's word and nonword reading. *Journal of Educational Psychology*, *107*(1), 141. DOI:

Van den Bos K.P., Zijlstra B.J.H. & Spelberg H.C. (2002). Life-span data on continuous-naming speeds of numbers, letters, colors, and pictured objects, and word-reading speed. *Scientific Studies of Reading.* 6:25–49. DOI: 10.1207/S1532799XSSR0601_02

Veenendaal, N. J., Groen, M. A., & Verhoeven, L. (2014). What oral text reading fluency can reveal about reading comprehension. *Journal of Research in Reading*. DOI: 10.1111/1467-9817.12024

Wang, S., & Gathercole, S. E. (2013). Working memory deficits in children with reading difficulties: memory span and dual task coordination. *Journal of experimental child psychology*, *115*(1), 188-197. DOI: 10.1016/j.jecp.2012.11.015

Wechsler, D. (2003). Wechsler intelligence scale for children–Fourth Edition (WISC-IV). *San Antonio, TX: The Psychological Corporation*.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology: Section A*, *50*(2), 439-456. DOI: 10.1080/713755710

Whitney, D., & Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences*,*15*(4), 160-168. DOI: 10.1016/j.tics.2011.02.005

Wilson, A.J., Andrewes, S. G., Struthers, H., Rowe, V. M., Bogdanovic, R., & Waldie, K. E. (in press). Dyscalculia and dyslexia in adults: Cognitive bases of comorbidity.*Learning and Individual Differences.*

Wimmer, H. & Schurz, M. (2010). Dyslexia in regular orthographies: Manifestation and causation. *Dyslexia, 16, 283–299.* DOI: 10.1002/dys.411

Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities. *Journal of Speech, Language, and Hearing Research*, *50*(4), 1093-1109. DOI: 10.1044/1092-4388(2007/076)

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III.

Wolf, M. & Katzir-Cohen, T. (2001). Reading Fluency and Its Intervention. *Scientific Studies of Reading, 5:3, 211-23.* DOI: 10.1207/ S1532799XSSR0503_2

Yang, Y. T. C. (2012). Building virtual cities, inspiring intelligent citizens: Digital games for developing students' problem solving and learning motivation.*Computers & Education*, *59*(2), 365-377. DOI: 10.1016/j.compedu.2012.01.012

Yu, D., Cheung, S. H., Legge, G. E., & Chung, S. T. (2007). Effect of letter spacing on visual span and reading speed. *Journal of Vision*, *7*(2), 2. DOI: 10.1167/7.2.2

Ziegler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3–29. DOI:10.1037/0033-2909.131.1.3

Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal?. *Journal of experimental child psychology*, *86*(3), 169-193. DOI: 10.1016/S0022-0965(03)00139-5

Zoccolotti, P., De Luca, M., Judica, A., & Spinelli, D. (2008). Isolating global and specific factors in developmental dyslexia: A study based on the rate and amount model (RAM). *Experimental Brain Research*, *186*(4), 551-560. DOI: 10.1007/s00221-007-1257-9

Zoccolotti, P., De Luca, M., Marinelli, C. V., Spinelli, D., Unit, N., & Santa Lucia, I. F. (2014). Modeling individual differences in text reading fluency: a different pattern of predictors for typically developing and dyslexic readers. *Frontiers in Human Neuroscience*, *8*, 793. DOI: 10.3389/fpsyg.2014.0137

Zorzi, M., Barbiero, C., Facoetti, A., Lonciari, I., Carrozzi, M., Montico, M., ... & Ziegler, J. C. (2012). Extra-large letter spacing improves reading in dyslexia. *Proceedings of the National Academy of Sciences*, *109*(28), 11455-11459. DOI: 10.1073/pnas.12055661