Author(s): Kronberg, Nea; Aro, Mikko; Eklund, Kenneth; Lehecka, Tomas; Vataja, Pia; Salmi, Paula

Title: Computer-assisted morphology training of reading with grade 2 and grade 3 poor readers

Year: 2024

Version: publishedVersion

Please cite the original version:

Computer-assisted morphology training of reading with grade 2 and grade 3 poor readers

Nea Kronberg[1,2] · Mikko Aro[1,3] · Kenneth Eklund[4] · Tomas Lehecka[5] · Pia Vataja[1,2] · Paula Salmi[2,6]

© The Author(s) 2024

Abstract

Explicit morphology instruction has been recommended for poor readers. Still, hardly any studies have compared the effectiveness of both explicit and implicit morphology training in relation to business-as-usual support in improving reading proficiency. In this study, 277 poor readers in Grades 2 and 3 participated in a computer-assisted eight-week intervention focusing on suffixes in a randomised control trial, with pre-, post- and follow-up assessment in schools for Swedish-speaking children in Finland. Students were randomly allocated into two experimental groups and one control group. The first experimental group played a digital game with explicit instruction in lexical morphology (explicit). The other experimental group played the game without said instruction (implicit). Intervention and Grade effects were analysed using mixed-design analysis of variance (ANOVA). A significant time x group intervention effect was found only in the implicit vs. control group comparison between pre- and post-assessments; a moderate gain in identification of trained words assessed in a game-like manner was found in the implicit group unlike in controls. The difference in gain remained significant at follow-up for both grades. Additionally, third graders in the implicit group demonstrated a more significant improvement than controls in reading aloud pseudowords with trained suffixes, both in the pre–post comparison and at follow-up. Results suggest that poor readers benefitted from implicit morphology training, but explicit instruction in lexical morphology (animations with verbal explanations) did not have a significant effect on reading proficiency.

Extended author information available on the last page of the article

∅ Springer

Introduction

Recent research has highlighted morphology training (MT) as a method for supporting reading development after basic decoding skills are acquired (Castles et al., 2018; Rastle, 2019; Tijms et al., 2020). MT is a broad concept that covers various means of training aimed at improving reading skills. In this article Morphology Training primarily focuses on enhancing reading skills, through training reading and identification of morphemes as orthographic units (cf. morphological decoding, Levesque et al., 2021). A distinct categorisation of MT approaches is challenging because different approaches overlap with each other (Goodwin & Ahn, 2013). One approach of MT has aimed at enhancing the recognition of morphemes as orthographic units (Tijms & Hoeks, 2005). A second approach has focused on the semantic aspects of morphemes (Arnbak & Elbro, 2000). A third approach has emphasised morphological rules and syntax as the core of training (Berninger et al., 2008; Nunes et al., 2003). These three perspectives resemble the three dimensions (morphological decoding, morphological awareness and morphological knowledge) in the Morphological Pathways Framework by Levesque et al. (2021) which has been implemented to other languages e.g. Norwegian by Kristensen (2023). More research is, however, needed on how, when and why MT can support reading development (Bowers et al., 2010; Carlisle et al., 2010; Rastle, 2019; Reed, 2008). When considering the form of MT (how), an essential question pertains to the role and need of instruction: Is explicit instruction needed (and in what form), or can implicit MT also be an alternative? Intervention studies are needed to examine the causal links between increased morphological knowledge and reading development (Bowers et al., 2010). In this article, we present an intervention study where the effects of targeted reading practice, i.e. training of decoding skills, with a focus on morphology (i.e. suffixes), on reading skill was investigated using an online educational game with Swedish-speaking poor readers in Grades 2 and 3. We focused especially on answering two of the questions proposed for future research (Rastle, 2019; Reed, 2008). These were: first, the question of how—i.e., which type of support (implicit vs. explicit) would be effective for students—and second, the question of when—i.e., in which grade (2nd or 3rd) this support could be effective. First, we examined how the form of the training, i.e. implicit training or explicit instruction, affected reading outcomes, especially accuracy and speed in reading words. Second, we examined whether the grade, in other words, the instructional level of individuals, is related to the effects of the training by including both second and third graders as participants in the study and using the participants' grade as a control variable in the models.

The participants in this intervention study were poor readers beyond the initial stages of reading instruction. Once readers master basic decoding, reading develops when readers gather orthographic knowledge through independent reading (Share, 1995). For some readers, this process is hard and difficulties often persist (Eklund et al., 2015; Landerl & Wimmer, 2008; Torppa et al., 2015). Finnish–Swedish children with poor skills in both reading fluency and spelling at the end of Grade 1 have also been shown to remain poor at both skills at the end of Grade 3 (Risberg et al., 2023). Among Swedish children, 60% of those experiencing difficulties in reading in Grade 2 remained poor readers in Grade 9 (Jacobson, 1999). Similarly, 59% of

Finnish-speaking children with poor reading skills in Grade 2 still had reading difficulties in Grade 8 (Torppa et al., 2015). The persistence of difficulties (Eklund et al., 2015), despite the support offered at school, underlines the need for more efficient approaches for supporting struggling readers (Risberg et al., 2023).

Morphological knowledge supporting reading development

Levesque et al. (2021) have constructed the Morphological Pathways Framework in order to illustrate how morphological knowledge influences literacy and the development of reading and spelling. A study by Kristensen et al. (2023) established that this framework is a good model also for the Norwegian language, whereby we can assume that it also might fit the closely related Swedish language, since Swedish and Norwegian are mutually intelligible North Germanic languages, sharing similar grammatical categories and affixes (Haugen, 1976; Wessén, 1969). In this model there are three, reciprocally related dimensions between morphology and literacy. In morphological decoding separate morphemes in words are recognized and in morphological analysis the meaning of morphemes within words are recognized. The second dimension, morphological awareness (MA) implies a conscious reflection on and manipulation of morphemes. Morphological knowledge (MK) refers to the ability to recognize, understand, manipulate and produce spoken and written morphemes (Kristensen et al., 2023). These two terms are closely related as explicit MK is often called MA (Castles et al., 2018). In this article the concepts of morphological knowledge (MK) and morphological awareness (MA) are used according to the references.

Other studies also show that the relationship between MK and reading can be reciprocal (Kuo & Anderson, 2006; Nunes et al., 2006). After basic decoding skill is acquired, further reading development requires a shift from decoding based on single graphemes to identifying larger graphophonemic units, such as morphemes (Ehri, 1998, 2020; Seymour et al. 2003). By consolidating larger units, the reader decreases the load on the working memory and is able to attain faster and more accurate reading (LaBerge & Samuels, 1974). Usually, as basic reading skill is mastered, MK is gathered implicitly through reading and writing experience (Castles et al., 2018; Nunes et al., 2006). Though MK often develops without explicit instruction (Bowers et al., 2010), the implicitly acquired skills can result in explicit knowledge (Torkildsen et al., 2021). For poor readers, however, the development of MK might be slower due to less exposure to printed words (Castles et al., 2018; Nunes et al., 2006; Reed, 2008) compared with typical readers.

The first morphological elements to be consolidated are thought to be common affixes (prefixes and suffixes) (Ehri, 1998). However, struggling with consolidation of morphological units can form an obstacle for further reading development (Tijms, 2004), so MT can benefit students' reading through enhancing word recognition skills (Bowers et al., 2010). The consolidation of morphological units can be supported by increasing children's understanding of the features of morphology on a sublexical level in both written and oral language, which in turn influences reading skills (Bowers et al., 2010). Once the reader recognises the regularity between the morpheme, e.g. an affix, and its meaning, orthographic learning is then no longer word-specific

and decoding of unfamiliar words can be supported by the familiar morphemes in the words (Castles et al., 2018; Goodwin & Ahn, 2013).

A meta-analysis by Goodwin and Ahn (2013) showed that the relationship between MA and reading skill was stronger for younger students than middle school and upper elementary school students. The effect of MA on reading has also been studied in correlative designs (Deacon, 2012; Kirby et al., 2012). Kirby et al. (2012) showed that among English-speaking children, the effect of MA on reading tended to increase from Grade 1 to 3, in other words, with increasing reading level and reading experience. In the study by Kirby et al. (2012) MA was related to all five measures of reading accuracy and speed in Grade 3, whereas in Grade 1, it made no significant contribution to reading measures, and in Grade 2, MA was related to only four measures with smaller effects than in Grade 3. However, Deacon (2012) found no differences between grade 1 and 3 regarding how MA affects reading among Canadian children.

With reference to the Morphological Pathways Framework Kristensen et al. (2023) state that growth in MA could also impact morphological decoding, which can enhance word reading. Consequently, they recommend the utilization of a morphological approach in reading interventions. Several meta-analyses have studied the gains of MT for students with varied reading skills and effects of MT have often been stronger for poor than average readers (Bowers et al., 2010; Goodwin & Ahn, 2010). However, we do not know exactly why this is the case or through which mechanisms MK influences reading (Castles et al., 2018; Levesque et al., 2021). It has been suggested that MK could function as a compensatory approach to reading for poor readers (Bowers et al., 2010; Casalis et al., 2004; Goodwin & Ahn, 2010), especially for those who may not have benefited from phonological training (Reed, 2008). It is not clear, either, whether and to what extent students' reading level moderates the association between MK and reading (Castles et al., 2018; Levesque et al., 2021). A correlational study by Deacon (2012) considered the impacts of both phonological and morphological awareness on early word reading, in grades 1–3. The results showed that for students with weaker phonological awareness the contribution of MA was higher on both real word and nonword reading. These results support the idea presented by Bowers et al. (2010), Casalis et al. (2004) and Reed (2008) that MK and MA (Deacon, 2012) could function as a compensatory strategy for poor readers.

Explicit and implicit approaches to morphology training

Most research on morphological training (MT) focuses on implicit learning, whereas studies on explicit MT are limited. Both approaches have also been applied to enhance reading acquisition (Rastle et al., 2021). Explicit instruction involves teaching and explaining the learning content prior to independent practice, while implicit training relies on learning through experience, where exercises are completed without prior explanation of their purpose.

Implicit learning has been shown to be beneficial, especially if specific meanings of morphological units are difficult to explain and the units occur frequently in variable contexts (Torkildsen et al., 2021). Thus, learning may be promoted by arranging systematic encounters with these morphological structures, for example, as interac-

tive exercises in learning apps. Bar-Kochva et al. (2020) found effects on spelling accuracy after computer-assisted implicit MT with poor spellers in German, but no effects on reading speed and accuracy in either trained or untrained words. A larger study with Norwegian second and third graders examined the effects of a digital game consisting of various tasks with a focus on morphological regularities (e.g. affixes) without any explicit instruction (Torkildsen et al., 2021). Gains on reading speed and accuracy of trained words were found after the intervention and the effects remained at follow-up. Additionally, transfer effects were found for reading untrained words with trained affixes directly after the intervention, but not at follow-up. An implicit approach was also used in a small study with poor spellers in Greek in Grades 5 and 6 (Tijms et al., 2020). The computerised programme visualised derivational and inflectional changes, but no separate instruction was involved. Directly after the intervention, significant transfer effects with large effect sizes were found on two standardised measures of word reading accuracy and speed and text reading speed.

Explicit instruction of morphological regularities in words has been recommended, especially for poor readers (Bowers et al., 2010), based on the presumption of their slower development of MK (Bowers et al., 2010; Reed, 2008). Poor readers are less exposed to words, which is needed for the ability to generalise MK to other words (Reed, 2008). They might also lack the ability to use implicit knowledge of morphology as consistently as average readers do (Bar-Kochva et al., 2020; Tijms & Hoeks, 2005). The literature review by Bowers et al. (2010) demonstrated that, indeed, there are stronger effects of explicit MT among poor readers than among average readers. However, whether this could be also related to a lower level of MK among poor readers remains unknown, as they did not report outcome levels of MA. Their hypothesis was that explicit instruction could endorse faster learning and help form morphological strategies that are more effective for poor readers. One of the few studies concerning explicit MT regarding reading was carried out with Dutch dyslexic children (Tijms & Hoeks, 2005). Explicit instruction of phonological and morphological structures was provided mainly in a graphic form, followed by training in the recognition and use of these structures. At the post-test, the children had significantly improved in both word and text reading rate as well as text reading accuracy. They had even achieved normal levels of text reading accuracy, but not word and text reading rate, on standardised tests. However, it is not possible to determine what was specifically responsible for the gains in reading, phonological training or MT or both.

To the best of our knowledge, only few studies have compared the explicit and implicit approaches with regard to reading outcomes, and neither of these have focused on children with poor reading skills. In a relatively small study with 16 children diagnosed with language impairment, explicit MT focusing on orthographic patterns and meanings of affixes did not yield any effects on the reading accuracy of either trained or untrained words as compared with the controls who received implicit training (Good et al., 2015). However, in a study of adults learning to read words in an artificial language, the benefits of explicit instruction of spelling-sound-regularities over implicit learning were shown to be strong (Rastle et al., 2021). With regard to spelling, explicit and implicit approaches have been compared in two studies (Kemper et al., 2012; Bryant et al., 2006; Burton et al. 2021). Both studies showed

larger gains for the students who had received explicit instruction of morphological spelling rules in comparison to students who received implicit training.

The aforementioned studies have investigated the effects of the morphology interventions using various measures. The studies differ in whether the measures used assess trained, untrained (near transfer) or standardized (far transfer) outcomes, or combinations of these. Additionally, research varies based on the type of items measured (i.e. words vs. pseudowords). The measures used also differ in their focus, with some examining reading accuracy and others reading time or a combination of these. Studies examining the effects of interventions focusing on word and text reading have yielded results in word reading using trained measures and untrained measures (Torkildsen et al., 2021). Additionally, results have been observed in standardized tests for both word reading (Tijms et al., 2020) and text reading (Tijms et al., 2020; Tijms & Hoeks, 2005). In some studies, different types of measures have been combined, making it difficult to precisely assess the nuanced effects of the intervention. For example, some studies have combined reading and spelling (Bar-Kochva et al., 2020; Torkildsen et al., 2021), while others have combined near transfer and far transfer within the same measure (Good et al., 2015). Additionally, there are studies that have merged time and accuracy into a single measure, assessing the number of words read correctly within a time limit (Bar-Kochva et al., 2020; Tijms et al., 2020; Torkildsen et al., 2021).

To our knowledge, no research has specifically investigated reading progress using pseudowords as a measure or separately examined the impact of all the factors mentioned above. Additionally, only a few studies have compared explicit and implicit methods for reading outcomes.

The effects of age and reading level on the outcomes of morphology training

Children seem to possess MA even before learning to read, and there is evidence for explicit MK in the initial stages of reading acquisition (Kirby et al., 2012; Rastle, 2019). MT has also been recommended for older children who have passed the initial stages of literacy development (Rastle, 2019). Tijms et al. (2020) completed a computer-based intervention with 11–12-year-old (Grades 5 and 6) poor spellers in the transparent Greek language. The participants showed significant gains in word reading and text reading fluency (accuracy and speed) after the intervention where morphological changes were visualised in a computerized training programme. Recommendations for an older target group could be based on the presumption that the ability to generalise MK increases with age and exposure to words that are morphologically more complex (Reed, 2008). In other words, the reader needs to master basic decoding and have acquired some experience with varied texts.

Evidence varies on the effects of MT on reading skills for different age groups (Rastle, 2019). A literature review by Bowers et al. (2010) comprised studies of the effects of MT on literacy skills in several language groups (i.e. English, Norwegian, Danish), still the number of studies in other languages than English was too small for cross-linguistic comparisons. In this review, positive effects of MT interventions on reading were found among students in Grades 1–3, as well as in Grades 3–8, with higher effect sizes in the lower grades. In a meta-analysis with a focus on morpho-

logical interventions in English with children of different age groups, large effects on reading were found already in preschool and early elementary through Grade 2 (Goodwin & Ahn, 2013). Smaller effects were also found for interventions targeting students in middle school and upper elementary school.

The process of reading acquisition is dependent on the characteristics of orthography (Aro, 2005; Share, 2008). Similar to the transparency of orthography, the characteristics of morphology also vary between languages (Borleffs et al., 2017). It has been suggested that in less transparent orthographies, morphology may play a more important role in reading acquisition (Reed, 2008). Swedish is thought to be positioned in the middle of the continuum from opaque to transparent orthographies, though Swedish readers reach high accuracy in reading after Grade 1, such as readers of more transparent orthographies (Aro & Wimmer, 2003). Hence, Swedish-speaking readers seem to possess sufficient decoding skills already after Grade 1 and reading fluency will be their main challenge after basic decoding skills are acquired. Therefore, MT might be applicable already for students in Grade 2.

It has been suggested that MT needs to be adapted to the age and reading level of the students (Carlisle et al., 2010; Reed, 2008). However, this seems to be complicated as several features in both the target group and the training itself need to be considered. One approach for matching a specific form of training to the right age group is to test the effects of that training with children of different ages.

Aims of the study

The primary purpose of this study was to investigate the effects of two types of game-like interventions for Swedish-speaking poor readers with a game supporting reading through MT. We especially focused on the development of reading accuracy and speed. The intervention was implemented in a school context by special education teachers of students in Grades 2 and 3. These poor readers playing the game in two different conditions with implicit training and explicit instruction were compared with a control group of poor readers not playing the game. The groups with implicit training and explicit instruction were also compared to each other. The controls participated in school instruction and received support as seen necessary by the teachers (business as usual condition). Hence, we studied whether an implicit or explicit training condition in the game, as compared with the controls and each other, was associated with growth in reading skills. In order to examine the long-term intervention effects, we also included follow-up measures.

The specific research questions were as follows

1   What are the short-term effects of the morphology training of reading?
    a   Does either the implicit or the explicit intervention result in greater gain in reading skill when compared with the controls and each other immediately after the intervention?
    b   Do the intervention effects differ according to grade or item (word/pseudoword) and transfer type (near/far)?
2   What are the long-term effects of the morphology training of reading?

a Is reading skill better in either the implicit or explicit group six weeks after the end of the intervention when compared with the controls and each other?

b Do the long-term effects differ according to grade or item (word/pseudo-word) and transfer type (near/far)?

Our hypothesis was that effects would be found in the explicit group rather than in the implicit group, as earlier research has recommended explicit instruction for poor readers and also shown gains from it (Arnbak & Elbro, 2000; Tijms et al., 2020). For the explicit group, we expected gains on the measures in identification of trained words but also on reading aloud the transfer words (words with trained affixes combined with root words with structure and frequency that matched the root words in the training) and pseudowords (words with trained affixes combined with pseudoroot words with a structure matching the trained root words). We expected that the students in the explicit group might be able to more easily generalise the explicit knowledge of the morphological features learned (Tijms et al., 2020).

A hypothesis was not formulated for the research question considering grade effects or item type. There is no previous research comparing specifically grade levels two and three, although there is evidence showing positive relations between MA and reading skill in lower grades (Deacon, 2012; Goodwin & Ahn, 2013). Additionally, there is no intervention research addressing the effects on pseudoword reading. Very few intervention studies in this field have included follow-up measures. We examined the long-term effects as further evidence of the effectiveness of the eventual gains from the intervention.

Method

In this study, MT was implemented in an online computer-assisted game. The Finnish Graphogame platform, with its origin in the findings of the Jyväskylä Longitudinal Study of Dyslexia (Lyytinen et al., 2006), formed the basis for the game. In Finland, a high frequency of digital appliances in public schools is a good premise for utilising digital based-practices (Deloitte & The European Commission, 2019). This intervention study was implemented with Swedish-speaking children in public schools, steered by the national curriculum, with tuition in Swedish. The Finland–Swedish population comprises 5.4% of the population.

The content of the online game was developed especially for this study with game-like training exercises focusing on 16 frequent morphemes (suffixes) in the Swedish language. The intervention consisted of 13 inflectional suffixes and 3 derivational suffixes. Compared with English, Swedish has a more complex inflectional and derivational morphology. Most inflected and derived forms of words are marked through suffixes (i.e. endings). Swedish nouns can be broadly divided into five declination types according to the plural suffix they take: -or, -ar, -er, -n, or no suffix (Teleman et al., 1999). In addition, nouns are inflected for definiteness, where definite forms are marked with the suffixes -en/-n (nouns with gender uter) or -et/-t (nouns with gender neuter) in singular and -na, -a, or -en in plural. Verbs have three tenses: present (marked typically with -er or -r), past (marked typically with -de, -te, or -dde)

and past participle (marked typically with -*t*, -*tt*, or -*it*). Adjectives are inflected for number, gender, definiteness and degree of comparison (typically -*are/-re* in comparative and -*ast/-st* in superlative). New nouns, adjectives and verbs can be derived using derivational affixes (e.g. -*lig/-ig* to form adjectives from nouns, and -*are* or -*ing/-ning* to form nouns from verbs).

Procedure

This study was part of a longitudinal study conducted in Swedish-speaking schools in Finland (Vataja et al., 2021). The current study was implemented as a randomised controlled trial with pre-, post- and follow-up assessments for three groups, with two experimental groups and a control group. The study was carried out in the spring of 2021 in Finland–Swedish public primary schools in Grades 2 and 3, implemented by special education teachers who supervised training and carried out all measurements. By involving second and third graders instead of first graders, we wanted to ensure that the students had consolidated the basic grapheme–phoneme correspondences and mastered basic decoding skill, which is usually reached during Grade (1) More advanced reading instruction, targeting reading fluency and comprehension, starts in the spring of Grade 1 and continues in Grade (2) Within the Finnish curriculum, there is a marked shift after the first two grades from focusing on learning to read to reading to learn, as children are expected to have developed independent reading skills for learning new content.

Assessments were carried out on four different occasions (see Table A1). Screening tests were administered during December 2020. Selected participants completed pre-tests during January–February 2021. Post-tests were administered during March–April 2021, and follow-up test was completed in May 2021.

Participants Recruitment of students

Special education teachers were asked to participate in implementing the intervention at the schools. After the drop-out of three teachers, 44 special education teachers from 31 schools remained until the end of the study.

The screening procedure (see Fig. 1) was carried out by the special education teachers. Results in all three tasks from the Individuell Läsning och skrivning (ILS)-test (Risberg et al., 2019) used in screening were compared with nationally standardised grade-level norms. All students in Grades 2 and 3 (*N*=1598) in the schools of the 44 teachers, excluding children studying Swedish as a second language (S2), were first screened with a group-administered reading fluency task (sentence verification task). Students scoring in the 15th percentile and lower were then assessed indi-

| N = 1,598 | Sentence Verification Task (ILS) Group-administered | 15th percentile n = 500 | Word Reading (ILS) & Text Reading (ILS) Individually administered | 20th percentile n = 296 | Parental consent & eligible | n = 277 |
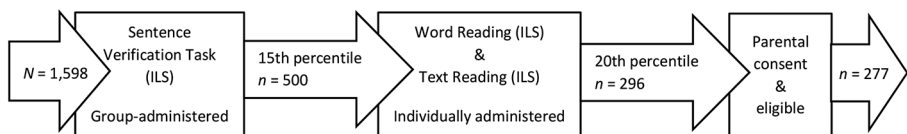
Fig. 1 Screening procedure

vidually with word reading and text reading tasks. Parental consent letters were sent to participants scoring at the 20th percentile and lower in both individually administered reading tasks.

Of the 296 children fulfilling the inclusion criteria, the parents of five students declined consent, and the consent forms of 11 students were not returned. Teachers and researchers agreed to exclude one student due to a lack of motivation and another due to extensive difficulties in decoding. One student with parental consent declined to participate due to reasons unrelated to this study. Thus, the final sample comprised 277 participants. Out of the final 277 participants, 21 students (7.58%) were special needs education students supported at the most intensive level in the Finnish three-tier system of support of learning and schooling. All students studied in mainstream schools, but might have received part-time special education during some point of their time in school. Finland has an extensive special educational system where many children receive special education. The decisions concerning such a low-threshold special educational support are pedagogical and do not require any formal diagnosis. Hence, over 20% of the students receive special educational support at some point of the school year in grades 1–9 (Official Statistics of Finland, 2020). Special education for students with e.g. reading difficulties is given as a pull-out service.

Grouping

After screening, participants were grouped into three groups. To control for skill level, students were first grade-wise sorted in ascending order according to a sum variable of standardised mean scores of both individual screening measures. These grade-wise lists were divided into consecutive triplets of students. Within every triplet, the students were randomly assigned to one of the three groups. No significant differences were found in reading skill and age (in months) between the three groups in Grades 2 and 3 (see Table 1). Experimental Group E (Explicit group) received

Table 1 Descriptive characteristics of the participants and the scores of the screening measures

Variable	Group E (Explicit)	Group I (Implicit)	Control Group	F
	Grade 2			
n	52	53	53	
Girls, *n* (%)	31 (59.61)	28 (52.83)	26 (49.00)	
Boys, *n* (%)	21 (40.38)	25 (47.17)	27 (50.94)	0.51
Age in months, *M* (*SD*)	99.75 (3.48)	100.91 (3.71)	100.70 (3.70)	1.12
ZILS [a], *M* (*SD*)	−1.50 (0.36)	−1.50 (0.37)	−1.52 (0.37)	0.03
Number of trained items *M* (*SD*)	1225.12 (309.86)	1320.57 (287.48)	-	2.68
	Grade 3			
n	40	39	40	
Girls, *n* (%)	23 (57.58)	22 (56.41)	20 (50.00)	
Boys, *n* (%)	17 (42.50)	17 (43.59)	20 (50.00)	0.32
Age in months *M* (*SD*)	113.38 (3.17)	112.33 (3.73)	113.13 (3.78)	0.84
ZILS [a], *M* (*SD*)	−1.65 (0.45)	−1.66 (0.48)	−1.66 (0.49)	0.01
Number of trained items *M* (*SD*)	1473.85 (366.77)	1512.51 (355.90)	-	0.23
n (%)	92 (33.21)	92 (33.21)	93 (33.57)	

[a] Means of the *z*-scores for Word Reading and Text Reading

explicit instruction embedded in the game. Experimental Group I (Implicit group) received exactly the same game tasks without the instruction. The third group, Group C (Control group), served as controls receiving school-provided support as ordinary (Official Statistics of Finland, 2020). The intervention teachers were also encouraged to use the training game after the study with Group C participants.

Measures

The timetable for assessments with instruments used are described in Table A2.

General reading measures

Three measures from the nationally standardised independent instrument, ILS (Risberg et al., 2019) were used for screening for participants in reading fluency. These measures were also used at post-test and follow-up to assess global reading skill and possible generalisation of the training to reading material not used in the training.

Word reading The task assesses students' ability to read highly common words aloud of increasing difficulty. The student read aloud words from a list with a total of 120 words. The test score was the number of correctly read words within a time limit of 45 s.

Text reading The student read aloud an informational text about the fox for one minute. The total number of words in the text is 272 and the test score was the number of correctly read words during the time limit.

Sentence Verification The group-administered task assessed reading fluency. Students read sentences and verified or rejected the truthfulness of each sentence. The test score was the number of correctly verified sentences within two minutes from a total of 70 sentences.

These three general reading measures where combined in the analyses. For each variable, a standard score was calculated based on its pre-test mean and standard deviation. These standard scores were then aggregated into a composite score by calculating the mean of the three standard scores. The reliability for the composite score based on Cronbach's alpha was 0,97.

Reading measures including trained morphology content

The learning and transfer of trained morphological contents were assessed with measures developed for this specific purpose. As a high congruency between training and outcome measures increases the transfer effect (Martin-Chang et al., 2007), two assessments of near transfer were carried out within the game context (the Measure Game), and two were read aloud from printed word lists to assess far transfer. By

transfer we imply whether the training has yielded gains in reading tasks that resemble the training partially (near transfer) or tasks that have less resemblance with the training (far transfer). For example, if reading training completed in the game (not requiring reading aloud), has also yielded gains in other types of reading tasks (reading aloud). The far transfer refers to possible changes in reading the trained suffixes in combination with untrained root-words.

The Word Identification tasks completed in the Measure Game required no reading aloud. In these tasks, the student listened to a word presented auditorily and chose the corresponding written word among four alternatives on the screen. The reading tasks required reading aloud a list of words. All the read-aloud tasks comprised 30 words. Four parallel word lists were used. The words were matched according to specific requirements with a parallel for each word in the other lists. We strived for parallel words to include the same suffix but to also match length, consonant-vowel structure (e.g. double consonants and vowel length), spelling (e.g. pronunciation of ä as the letter e) and corpus frequency (Borin et al., 2012). The parallel versions of each task were used for pre-, post- and follow-up assessment. Table A3 presents the contents of the tasks and Table A4 presents examples of the lists and how they were matched to the trained words.

Trained word identification In the Measure Game, 30 trained words (with both root word and suffix) were presented as a word identification task with three distractors, all in written form. The distractors were constructed of the same root word with other suffixes (i.e. other trained suffixes or not trained forms as genitive -s), in a few cases (8 out of 558) we had to construct word forms that are not used in the Swedish language (pseudowords). The instruction for the task was *Välj rätt ord. Var snabb och noggrann!* (Choose the right word. Be quick and careful.) The game then orally presented the target word for the student to look for among four alternatives. The test score was the number of correctly selected target words. Reliabilities based on Cronbach's alpha were 0.90, 0.87, and 0.87 at pre-, post- and follow-up assessments, respectively.

Transfer word identification This was a word identification task in the Measure Game with 30 words. The words were combined from trained suffixes and untrained root words with a relative corpus frequency of above 5 tokens per one million words. The test score was the number of correctly selected target words. Reliabilities based on Cronbach's alpha were 0.89, 0.91, and 0.89 at pre-, post- and follow-up assessments, respectively.

Transfer word reading A word list of 30 words, built from trained suffixes combined with untrained root words with a similar corpus frequency as trained root words, was read aloud. Scores for the number of correctly read words (accuracy) and the total time for reading were recorded. Reliabilities for the accuracy measure based on

Cronbach's alpha were 0.71, 0.74, and 0.71 at pre-, post- and follow-up assessments, respectively.

Transfer pseudoword reading A list comprising 30 pseudowords constructed of pseudoroot words (resembling authentic Swedish words) combined with trained suffixes was read aloud. The number of correctly read pseudowords (accuracy) and the total time for reading formed the scores. Reliabilities for the accuracy measure based on Cronbach's alpha were 0.78, 0.72, and 0.78 at pre-, post- and follow-up assessments, respectively.

Intervention programme

The training was conducted with the study-specific game for eight weeks with an additional one-week pause due to school holidays. Teachers or school assistants supervised 4–5 training game sessions weekly, automatically limited to 10 min of active training time. The training was conducted individually on tablet computers equipped with headphones during school hours in small groups or individually. The intervention timetable is described in Table A1.

Group E received explicit instruction via in-game animations. The trained suffixes were exemplified in short, funny stories, demonstrated by an animated wizard. For instance, the plural ending -*or* was introduced by the wizard who was feeling hungry and therefore bewitching a plural ending to a noun in singular form (*kaka – kakor*, cake – cakes), where he could make numerous cakes out of only one. In the animation a written root word was combined with the suffix (kaka + or = kakor) simultaneously as the cakes multiplied in the animation. This stressed the meaning of the suffix addition. The animations were followed by training tasks.

Group I completed exactly the same training tasks with implicit guidance, that is, without the animations. In task instructions the word suffix was used for Group E but left out or paraphrased for Group I, e.g. Group E was instructed: *Nu är det din tur att bygga ändelsen.* (It's your turn to build the suffix.) while Group I followed the instruction: *Nu är det din tur att bygga.* (It's your turn to build.). Another for Group E was: *Hittar du samma ord med olika ändelser?* (Can you find the same word with different suffixes?) and for Group I: *Hittar du liknande ordpar?* (Can you find the similar pairs of words?).

The third group (Group C) functioned as controls, receiving school-provided support as ordinary.

The training game design

The in-game training focused on morphology as the repetition of orthographic units (multilettered suffixes). Content was retrieved from the LäSBarT corpus at Språkbanken (*spraakbanken.gu.se*; Borin et al., 2012), constructed of Swedish easy-to-read texts and children's literature. We selected 16 frequent suffixes combined with the most frequent root words that also were considered to be age-group relevant accord-

ing to teachers. By utilising frequent suffixes and root words, predictability can be increased and context will be easier to harness (Martin-Chang et al., 2007; Martin-Chang & Levy, 2006). Table A5 lists suffixes and examples of root words.

In the game, students created an alias that moved through a magic school and its surroundings, while encountering training tasks and receiving rewards. The game comprised 16 streams, each including 3–10 task types (with several exercises of a certain kind). For each of the 16 streams, a suffix was added to the training. For Group E, each stream started with an animated introduction of the new suffix which was followed by a suffix identification task. The animation demonstrated by the wizard is described above. In the task, the students were shown how the letters in the suffix formed the suffix, after which the players were asked to identify this suffix among other suffixes functioning as distractors. The suffix term was used in the instruction in Group E. For Group I no animation was presented but the formation of the new suffix was introduced to the players similarly letter by letter as for Group E. However, the term suffix was not used and in the task the players were instructed only to choose the correct answer among different alternatives. The training within each stream focused on the new suffixes as well as suffixes presented in previous streams. The streams comprised several task types, each containing multiple similar exercises. The player selected a task type out of a pool of three alternatives. The task type was presented repeatedly in the pool until the student correctly completed 75% of the exercises within the task types correctly. All task types within one stream had to be completed before entering the next stream.

There were 12 different forms of task types (see Table A6). One task type included similar exercises with 3 to 16 single words (root word+suffix), depending on task type. In all task types, the student heard a verbal stimulus, i.e. the trained word or a sentence, into which a word should be added. The exercise was solved by clicking on, for example, the correct word form out of four alternatives or by combining two words or word parts. After a correct response, the game gave the player a specific sound feedback and turned the colour of the written item green. The response to false answers was a different sound and a visual red mark on the falsely chosen alternative. After this, the same exercise was repeated to offer the student a chance to correct his/her answer. Figures 2, 3, 4, 5, 6 and 7 show images of some task types and the instructions for them.

Fidelity

The intervention fidelity was ensured mainly through teacher training and support. Ahead of each testing session, the teachers participated in group video meetings where the test administration was instructed and modelled. The teachers were provided with printed manuals with word-to-word instructions for students and test materials. All individual test sessions were recorded. Throughout the intervention period, teachers had access to further support by telephone or e-mail and a weekly video meeting.

Student´s training times were automatically logged by the game. In order to enhance motivation, each student documented the playing sessions in a personal Player Card with a visual overview of the training period.
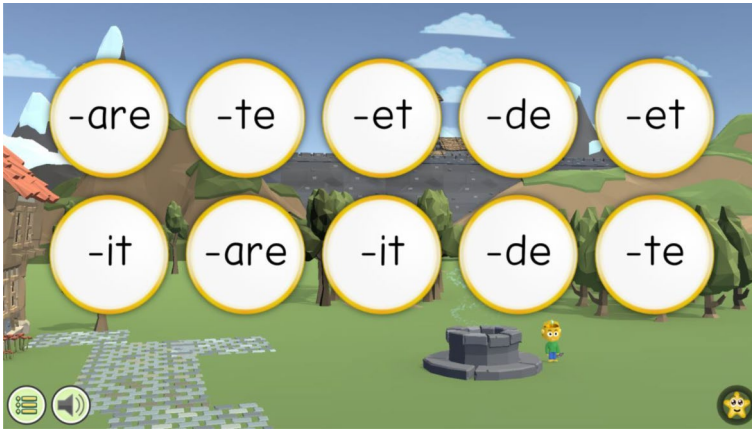
Fig. 2 Screenshot of the task type identifying the suffix.
Note. Task instruction: Explicit Group: *Välj rätt ändelser* [Choose the correct suffixes] Implicit group: *Välj rätt ändelser* [Choose the correct alternatives]. After this instruction the game orally presents the suffix that the player needs to find. All alternatives are existing Swedish suffixes
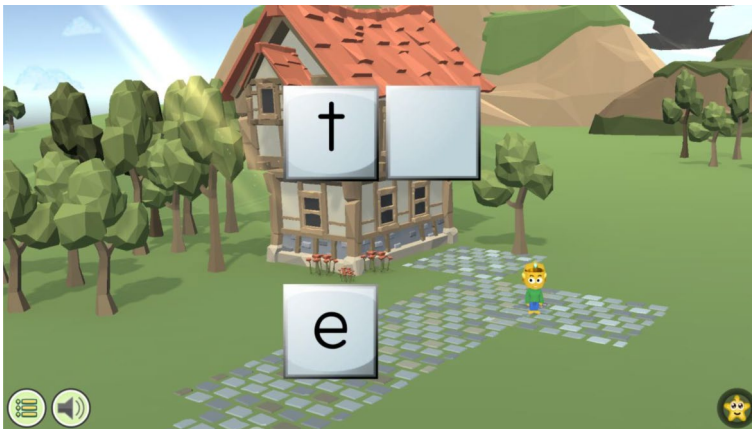
Fig. 3 Screenshot of the task type identifying the suffix with task instruction below.
Note. Task instruction: Explicit Group: *Nu är det din tur att bygga ändelsen* [It's your turn to build the suffix] Implicit group: *Nu är det din tur att bygga*. [It's your turn to build]

Due to the COVID-19 pandemic situation, time points for the pre- and post-tests varied for a few students. One student completed part of the testing through video contact with her teacher. Four students also had the option to carry out training during quarantine periods of approximately one week each.

Each playing session was limited to a total of 10 min, this limitation was programmed in the game. The teachers were instructed to guide the students to train at least 4 sessions a week for 8 weeks (altogether 36 sessions). Hence, after fully completing the training period, the students would have played for at least 320 min (5 h 20 min). The actual playing times for the participants varied from 158.4 min

Fig. 4 Screenshot of the main task type finding a false word with task instruction below.
Note. Task instruction: *Klicka på ordet som inte passar in* [Mark the word that doesn't fit the sentence]

Fig. 5 Screenshot of the main task type recognizing a suffix in the word with task instruction below.
Note. Task instruction: *Hur slutar ordet?* [How does the word end?]. The player then sees a word in the flashlight beam, after this several suffixes are shown and the player needs to choose the one he/she saw in the word in the flashlight beam. All alternatives are existing Swedish suffixes

(2 h 39 min) to 339.6 min (5 h 40 min). The average playing time for Group E was 238.47 min ($SD=39.48$ min) and 244.01 ($SD=37.79$ min) for Group I. There was no difference in the amount of training, measured as the number of trained items, between the two experimental groups (see Table 1).

Statistical analyses

A mixed-design univariate analysis of variance (ANOVA), separately for each reading task, was conducted with IBM SPSS Statistics 28 (IBM, Armonk, NY, USA) to compare reading development in the experimental groups and the controls during the

Fig. 6 Screenshot of the Main Task Type Word identification (selecting three correct words).
Note. Task instruction: *Välj tre ord som passar* [Choose three words that fit]

Fig. 7 Screenshot of the Repetitive Task Type Combining root words and their inflected forms.
Note. Task instruction: Explicit Group: *Hittar du samma ord med olika ändelser?* [Can you find the same word with different suffixes?] Implicit group: *Hittar du liknande ordpar?* [Can you find the similar pairs of words?].

intervention. In the analysis, training condition (Implicit, Explicit and Control) and Grade (2 and 3) were used as between-subject factors and time (pre- and post-test) as a within-subject factor. Separate analysis was performed for each of the reading scores because they were supposed to measure the near and far transfer of the training effects. A univariate analysis of variance was conducted to analyse reading outcomes in the follow-up assessment. In these analyses, the initial level of reading and grade were controlled. Effect sizes (Cohen's *d*) within each group were calculated to describe the magnitude of change in reading measures separately from pre- to post-test and from pre-test to follow-up. Group differences in effect size (Cohen's *d*) were calculated pairwise for all three pairs (E vs. C, I vs. C, and E vs. I).

Results

In Table 1, we present descriptive characteristics of the participants and the scores at screening for the three groups. In Table 2, we present the means, standard deviations, effect sizes (Cohen's *d*) and missing data for reading measures for the groups at all measure points. The missing data per measure on a specific time-point ranges from 0 to 3 (3,2%) of the participants per group. The correlations between the measures at each time-point are presented in Tables A7–A9.

The distributions were mostly normal, except for the time measures (Transfer Word and Transfer Pseudoword Time) that had a high kurtosis (Transfer Word, $K=7.94$–12.59) (Transfer Pseudoword Time, $K=9.62$–16.94), which is to be expected in this kind of selected sample of poor readers.

Intervention effects (pre- and post-test development)

A significant Time x Group and a Time x Grade effect were found in Trained Word Identification (see Table 3). A further analysis separately by group showed that in Group I, there was a significant increase in Trained Word Identification ($F(1, 90)=0.16.141$, $p=<0.001$, $\eta_p^2 =0.152$), whereas in Group C, there was not ($F(1, 90)=0.283$, $p=.596$, $\eta_p^2 = 0.003$). The effect size (Cohen's *d*) between pre- and post-assessment was moderate in Group I and negligible in Group C (see Table 2). The difference in effect sizes (Cohen's *d*) between Group I and Group C was small. Effect sizes were interpreted according to Cohen's criteria (Cohen, 1988): 0.010–0.059 for small, 0.060–0.139 for moderate and over 0.140 for large effect.

Similarly, in Transfer Pseudoword Reading Accuracy, a significant Time x Grade x Group interaction was found. Further analysis separately by Grade showed that the Time x Group interaction was significant only in Grade 3 ($F(1,77)=5.728$, $p=.019$, np2$=0.069$). In Grade 3, Group I improved their reading accuracy of pseudowords significantly ($F(1,38)=7.195$, $p=.011$, np2$=0.159$), whereas no significant change was found in Group C. The effect size (Cohen's *d*) between pre- and post-assessment was small in Grade 3 in Group I (see Table 2). However, the difference in effect sizes (Cohen's *d*) between Group I and Group C was moderate due to the increase in reading accuracy in Group I in contrast to no change in reading accuracy in Group C.

In the other measures, no significant Time x Group effect was found. The effect sizes (Cohen's *d*) reflecting either changes in the skill within each group or differences in the effect sizes between the groups were small or negligible.

No differences in development were found in comparisons of Group E to Group C and to Group I, either in the identification tasks or tasks regarding reading accuracy or time.

Reading outcomes of the groups in the follow-up assessment

We used ANOVA to compare the reading outcomes in the groups at the follow-up assessment. Analyses were performed separately for each reading measure. Reading level and grade were used as covariates in all analyses. A significant main effect of group, due to a higher value in either of the experimental groups compared with

controls, was considered a sign of sustained effect of the intervention. The results for the univariate analyses are presented in Table 4.

Pre-test to follow-up comparisons confirmed the earlier findings regarding the intervention effects in the pre- to post-test analyses, i.e. at follow-up, a group effect was found in the same tasks in which the intervention resulted in gains in Group I at the post-test. In Trained Word Identification, the main effect of Group approached significance ($p=.051$). According to further analyses separately by group, Trained Word Identification accuracy increased in Group I ($t = -4.705$, $p<.001$, one-tailed), whereas a decreasing trend was found in Group C ($t = -1.547$, $p=.063$, one-tailed). The effect size from pre-test to follow-up was moderate for Group I and negligible for Group C (see Table 2). The difference in effect sizes from post-test to follow-up between Group I and Group C was small.

In Transfer Pseudoword Reading Accuracy, a significant Grade x Group interaction effect was found. A further analysis separately by Grade showed a significant effect of Group only in Grade 3 ($F(1, 73)=4.817$, $p=.031$, $\eta_p^2 = 0.062$). Further analyses separately by group within third graders showed an increasing trend in Pseudoword Reading Accuracy in Group I ($t = -1.658$, $p=.053$, one-tailed) and a decrease in Group C ($t=2.305$, $p = .014$, one-tailed).

The effect size (Cohen's d) between the pre-test and follow-up for Group I in Grade 3 was small (see Table 2). For Group C, there was a small but negative difference in effect size (Cohen's d) from pre-test to follow-up. However, there was a moderate difference in the effect sizes (Cohen's d) for Group I and Group C between the pre-test and follow-up.

No further significant differences between Group I and Group C at follow-up were found. Similarly, significant differences were not found either between Group E and Group C or between Group E and Group I.

Discussion

The goal of this study was to explore the effects of computer-assisted MT with poor readers in Grades 2 and 3, both directly after the intervention and after a six-week follow-up period. Reading training with a focus on morphology with explicit instruction (Group E) or without instruction (Group I) was provided for two respective experimental groups and compared with business-as-usual controls, additionally comparing the experimental groups with each other. Grade as a proxy for reading level was explored as a possible moderator for the response to the intervention. The training methods aimed at improving morphological decoding of suffixes, with the explicit version emphasizing the specific meanings of morphological units, while the implicit version relied on reading experience. The results showed that poor readers benefited from an intensive intervention with MT. However, the improvement in reading skills was dependent on how the training was implemented in the game (with explicit instructions or as implicit training) and the students' grade level. All students within Group I gained in identification of trained words and these effects prevailed to follow-up. Further, grade was found to be a moderator for transfer to pseudoword reading accuracy, where gains were found only among third graders within Group

Table 2 Descriptives and effect sizes for reading measures

Measure	Grade	Time point	Control group n=93				Implicit game group n=92				Explicit game group n=92				Group differences in effect size between the groups [a]		
			n	M	SD	d	n	M	SD	d	n	M	SD	d	I–C	E–C	I–E
Trained Word Identification	2 & 3	Pre-test	92	24.67	5.10		92	23.67	5.44		92	24.14	5.08				
		Post-test	92	24.91	4.17	−0.05	91	25.75	3.84	−0.44	92	25.12	5.39	−0.19	−0.39	−0.14	−0.25
		Follow-up [b]	92	25.49	4.85	−0.16	91	26.21	3.80	−0.54	90	25.98	4.63	−0.38	−0.38	−0.21	−0.16
	2	Pre-test	53	23.66	6.26		53	22.30	5.98		52	22.54	5.18				
		Post-test	52	24.44	4.95	−0.14	53	25.28	4.17	−0.58	52	24.27	6.40	−0.30	−0.44	−0.16	−0.25
		Follow-up	53	25.42	4.90	−0.31	52	25.75	4.24	−0.66	51	25.63	4.92	−0.61	−0.35	−0.30	−0.05
	3	Pre-test	39	26.05	2.31		39	25.54	3.95		40	26.23	4.16				
		Post-test	40	25.53	2.82	0.20	38	26.39	3.28	−0.24	40	26.23	3.48	0.00	−0.44	−0.20	−0.25
		Follow-up	39	25.59	4.85	0.12	39	26.82	3.07	−0.36	39	26.44	4.24	−0.05	−0.48	−0.17	−0.31
Transfer Word Identification	2 & 3	Pre-test	91	23.76	5.19		91	23.45	5.29		92	23.62	5.23				
		Post-test	92	24.04	5.22	−0.05	91	24.41	5.36	−0.18	92	23.73	6.26	−0.02	−0.12	0.04	−0.25
		Follow-up	92	23.95	5.51	−0.04	91	25.13	5.42	−0.31	90	25.19	5.19	−0.30	−0.28	−0.27	−0.01
	2	Pre-test	53	22.58	5.80		52	21.98	5.63		52	22.29	5.44				
		Post-test	52	23.69	5.28	0.20	53	23.96	5.75	0.35	52	22.98	6.96	−0.11	0.15	−0.31	−0.25
		Follow-up	53	23.92	5.94	−0.23	52	24.79	5.65	−0.50	51	25.10	4.85	−0.55	−0.27	−0.32	0.05
	3	Pre-test	38	25.39	3.67		39	25.41	4.11		40	25.35	4.45				
		Post-test	40	24.50	5.18	0.20	38	25.03	4.77	0.09	40	24.70	5.13	0.14	−0.11	−0.06	−0.25
		Follow-up	39	23.97	4.95	0.33	39	25.59	5.12	−0.04	39	25.31	5.67	0.01	−0.36	−0.32	−0.05
Transfer Word Reading Accuracy	2 & 3	Pre-test	93	19.77	4.07		92	18.54	6.08		92	20.12	3.86				
		Post-test	93	22.66	4.30	−0.69	92	21.61	5.89	−0.51	92	22.54	4.80	−0.56	0.18	0.13	0.04
		Follow-up	90	23.64	4.17	−0.94	91	22.37	5.75	−0.65	92	24.09	3.28	−1.11	0.29	−0.17	0.46
	2	Pre-test	53	19.26	3.97		53	17.19	6.07		52	19.35	3.97				
		Post-test	53	22.38	4.57	−0.73	53	20.70	6.12	−0.58	52	21.63	5.52	−0.48	−0.15	−0.25	−0.10
		Follow-up	53	23.25	4.61	−0.93	52	21.35	5.82	−0.70	52	23.44	3.47	−1.10	0.23	−0.17	0.40
	3	Pre-test	40	20.45	4.16		39	20.38	5.65		40	21.13	3.52				
		Post-test	40	23.03	3.95	−0.64	39	22.85	5.41	−0.45	40	23.73	3.39	−0.75	0.19	−0.12	0.31
		Follow-up	37	24.22	3.44	−0.99	39	23.74	5.42	−0.61	40	24.93	2.84	−1.19	0.38	−0.20	0.58

Table 2 (continued)

Measure	Grade	Time point	Control group n=93				Implicit game group n=92				Explicit game group n=92				Group differences in effect size between the groups [a]	
Transfer Word Reading Time	2 & 3	Pre-test	93	114.38	51.29		92	111.10	49.78		91	112.92	61.25		0.05	−0.09
		Post-test	93	93.59	39.48	0.45	92	89.16	36.60	0.50	92	92.74	48.79	0.36	0.13	0.06
		Follow-up	90	87.34	43.27	0.57	91	80.35	37.75	0.70	91	80.23	39.63	0.63		
	2	Pre-test	53	127.40	54.42		53	124.49	52.79		51	126.62	73.14		0.11	−0.13
		Post-test	53	104.15	39.32	0.49	53	96.98	38.64	0.59	52	103.44	55.00	0.36	0.17	0.02
		Follow-up	53	97.68	43.47	0.60	52	88.37	39.71	0.77	51	88.61	45.12	0.63		
	3	Pre-test	40	97.13	41.47		39	92.90	39.11		40	95.47	35.22		−0.05	0.02
		Post-test	40	79.60	35.53	0.45	39	78.54	31.06	0.41	40	78.83	35.34	0.47	0.04	0.20
		Follow-up	37	72.54	38.95	0.61	39	69.67	32.46	0.65	40	69.55	28.37	0.81		
Transfer Pseudoword Reading Accuracy	2 & 3	Pre-test	93	20.32	5.16		92	19.09	7.27		92	21.51	5.06		0.03	0.11
		Post-test	93	21.57	4.82	−0.25	92	20.6	6.23	−0.22	91	22.21	4.68	−0.14	−0.13	0.05
		Follow-up	90	20.54	5.15	−0.04	91	20.26	6.34	−0.17	92	21.50	4.49	0.00		
	2	Pre-test	53	19.23	5.49		53	18.66	7.94		52	20.06	5.56		0.40	0.35
		Post-test	53	22.09	4.57	−0.57	53	19.89	6.91	−0.16	51	21.25	5.54	−0.22	0.14	0.11
		Follow-up	53	20.79	5.01	−0.30	52	19.79	6.50	−0.16	52	21.02	4.63	−0.19		
	3	Pre-test	40	21.78	4.33		39	19.67	6.30		40	23.40	3.59		−0.52	−0.20
		Post-test	40	20.88	5.10	0.19	39	21.56	5.08	−0.33	40	23.43	2.91	−0.01	−0.52	0.00
		Follow-up	37	20.19	5.38	0.32	39	20.90	6.15	−0.20	40	22.13	4.29	0.32		

Table 2 (continued)

Measure	Grade	Time point	Control group n=93				Implicit game group n=92				Explicit game group n=92				Group differences in effect size between the groups [a]		
Transfer Pseudo-word Reading Time	2 & 3	Pre-test	93	108.27	42.25		92	108.46	50.38		92	112.66	60.49				
		Post-test	92	98.22	45.66	0.23	92	92.13	34.57	0.38	91	92.05	33.27	0.42	0.15	0.19	-0.04
		Follow-up	90	95.04	44.79	0.30	91	88.47	36.84	0.45	92	89.27	41.64	0.45	0.15	0.15	0.00
	2	Pre-test	53	118.72	43.55		53	96.28	39.62		52	123.27	70.08				
		Post-test	53	110.32	51.30	0.18	53	82.85	29.00	0.39	51	98.39	34.05	0.45	0.21	0.28	-0.06
		Follow-up	53	106.96	49.59	0.25	52	95.17	39.90	0.03	52	96.38	47.49	0.45	-0.22	0.20	-0.42
	3	Pre-test	40	94.42	36.58		39	117.42	55.69		40	98.87	42.09				
		Post-test	39	81.77	30.16	0.38	39	98.96	36.94	0.39	40	83.97	30.78	0.40	0.01	0.03	-0.01
		Follow-up	37	77.97	29.98	0.49	39	79.54	30.57	0.84	40	80.03	30.70	0.51	0.35	0.02	0.33
General Reading Skill [c]	2 & 3	Pre-test	92	0.01	0.94		91	-0.00	0.92		92	-0.01	0.88				
		Post-test	92	0.66	1.03	-0.66	91	0.64	1.11	-0.63	92	0.71	1.02	-0,76	0,03	-0,1	0,13
		Follow-up	91	1.02	1.11	-0.98	91	1.06	1.15	-1.02	91	1.03	1.04	-1,08	-0,04	-0,1	0,06
	2	Pre-test	52	-0.50	0.65		52	-0.53	0.63		52	-0.52	0.59				
		Post-test	52	0.21	0.85	-0.94	52	0.15	0.90	-0.88	52	0.27	0.85	-1.08	0,06	-0,14	0,2
		Follow-up	53	0.55	0.89	-1.35	52	0.56	0.94	-1.36	52	0.62	0.83	-1.58	-0,01	-0,23	0,22
	3	Pre-test	40	0.68	0.84		39	0.70	0.77		40	0.67	0.72				
		Post-test	40	1.24	0.96	-0.62	39	1.28	1.05	-0.63	40	1.29	0.95	-0.74	-0,01	-0,12	0,11
		Follow-up	38	1.67	1.06	-1.04	39	1.72	1.07	-1.09	39	1.58	1.05	-1.01	-0,05	0,03	-0,08

Note. Effect size (Cohen's d) was measured as standardised mean differences between pre-test and post-test or pre-test and follow-up

[a] Calculated by computing the difference between the effect sizes in groups at post-test and follow–up. The sign in front of Cohen's d indicates the direction of the difference in the effect size. I=Implicit Group. C=Control Group. E=Explicit Group

[b] Follow-up effect sizes are based on pre-test to follow-up measures

[c] The values for the composite score General Reading Skill are based on a mean of the z-scores for the three measures: Word reading, Text Reading and Sentence Verification

Table 3 F-values related to mixed-design anovas evaluating possible intervention effects between pre- and post-assessment

Compared groups	Measure	Main effect			Interaction effect		
		Time	Group	Grade	Time*Group	Time*Grade	Time*Grade*Group
Implicit–Control	Trained Word Identification	$F(1, 178)=8.78**$	$F(1, 178)=0.01$	$F(1, 178)=11.17$	$F(1, 178)=6.22*$	$F(1, 178)=6.17*$	$F(1, 178)=0.66$
	Transfer Word Identification	$F(1, 176)=1.32$	$F(1, 176)=0.00$	$F(1, 176)=9.77**$	$F(1, 176)=0.55$	$F(1, 176)=6.86**$	$F(1, 176)=0.20$
	Transfer Word Reading Accuracy	$F(1, 181)=61.03***$	$F(1, 181)=2.31$	$F(1, 181)=7.44**$	$F(1, 181)=0.04$	$F(1, 181)=1.13$	$F(1, 181)=0.12$
	Transfer Word Reading Time	$F(1, 181)=123.12***$	$F(1, 181)=0.40$	$F(1, 181)=18.42***$	$F(1, 181)=0.02$	$F(1, 181)=6.42*$	$F(1, 181)=1.00$
	Transfer Pseudoword Reading Accuracy	$F(1, 181)=11.30***$	$F(1, 181)=1.75$	$F(1, 181)=1.60$	$F(1, 181)=0.58$	$F(1, 181)=4.18**$	$F(1, 181)=8.59**$
	Transfer Pseudoword Reading Time	$F(1, 180)=48.42***$	$F(1, 180)=0.14$	$F(1, 180)=14.15***$	$F(1, 180)=2.34$	$F(1, 180)=0.04$	$F(1,180)=1.32$
	General Reading Skill	$F(1, 179)=331.60***$	$F(1, 179)=0.00$	$F(1, 179)=91.27***$	$F(1, 179)=0.47$	$F(1, 179)=0.97$	$F(1, 179)=0.14$
Explicit–Control	Trained Word Identification	$F(1, 179)=1.89$	$F(1, 179)=0.06$	$F(1, 179)=13.79***$	$F(1, 179)=0.84$	$F(1, 179)=3.71$	$F(1, 179)=0.16$
	Transfer Word Identification	$F(1, 178)=0.045$	$F(1, 178)=0.17$	$F(1, 178)=9.84**$	$F(1, 178)=0.03$	$F(1, 178)=3.17$	$F(1, 179)=0.06$
	Transfer Word Reading Accuracy	$F(1, 181)=54.21***$	$F(1, 181)=0.12$	$F(1, 180)=7.67**$	$F(1, 181)=0.31$	$F(1, 181)=0.03$	$F(1, 181)=0.35$
	Transfer Word Reading Time	$F(1, 180)=68.02***$	$F(1, 180)=0.02$	$F(1, 180)=16.21***$	$F(1, 180)=0.02$	$F(1, 180)=1.49$	$F(1, 189)=0.00$
	Transfer Pseudoword Reading Accuracy	$F(1, 180)=3.74$	$F(1, 180)=3.41$	$F(1, 180)=7.57**$	$F(1, 180)=0.44$	$F(1, 180)=9.53**$	$F(1, 180)=3.53$
	Transfer Pseudoword Reading Time	$F(1, 179)=32.69***$	$F(1, 179)=0.05$	$F(1, 179)=13.37***$	$F(1, 179)=2.15$	$F(1, 179)=0.01$	$F(1, 179)=0.74$
	General Reading Skill	$F(1,)=331.60***$	$F(1, 180)=0.026$	$F(1, 180)=92.52***$	$F(1, 180)=0.00$	$F(1, 180)=2.79$	$F(1, 180)=0.11$

Table 3 (continued)

Compared groups	Measure	Main effect				Interaction effect	
Explicit – Implicit	Trained Word Identification	F(1,179)=12.33***	F(1, 179)=0.02	F(1, 179)=17.25***	F(1, 179)=1.64	F(1, 179)=6.57	F(1, 179)=0.11
	Transfer Word Identification	F(1,178)=0.82	F(1, 178)=0.18	F(1, 178)=11.38***	F(1, 178)=0.73	F(1, 178)=4.85	F(1, 178)=0.43
	Transfer Word Reading Accuracy	F(1, 180)=46.76 ***	F(1, 180)=3.26	F(1, 180)=12.45***	F(1, 180)=0.47	F(1, 180)=0.22	F(1, 180)=0.73
	Transfer Word Reading Time	F(1,179)=75.95***	F(1, 179)=0.19	F(1, 179)=15.14***	F(1, 179)=0.07	F(1, 179)=4.28*	F(1, 179)=0.57
	Transfer Pseudoword Reading Accuracy	F(1, 179)=7.90 **	F(1, 179)=7.54**	F(1, 179)=6.40*	F(1, 179)=2.20	F(1, 179)=0.03	F(1, 179)=1.19
	Transfer Pseudoword Reading Time	F(1, 179)=52.94***	F(1, 179)=0.02	F(1, 179)=8.67**	F(1, 179)=0.08	F(1, 179)=1.12	F(1, 179)=0.00
	General Reading Skill	F(1, 179)=351.18***	F(1, 179)=0.04	F(1, 179)=95.57***	F(1, 179)=1.10	F(1, 179)=0.35	F(1, 179)=0.35

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

Table 4 F-values related to univariate analysis of variance evaluating possible intervention effects between pre- and follow up-assessment

Compared groups	Measure	Main effect Start level	Group	Grade	Interaction effect Grade*Group
Implicit – Control	Trained Word Identification	$F(1, 177)=50.19$ ***	$F(1, 177)=3.86$	$F(1, 177)=0.69$	$F(1, 177)=0.33$
	Transfer Word Identification	$F(1, 175)=54.87$***	$F(1, 175)=3.36$	$F(1, 175)=2.45$	$F(1, 175)=0.13$
	Transfer Word Reading Accuracy	$F(1, 176)=85.19$***	$F(1, 176)=0.49$	$F(1, 176)=0.22$	$F(1, 176)=0.26$
	Transfer Word Reading Time	$F(1, 176)=343.76$***	$F(1, 176)=1.16$	$F(1, 176)=0.20$	$F(1, 176)=1.47$
	Transfer Pseudoword Reading Accuracy	$F(1, 176)=112.94$***	$F(1, 176)=1.23$	$F(1, 176)=1.75$	$F(1, 176)=3.99*$
	Transfer Pseudoword Reading Time	$F(1, 176)=330.97$***	$F(1, 176)=2.26$	$F(1, 176)=3.07$	$F(1, 176)=2.91$
	General Reading Skill	$F(1, 177)=544.71$***	$F(1, 177)=0.12$	$F(1, 177)=7.95$**	$F(1, 177)=0.00$
Explicit – Control	Trained Word Identification	$F(1, 176)=57.93$***	$F(1, 176)=1.59$	$F(1, 176)=2.23$	$F(1, 176)=0.01$
	Transfer Word Identification	$F(1, 175)=31.16$***	$F(1, 175)=3.28$	$F(1, 175)=1.94$	$F(1, 175)=0.00$
	Transfer Word Reading Accuracy	$F(1, 177)=24.41$***	$F(1, 177)=0.63$	$F(1, 177)=1.36$	$F(1, 177)=0.64$
	Transfer Word Reading Time	$F(1, 176) = 231.03$***	$F(1, 176)=1.65$	$F(1, 176)=1.69$	$F(1, 176)=0.78$
	Transfer Pseudoword Reading Accuracy	$F(1, 177)=47.57$***	$F(1, 177)=0.083$	$F(1, 177)=3.02$	$F(1, 177)=1.34$
	Transfer Pseudoword Reading Time	$F(1, 177)=214.00$***	$F(1, 177)=2.56$	$F(1, 177)=3.48$	$F(1, 177)=2.38$
	General Reading Skill	$F(1, 177)=482.89$***	$F(1, 177)=0.07$	$F(1, 177)=11.07$**	$F(1, 177)=1.18$
Implicit - Explicit	Trained Word Identification	$F(1, 176)=50.98$***	$F(1, 176)=0.50$	$F(1, 176)=0.50$	$F(1,176)=0.18$
	Transfer Word Identification	$F(1, 175)=29.98$***	$F(1, 175)=0.00$	$F(1, 175)=0.96$	$F(1,175)=0.13$
	Transfer Word Reading Accuracy	$F(1, 178)=90.31$***	$F(1, 178)=2.41$	$F(1, 178)=1.16$	$F(1, 178)=0.03$
	Transfer Word Reading Time	$F(1, 177)=220.60$***	$F(1, 177)=0.07$	$F(1, 177)=0.35$	$F(1, 177)=0.02$
	Transfer Pseudoword Reading Accuracy	$F(1, 178)=111.33$***	$F(1, 178)=0.11$	$F(1, 178)=0.04$	$F(1, 178)=0.90$
	Transfer Pseudoword Reading Time	$F(1, 178)=273.40$***	$F(1, 178)=0.13$	$F(1, 178)=0.37$	$F(1, 178)=0.92$
	General Reading Skill	$F(1, 177)=434.22$***	$F(1, 177)=0.00$	$F(1, 177)=12.70$***	$F(1, 177)=1.16$

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

I, and these gains also remained to follow-up. No effects were found in either group with the standardised measures of generalised reading skills.

Our first research question concerned the differences in gains in reading skills in Group I and Group E when compared pairwise to the control group and to one another. Our hypothesis of seeing larger gains for Group E was not verified. Students in Group E made progress in all measures, but this progress did not differ significantly when compared with Group C. Greater gains compared with controls, regardless of grade, were found only for Group I on the task of identifying trained words. The dissimilarity of training effects between the two experimental groups in relation to Group C cannot be explained by a difference in the amount of training between the two experimental groups. Neither can it be explained by a difference in their reading level at the beginning of the training (see Table 1). Hence, explicit training in the form that we used did not seem to be effective.

Our hypothesis of larger gains for Group E was not fulfilled. It is possible that the animated instructions were not sufficient for supporting an explicit understanding of the relationships between morphological units and reading. Tijms and Hoeks (2005) found the LEXY programme to be successful for improving reading among Dutch children with dyslexia. In LEXY, explicit MT was mainly graphic, rather than verbal, in order to avoid verbal information processing, which can be hard for dyslexics. However, participants were required to articulate their actions in order to enhance the internalisation of the training. The instruction in the present study was designed by experts in the fields of special education and speech therapy, but we could not control for whether the students understood the central concepts presented during the practice. It is also worthwhile to note that concepts related to morphology do not play a part in typical reading instruction. A more effective form of explicit instruction could be the direct and personal instruction by a teacher, which also can activate and involve students to a higher degree, contrary to our animations. It would also allow the teacher to monitor student comprehension and to clarify instructions when needed (McTigue et al., 2020).

The gains in identification of trained words in Group I can be compared with the results of the study by Good et al. (2015) in which third grade students showed similar gains in reading both with explicit MT and implicit training. They concluded that repeated exposure to stimuli, even without explicit instruction, may have increased the students' word recognition skills. Although this repetition was similar in our training in both experimental conditions, only the implicit group gained in comparison to controls. Thus, it is even possible that the animations were a bit detrimental. The animations may have distracted the students by drawing their attention to irrelevant details. Another possibility is that the students perceived the animations as a form of entertainment and a break from their training.

Further, we studied whether the intervention effects would differ according to grade level in order to find out for which grade our training was most suitable. Grade did not alter the training effect in the identification of trained words, and gains were found among both second and third graders. Hence, third graders' better skill level in reading or the longer reading instruction they had received did not seem to affect their ability to benefit from the training. However, grade level did have a moderating effect in the Transfer Pseudoword Reading Accuracy task, as significant gains were found

only among third graders in the implicit intervention as compared with controls. It could be that due to the weaker decoding skills of the second graders, they might lack the capacity to identify the trained suffixes in the pseudowords.

In general, our results are in line with the conclusion of Rastle et al. (2019) that MT suits students who have passed the initial stages of reading. Likewise, Tijms et al. (2020) found MT to be effective for 11–12-year-old students. At that stage, readers have gathered knowledge of more complex morphological structures (Reed, 2008). In our study, the third graders were able to utilise the morphological features learned in the training in other contexts, as the findings in the Transfer Pseudoword Reading task show. In contrast, second graders in the implicit condition benefitted only in the Trained Word Identification task, suggesting a more limited training effect compared with that of third graders. The third graders' reading-level (Carlisle et al., 2010; Tijms et al., 2020; Reed, 2008) and longer experience of reading instruction can both be prerequisites contributing to increased receptiveness to the MT overall and to this training in particular. On the other hand, in the study by Torkildsen et al. (2021), in a relatively similar language as Swedish, namely Norwegian, transfer to other forms of reading was already found among second graders. However, these second graders comprised readers of all levels, not only poor readers, as was the case in our study. Therefore, it is likely that the reading level of our Grade 2 poor readers was lower and the reading level of our Grade 3 poor readers was more similar to the second graders' average reading level in Torkildsen et al.'s study. In line with this, Eklund et al. (2015) found that Finnish-speaking typical readers in Grade 2 outperformed poor readers in Grade 3.

Grade-level results were also observed as a transfer from identification of words (a silent reading task used in the training) to reading aloud from print in the Transfer Pseudoword Reading measure. Nevertheless, no transfer effect was found for the Transfer Word Reading measure, which also required reading aloud from print. An explanation for the transfer could be that third graders have encountered more complex words than second graders, which enhanced their ability to generalise the MK (Reed, 2008). The pseudoword task may also be a better measure for transfer than reading real words, as readers cannot have previously encountered the pseudowords. This result differs from the study in Norwegian, where transfer effects to unexposed, real words with trained affixes were noted (Torkildsen et al., 2021). In our study, transfer was observed to pseudowords as an item type, rather than to the reading of real words. One possible explanation could be that relatively frequent real words, along with high-frequency suffixes, have already been lexicalized by the third graders. The practice during the intervention may have supported especially the phonological assembly skills required especially in reading pseudowords. Regarding reading aloud from print, we noted effects in the accuracy measures rather than the reading time measures. Similarly, Tijms and Hoeks (2005) noted the highest gains on a measure of accuracy in text reading in addition to smaller, significant gains on word and text reading rate. As no separate accuracy measure for word reading was included in their study, a detailed comparison of the results is not possible.

Our second research question concerned the long-term effects of the intervention in order to confirm eventual intervention effects noted at the post-test. Taking the results related to the pre–post comparisons into account, we were primarily inter-

ested in whether the effects found in Group I would prevail to follow-up and whether there were differences between Grades 2 and 3. For Group I, the gains on the task of identifying trained words remained also at follow-up, suggesting that the change was more permanent. Few intervention studies on MT have included follow-up measures. However, Torkildsen et al. (2021) found similar long-term gains on the reading of practiced words, even in another form, reading aloud from print, still present at follow-up. In our study, another long-term effect was found on the accuracy of reading aloud pseudowords in Group I. Here, Grade functioned as a moderator, as the effect was found specifically among third graders. In the study by Torkildsen et al. (2021), gains in reading of unexposed words (real words) among second grade average readers did not endure to follow-up. However, that follow-up six months later was notably later than the follow-up in the present study. Our results at follow-up raise the possibility that this kind of training could be more suitable for the reading level of the third graders.

Limitations

Characteristics of morphology vary across languages, and the development of MK varies accordingly (Reed, 2008). The Swedish language used in the training belongs to Germanic languages. The MT was directed at suffixes that are very common morphemes in Swedish. Our results might be generalisable, but only to other similar languages with similar characteristics of morphology and orthographic transparency, but naturally, this needs to be established empirically. Generally, in order to understand universalities of reading development and its support, studies carried out in various language contexts are needed.

In this study we used several measures to analyze finely tuned intervention effects on reading. We conducted many analyses of variance without correcting the p-value for significance. Therefore, the results of this study should be interpreted with caution. However, two different facts suggest that our findings are reliable and not merely coincidental; first, significant effect was found with the same measures between the same groups repeatedly in two time intervals. Second, the differences in changes between the groups were moderate or close to moderate in magnitude measured with effect size. We did not measure the potential transfer from identifying words (which was the form of reading in the training) to reading aloud the trained words from print. Therefore, we cannot say if the gains in identification of trained words could have also transferred to reading trained words aloud from print. Another limitation is that the gains in reading were noted only on measures developed specifically for this intervention. Positive outcomes are more often observed with study-specific measures than with standardised measures focusing on the general level of the skill (Carlisle et al., 2010; Goodwin & Ahn, 2013). Hence, the effects of the implicit intervention were limited, and no effects on untrained words or texts were found. To reach effects on general reading skills, a longer intervention period and more extensive content may be needed (i.e. Tijms et al., 2020; Torkildsen et al., 2021). The mean playing times for both groups were lower than the recommended time for playing. The intervention results might have been different if the recommendation had been

fulfilled, but we do not have evidence of how increased training time would affect the intervention results.

We can assume that in Group C, the business-as-usual group, most of the children were offered special educational support during the intervention period. This is due to the fact that in the Finnish school system, special educational support is provided to every student identified as needing it as part of general instruction. More than 20% of the students within basic education grades 1–9 receive special educational support for learning at some point of the school year.

Arrangements involving external personnel at the schools were impossible due to the pandemic situation. Including teachers in an intervention involved both risks and gains. In order to control for teacher and school effects, a majority of the participating teachers had both experimental and control group children in their schools. Therefore, teachers could not be kept blind to the study design. Though advised to continue regular special education for the control group children, teachers may have been tempted to offer more intensive support also to the controls and apply the intervention principles also to the control group children receiving business-as-usual support. On the other hand, involving teachers in training and as part of schooling increases the ecological validity of the study. One important reason for involving teachers is that they can be considered experts in conducting training in the school context (Samuels, 2009). Intervention research in learning has often overlooked these experts (Fischer, 2009). Hawes et al. (2020) have suggested even more intensive collaboration with teachers in creating and implementing study designs.

Further research and implications for practice

Our results imply that the Grade 3 students were more receptive to the training than the Grade 2 students, but further research is needed to study the appropriate level of reading skill required to benefit from MT. In addition, specific linguistic competencies among the participants could be included in order to study whether they correlate with positive intervention and transfer effects. For example, the level of MA among participants could be studied by measuring the level at the start and end of the intervention in order to conclude that an increase in MA can moderate or mediate eventual gains in reading. In this study, different types of morphemes were used in the training, though they might not be processed in the same way during learning, nor are they equally frequent in language data. There could be a difference between different kinds of morphemes (e.g. derivational and inflectional morphemes as well as morphemes relevant for nouns vs. verbs) (e.g. Kuo & Anderson, 2006; Lee et al., 2023). A more thorough analysis of the learning processes of different morphemes would be relevant.

The primary goal of intervention studies is to develop evidence-based practices for school use (Naveenkumar et al., 2022). The ecological validity of this study is high, and it confirmed that the training can be applied well as a part of special education practices and that intensive interventions are feasible within school hours. The training design seems applicable for poor readers, especially in Grade 3. Both in combination with teacher-led tuition or purely as an implicit form of training, this

computerised version can free teacher resources for individual support of additional students.

In this study we primarily targeted reading skill through morphological decoding (Levesque et al., 2021). We expected that the explicit training would show larger gains than the implicit training as the explicit training might have targeted also the two other dimensions, MA and MK in the Morphological Pathways Framework. These two dimensions could be expected to support the development of morphological decoding. Instead the results now seemed to support implicit training which might imply that the other dimensions (MA and MK) did not play a significant role in enhancing morphological decoding skills. However, further research with specific focus on the Morphological Pathways Framework in relation to enhancing reading is needed.

Conclusion

In line with the adaptation of the Morphological Pathways Framework by Kristensen et al. (2023) our results imply that MT can be utilized in reading interventions. The results in our study show that MT, in other words, learning to recognise and read suffixes, can provide gains in word reading among students who have passed the initial stages of their reading acquisition. Students benefit from an implicit form of MT in reading, but the design of the training needs to be matched to students' age and reading level. Hence, intervention programmes with a focus on MT could be further developed as a means of support for poor readers. When considering adding explicit elements, there might be a need to complement computerised training with teacher-led tuition.

Appendix

Table A1 Intervention Timetable

Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Activity	Screen-ing					Pre-test	Intervention period 1				Holiday	Intervention period 2				Post-test	Pause						Follow- up test

Table A2 Reading Measures

Form of assessment		Measure	Time point			
Individual			Screening	Pre-test	Post-test	Follow-up
Silent reading	Digital	Trained Word Identification		x	x	x
		Transfer Word Identification		x	x	x
Oral reading	In print	Transfer Word Reading Accuracy		x	x	x
		Transfer Word Reading Automatization		x	x	x
		Transfer Pseudoword Reading Accuracy		x	x	x
		Transfer Pseudoword Reading Automatization		x	x	x
		Word Reading	x		x	x
		Text Reading	x		x	x
Group						
Silent reading	In print	Sentence Verification	x		x	x

Table A3 Measures of Trained Morphology Content

Measure	Version A[a]		Version B[a]	
	Root word	Suffix	Root word	Suffix
Trained Word Identification	trained	trained	trained	trained
	blomm	*or*	*trapp*	*or*
	[flowers]		[stairs]	
Transfer Word Identification	not trained	trained	not trained	trained
	flagg	*or*	*klipp*	*or*
	[flags]		[cliffs]	
Transfer Word Reading	not trained	trained	not trained	trained
	kvinn	*or*	*skugg*	*or*
	[women]		[shadows]	
Transfer Pseudoword Reading	pseudoword	trained	pseudoword	trained
	glumm	*or*	*slåpp*	*or*

[a] Example of the parallel versions.

Table A4 Examples of the Parallel Word Lists Used for the Different Reading Measures

Trained words	Transfer words [1][2]	Pseudowords[2]	Length (number of letters)	CV structure
handen [the hand]	*burken* [the can]	*sinden*	6	cvccvc
bilen [the car]	*magen* [the belly]	*dålen*	5	cvcvc
fötterna [the feet]	*vännerna*[3] [the friends]	*fimmerna*	8	cvccvccv
hästarna[3] [the horses]	*lamporna*[4] [the lamps]	*dältarna*[3]	8	cvccvccv
bitar [pieces]	*burar* [cages]	*bemar*	5	cvcvc
dörrar [doors]	*bullar* [buns]	*tirrar*	6	cvccvc
gator [streets]	*sopor* [garbage]	*mudor*	5	cvcvc
blommor [flowers]	*flaggor* [flags]	*glummor*	7	ccvccvc
bilder [pictures]	*dikter* [poems]	*bulder*	6	cvccvc
datorer [computers]	*kameler* [camels]	*datiser*	6	cvcvcvc
trädet[3] [the tree]	*gruset* [the gravel]	*tröket*	6	ccvcvc
hålet [the hole]	*rådet* [the advice]	*sybet*	5	cvcvc
spåren [the traces]	*kraven*[4] [the demands]	*klyren*	6	ccvcvc
målen [the goals]	*fåren*[4] [the sheep]	*hölen*	5	cvcvc
domare [judge]	*dykare*[4] [diver]	*debare*	6	cvcvcv
jägare[3] [hunter]	*bagare* [baker]	*läpare*[3]	6	cvcvcv
städning[3] [cleaning]	*flätning*[3][4] [braiding]	*gludning*	8	ccvccvcc
mening [meaning/sentence]	*höjning* [raise]	*jating*	6–7	cvcvcc/cvccvcc
heter [is called]	*lever* [lives]	*fyter*	5	cvcvc
tycker [thinks]	*hinner* [makes it/reaches]	*nöcker*	6	cvccvc
bjudit [has invited]	*dragit* [has dragged]	*bjorit*	6	ccvcvc
tagit [has taken]	*farit*[4] [has gone]	*nagit*	5	cvcvc

⚫ Springer

Table (continued)

	Trained words	Transfer words [1][2]	Pseudowords[2]	Length (number of letters)	CV structure
	löpte [ran]	*lyfte* [lifted]	*nipte*	6	cvccv
	klippte [cut]	*knäppte*[3] [snapped, switched]	*styppte*	7	ccvcccv
	slutade [stopped]	*tvekade* [hesitated]	*klötade*	7	ccvcvcv
	letade [searched]	*hotade* [threatened]	*vosade*	6	cvcvcv
	ledig [free]	*fånig* [silly]	*hydig*	5	cvcvc
	farlig [dangerous]	*tydlig* [clear]	*perlig*	6	cvccvc
	dyrare [more expensive]	*vidare* [wider]	*tirare*	6	cvcvcv
	först [first]	*värst* [worst]	*mälst*[3]	5	cvccc
Intransparent orthographical feature[3]	4	3	3		

[1] Used for the measures Transfer Word Identification and Transfer Word Reading, different versions were used for the two measures at the same test-point and a version used at a previous test-point was not used at the consequent in neither measure .

[2] Used for the measure Pseudoword Reading.

[3] In Swedish the letter ä is sometimes pronounced as ä, and sometimes like the letter e.

[4] The frequency requirement was at least 5 attestations in the corpus, some words were less frequent but were chosen as they matched other requirements and were still considered to be familiar for the student.

Table A5 Trained Suffixes and their Functions

Suffix	Function	Type	n Root words for combinations	Example Combinations root word+suffix	English translation
-en	Noun singular definite	inflectional	25	*går**den***	the yard
-et	Noun singular definite	inflectional	22	*lag**et***	the team
-er	Noun plural indefinite	inflectional	22	*pris**er***	prices
-or	Noun plural indefinite	inflectional	18	*mamm**or***	mothers
-ar	Noun plural indefinite	inflectional	18	*stol**ar***	chairs
-na	Noun plural definite	inflectional	22	*stenar**na***	the stones
-en	Noun plural definite	inflectional	18	*ord**en***	the words
-er	Verb present tense	inflectional	18	*säg**er***	says
-te	Verb past tense	inflectional	18	*sök**te***	sought
-de	Verb past tense	inflectional	22	*ropa**de***	yelled
-it	Verb past participle	inflectional	18	*tag**it***	taken
-are	Adjective comparative	inflectional	14	*stark**are***	stronger
-st/-ast	Adjective superlative	inflectional	14	*snäll**ast***	kindest
-ig/-lig	Adjective	derivational	18	*hem**lig***	secret
-are	Noun	derivational	18	*arbet**are***	worker
-ing/-ning	Noun	derivational	18	*öv**ning***	training

Table A6 Task Type Descriptions

Aim	Task Types	Features				
		Reading	Combining letters/ word parts	Context-bound [a]	Time limited	Multi-choice
Introduction of new suffixes	Writing the suffix		x [b]			
	Identifying the suffix					x [b]
Main Task Types [c]	Word identification	x			x	x
	Word identification (selecting one correct word)			A		x
	Word identification (selecting three correct words)			A		x
	Finding a false word	x		W		x
	Finding a correct word	x		W		x
	Recognizing the suffix in a word	x			x	x
	Forming a word (root word+correct suffix)		x	A		x
Repetitive Task Types	Combining root words and their inflected forms	x				
	Identifying a word after reading root word and suffix separately				x	x [b]
	Word identification	x				x

[a] Context was given in auditive (A) or written (W) form with sentences that, for example, lacked one or several words, which the student needed to find, or as questions for the student to answer with one or several correct words.

[b] In these tasks the word "suffix" was used in the instructions for Group E but left out or reformulated in the instructions for Group I.

[c] All words (root word–suffix-combination) were trained at least once within Main Task Types and then repeated twice within either two Task Types from Main Task Types and/or Repetitive Task Types.

Table A7 Correlations Between Measures at Start

Variable	1	2	3	4ª	5	6 ª	7	8
1. Trained Word Identification								
2. Transfer Word Identification	.75**							
3. Transfer Word Reading Accuracy	.24**	.26**						
4. Transfer Word Reading Time	−.10	−.12*	−.07					
5. Transfer Pseudoword Reading Accuracy	.24**	.23**	.64**	−.28**				
6. Transfer Pseudoword Reading Time	−.07	−.10	.00	.86**	−.21**			
7. Word Reading	.27**	.27**	.32**	−.69**	.40**	−.64**		
8. Text Reading	.35**	.34**	.34**	−.62**	.36**	−.55**	.84**	
9. Sentence Verification	.31**	.29**	.23**	−.50**	.31**	−.47**	.63**	.73**

* p<.05; two tailed

** p<.01; two tailed

Note. Pearson correlations reported, except in associations in which either of the two measures was time (Transfer Word Reading Time or Transfer Pseudoword Reading Time), in which Spearman correlations are reported.

Table A8 *Correlations Between Measures at Post*

Variable	1	2	3	4ª	5	6 ª	7	8
1. Trained Word Identification								
2. Transfer Word Identification	.74**							
3. Transfer Word Reading Accuracy	.31**	.27**						
4. Transfer Word Reading Time	−.09	−.10	−.19**					
5. Transfer Pseudoword Reading Accuracy	.25*	.22**	.76**	−.22**				
6. Transfer Pseudoword Reading Time	−.08	−.07	−.16**	.88**	-19**			
7. Word Reading	.26**	.30**	.47**	−.77**	.40**	−.70**		
8. Text Reading	.31**	.33**	.47**	−.73**	.32**	−.63**	.87**	
9. Sentence Verification	.25**	.26**	.33**	−.52**	.18**	−.48**	.63**	.72**

* p<.05; two tailed

** p<.01; two tailed

Note. Pearson correlations reported, except in associations in which either of the two measures was time (Transfer Word Reading Time or Transfer Pseudoword Reading Time), in which Spearman correlations are reported.

Table A9 Correlations Between Measures at Follow-up

Variable	1	2	3	4[a]	5	6 [a]	7	8
1. Trained Word Identification								
2. Transfer Word Identification	.75**							
3. Transfer Word Reading Accuracy	.20**	.19**						
4. Transfer Word Reading Time	−.50	−.18**	−.19**					
5. Transfer Pseudoword Reading Accuracy	.09	.15*	.66**	−.15*				
6. Transfer Pseudoword Reading Time	−.02	−.11	−.14**	.88**	−.13*			
7. Word Reading	.18**	.23**	.41**	−.84**	.31**	−.75**		
8. Text Reading	.24**	.29**	.36**	−.76**	.25**	−.64**	.85**	
9. Sentence Verification	.18**	.19**	.29**	−.52**	.15*	−.47**	.63**	.71**

* p<.05; two tailed

** p<.01; two tailed

Note. Pearson correlations reported, except in associations in which either of the two measures was time (Transfer Word Reading Time or Transfer Pseudoword Reading Time), in which Spearman correlations are reported.

Declarations

Competing interests The authors declare that they have no conflicts of interest in regard to this work.

References

Arnbak, E., & Elbro, C. (2000). The effects of morphological awareness training on the reading and spelling skills of young dyslexia. *Scandinavian Journal of Educational Research*, *44*(3), 229–251. https://doi.org/10.1080/00313830050154485

Aro, M. (2005). Learning to read: The effect of orthography. In R. Malatesha, Joshi, & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 531–550). Routledge.

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*(4), 621–635. https://doi.org/10.1017/S0142716403000316

Bar-Kochva, I., Korinth, S., & Hasselhorn, M. (2020). Effects of a morpheme-based training procedure on the literacy skills of readers with a reading disability. *Applied Psycholinguistics*, *41*(5), 1061–1082. https://doi.org/10.1017/S0142716420000120

Berninger, V. W., Winn, W. D., Stock, P., Abbott, R. D., Eschen, K., Lin, S. J. (C), Garcia, N., Anderson-Youngstrom, M., Murphy, H., Lovitt, D., Trivedi, P., Jones, J., Amtmann, D., & Nagy, W. (2008). Tier 3 specialized writing instruction for students with dyslexia. *Reading and Writing*, *21*(1–2), 95–129. https://doi.org/10.1007/s11145-007-9066-x

Borin, L., Forsberg, M., & Roxendal, J. (2012). *Korp—The corpus infrastructure of Språkbanken* [The corpus infrastructure of the language bank]. In *Proceedings of LREC 2012. Istanbul: ELRA*, (pp. 474–478).

Borleffs, E., Maassen, B. A. M., Lyytinen, H., & Zwarts, F. (2017). Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: A narrative review. *Reading and Writing*, *30*(8), 1617–1638. https://doi.org/10.1007/s11145-017-9741-5

Bowers, P. N., Kirby, J. R., & Deacon, S. H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research*, *80*(2), 144–179. https://doi.org/10.3102/0034654309359353

Bryant, P., Deacon, S. H., & Nunes, T. (2006). Morphology and spelling: What do morphemes have to do with spelling? In R. Malatesha, & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 601–616). Routledge.

Burton, L., Nunes, T., & Evangelou, M. (2021). Do children use logic to spell logician? Implicit versus explicit teaching of morphological spelling rules. *British Journal of Educational Psychology*, *91*(4), e12414. https://doi.org/10.1111/bjep.12414

Carlisle, J. F., McBride-Chang, C., Nagy, W., & Nunes, T. (2010). Effects of instruction in morphological awareness on literacy achievement: An integrative review. *Reading Research Quarterly*, *45*(4), 464–487.

Casalis, S., Colé, P., & Sopo, D. (2004). Morphological awareness in developmental dyslexia. *Annals of Dyslexia*, *54*(1), 114–138. https://doi.org/10.1007/s11881-004-0006-z

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Deacon, S. H. (2012). Sounds, letters and meanings: The independent influences of phonological, morphological and orthographic skills on early word reading accuracy: Three influences on early reading. *Journal of Research in Reading*, *35*(4), 456–475. https://doi.org/10.1111/j.1467-9817.2010.01496.x

Deloitte & The European Commission (2019). *2nd survey of schools: ICT in education: objective 1 : benchmark progress in ICT in schools, final report.* Publications Office of the European Union. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57894

Ehri, L. C. (1998). Grapheme–phoneme knowledge is essential for learning to read words in English. In J. L. Metsala (Ed.), *Word recognition in beginning literacy* (1st ed., pp. 13–50). Routledge. https://doi.org/10.4324/9781410602718-6

Ehri, L. C. (2020). The science of learning to read words: A case for systematic phonics instruction. *Reading Research Quarterly*, *55*(S45-S60). https://doi.org/10.1002/rrq.334

Eklund, K., Torppa, M., Aro, M., Leppänen, P. H. T., & 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. https://doi.org/10.1037/a0037121

Fischer, K. W. (2009). Mind, brain and education: Building a scientific groundwork for learning and teaching. *Mind Brain and Education*, *3*(1), 3–16. https://doi.org/10.1111/j.1751-228X.2008.01048.x

Good, J. E., Lance, D. M., & Rainey, J. (2015). The effects of morphological awareness training on reading, spelling and vocabulary skills. *Communication Disorders Quarterly*, *36*(3), 142–151. https://doi.org/10.1177/1525740114548917

Goodwin, A. P., & Ahn, S. (2010). A meta-analysis of morphological interventions: Effects on literacy achievement of children with literacy difficulties. *Annals of Dyslexia*, *60*(2), 183–208. https://doi.org/10.1007/s11881-010-0041-x

Goodwin, A. P., & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. *Scientific Studies of Reading*, *17*(4), 257–285. https://doi.org/10.1080/10888438.2012.689791

Haugen, E. (1976). *The Scandinavian Languages: An Introduction to Their History*. Harvard University Press.

Hawes, Z., Cain, M., Jones, S., Thomson, N., Bailey, C., Seo, J., Caswell, B., & Moss, J. (2020). Effects of a teacher-designed and teacher-led numerical board game intervention: A randomized controlled study with 4- to 6-year-olds. *Mind Brain and Education*, *14*(1), 71–80. https://doi.org/10.1111/mbe.12215

Jacobson, C. (1999). How persistent is reading disability? Individual growth curves in reading. *Dyslexia*, *5*(2), 78–93. https://doi.org/10.1002/(SICI)1099-0909(199906)5:2<78::AID-DYS127>3.0.CO;2-8

Kemper, M. J., Verhoeven, L., & Bosman, A. M. T. (2012). Implicit and explicit instruction of spelling rules. *Learning and Individual Differences*, *22*(6), 639–649. https://doi.org/10.1016/j.lindif.2012.06.008

Kirby, J. R., Deacon, S. H., Bowers, P. N., Izenberg, L., Wade-Woolley, L., & Parrila, R. (2012). Children's morphological awareness and reading ability. *Reading and Writing*, *25*(2), 389–410. https://doi.org/10.1007/s11145-010-9276-5

Kristensen, J. K., Andersson, B., Bratlie, S. S., & Torkildsen, J. V. K. (2023). Dimensionality of Morphological Knowledge—Evidence from Norwegian Third Graders. *Reading Research Quarterly*, *58*(3), 406–424. https://doi.org/10.1002/rrq

Kuo, L., & Anderson, R. C. (2006). Morphological awareness and learning to read: A cross-language perspective. *Educational Psychologist*, *41*(3), 161–180. https://doi.org/10.1207/s15326985ep4103_3

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology*, *100*(1), 150–161. https://doi.org/10.1037/0022-0663.100.1.150

Lee, J., won, Wolters, A., & Grace Kim, Y. S. (2023). The relations of morphological awareness with language and literacy skills vary depending on orthographic depth and nature of morphological awareness. *Review of Educational Research*, *93*(4), 528–558. https://doi.org/10.3102/00346543221123816

Levesque, K. C., Breadmore, H. L., & Deacon, S. H. (2021). How morphology impacts reading and spelling: Advancing the role of morphology in models of literacy development. *Journal of Research in Reading*, *44*(1), 10–26. https://doi.org/10.1111/1467-9817.12313

Lyytinen, H., Erskine, J., Tolvanen, A., Torppa, M., Poikkeus, A. M., & Lyytinen, P. (2006). Trajectories of reading development: A follow-up from birth to school age of children with and without risk for dyslexia. *Merrill-Palmer Quarterly*, *52*(3), 514–546. https://doi.org/10.1353/mpq.2006.0031

Martin-Chang, S. L., & Levy, B. A. (2006). Word reading fluency: A transfer appropriate processing account of fluency transfer. *Reading and Writing*, *19*(5), 517–542. https://doi.org/10.1007/s11145-006-9007-0

Martin-Chang, S. L., Levy, B. A., & O'Neil, S. (2007). Word acquisition, retention, and transfer: Findings from contextual and isolated word training. *Journal of Experimental Child Psychology*, *96*(1), 37–56. https://doi.org/10.1016/j.jecp.2006.08.004

McTigue, E. M., Solheim, O. J., Zimmer, W. K., & Uppstad, P. H. (2020). Critically reviewing graphogame across the world: Recommendations and cautions for research and implementation of computer-assisted instruction for word-reading acquisition. *Reading Research Quarterly*, *55*(1), 45–73. https://doi.org/10.1002/rrq.256

Naveenkumar, N., Georgiou, G. K., Vieira, A. P. A., Romero, S., & Parrila, R. (2022). A systematic review on quality indicators of randomized control trial reading fluency intervention studies. *Reading & Writing Quarterly*, *38*(4), 359–378. https://doi.org/10.1080/10573569.2021.1961647

Nunes, T., Bryant, P., & Olsson, J. (2003). Learning morphological and phonological spelling rules: An intervention study. *Scientific Studies of Reading*, *7*(3), 289–307. https://doi.org/10.1207/S1532799XSSR0703_6

Nunes, T., Bryant, P., & Bindman, M. (2006). The effects of learning to spell on children's awareness of morphology. *Reading and Writing*, *19*(7), 767–787. https://doi.org/10.1007/s11145-006-9025-y

Official Statistics of Finland (OSF) (2020). *Support for learning*. Statistics Finland. http://www.stat.fi/til/erop/2020/erop_2020_2021-06-08_tie_001_en.html

Rastle, K. (2019). The place of morphology in learning to read in English. *Cortex; A Journal Devoted To The Study Of The Nervous System And Behavior*, *116*, 45–54. https://doi.org/10.1016/j.cortex.2018.02.008

Rastle, K., Lally, C., Davis, M. H., & Taylor, J. S. H. (2021). The dramatic impact of explicit instruction on learning to read in a new writing system. *Psychological Science*, *32*(4), 471–484. https://doi.org/10.1177/0956797620968790

Reed, D. K. (2008). A synthesis of morphology interventions and effects on reading outcomes for students in grades K–12. *Learning Disabilities Research & Practice*, *23*(1), 36–49. https://doi.org/10.1111/j.1540-5826.2007.00261.x

Risberg, A. K., Vataja, P., Plyhm, L., Lerkkanen, M. K., Aro, M., Westerholm, J., & Salmi, P. (2019). *ILS-Individuell läsning och skrivning. Kartläggningsmaterial för åk 1, 2, 3 och 5. Handledning* [ILS – Individual reading and writing. Tests for Grade 1, 2, 3 and 5. Manual]. Niilo Mäki Institute.

Risberg, A. K., Widlund, A., Hellstrand, H., Vataja, P., & Salmi, P. (2023). Profiles of reading fluency and spelling skills: Stability and change across the early school years. *Scandinavian Journal of Educational Research*, *68*(6),1231–1246. https://doi.org/10.1080/00313831.2023.2228822

Samuels, B. M., & Mind (2009). Can the differences between education and neuroscience be overcome by mind, brain and education?. *Brain and Education*, 3(1), 45–55. https://doi.org/10.1111/j.1751-228X.2008.01052.x.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174.

Share, D. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*(2), 151–218. https://doi.org/10.1016/0010-0277(94)00645-2

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of over-reliance on an 'outlier' orthography. *Psychological Bulletin*, *134*(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584

Teleman, U., Hellberg, S., & Andersson, E. (1999). *Svenska akademiens grammatik.* [The Swedish Academy's Grammar]. Norstedt.

Tijms, J. (2004). A process-oriented evaluation of a computerised treatment for dyslexia. *Educational Psychology*, *24*(6), 767–791. https://doi.org/10.1080/0144341042000271719

Tijms, J., & Hoeks, J. (2005). A computerized treatment of dyslexia: Benefits from treating lexico-phonological processing problems. *Dyslexia*, *11*(1), 22–40. https://doi.org/10.1002/dys.283

Tijms, J., Pavlidou, E. V., & Hoette, H. A. I. (2020). Improvements in reading and spelling skills after a phonological and morphological knowledge intervention in Greek children with spelling difficulties: A pilot study. *European Journal of Special Needs Education*, *35*(5), 711–721. https://doi.org/10.1080/08856257.2019.1709702

Torkildsen, J. K., Bratlie, S. S., Kristensen, J. K., Gustafsson, J. E., Lyster, S. A. H., Snow, C., Hulme, C., Mononen, R. M., Næss, K. A. B., López-Pedersen, A., Wie, O. B., & Hagtvet, B. (2021). App-based morphological training produces lasting effects on word knowledge in primary school children: A randomized controlled trial. *Journal of Educational Psychology, 114*(4), 833-854. https://doi.org/10.1037/edu0000688

Torppa, M., Eklund, K., van Bergen, E., & Lyytinen, H. (2015). Late-emerging and resolving dyslexia: A follow-up study from age 3 to 14. *Journal of Abnormal Child Psychology*, *43*(7), 1389–1401. https://doi.org/10.1007/s10802-015-0003-1

Vataja, P.,Lerkkanen, M. K., Aro, M., Westerholm, J., Risberg, A. K., & Salmi, P.(2021). The Predictors of Literacy Skills among Monolingual and BilingualFinnish–Swedish Children During First Grade. *Scandinavian Journal of Educational Research*, *66*(6), 960–976. https://doi.org/10.1080/00313831.2021.1942191

Wessén, E. (1969). *De nordiska språken* (9th ed.). AWE/Geber.

Authors and Affiliations

Nea Kronberg[1,2] · **Mikko Aro[1,3]** · **Kenneth Eklund[4]** · **Tomas Lehecka[5]** · **Pia Vataja[1,2]** · **Paula Salmi[2,6]**

✉ Nea Kronberg
  nea.kronberg@nmi.fi

[1] Department of Education, University of Jyväskylä, Jyväskylä, Finland

[2] Niilo Mäki Institute, Fin-40 101 Jyväskylä, P.O. Box 29, Jyväskylä, Finland

[3] Centre of Excellence in Learning Dynamics and Intervention Research (InterLearn), University of Jyväskylä and University of Turku, Turku, Finland

[4] Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland

[5] Faculty of Arts, Psychology and Theology, Åbo Akademi University, Turku, Finland

[6] Faculty of Medicine, University of Helsinki, Helsinki, Finland