Effects of a phonics intervention in a randomized controlled study in Swedish second-grade students at risk of reading difficulties

Hanna Lindström-Sandahl, Åsa Elwér, Stefan Samuelsson, Henrik Danielsson

Teaching phoneme awareness to children at risk for early reading difficulties has been recognized as successful in several studies. In this randomized controlled trial (RCT)-study, we add to this research by optimizing core procedural as well as teaching components in a phonics-directed intervention and extend the RCT reading intervention research into a semi-transparent language context. The aim of the present study was to evaluate the effects of a novel Swedish intensive phonics program. This randomized controlled pre-test and post-test intervention study targeted second-grade students with early reading difficulties. Students were identified by a repeated screening procedure and allocated to intervention (n = 34) and control (n = 34) conditions. A 9-week intensive phonics-based program was administrated one-to-one, by special education teachers in Swedish mainstream elementary schools. Results show an improvement in the intervention group, compared with the controls on all outcome measures. Findings indicate that the supplementary phonics program, delivered with high intensity, can significantly increase word reading skills and reading comprehension in second-grade students with early reading difficulties.

KEYWORDS
decoding, dyslexia, instruction, intervention, phonology, reading
The acquisition of reading is a foundational outcome of early school instruction that has long-term effects on subsequent academic performance (Cunningham & Stanovich, 1997; Sparks et al., 2014). Timely and well-planned teaching of word reading skills can have a substantial impact on the early acquisition of reading in general, not least for children struggling with the phonemic code of reading and at risk for developing reading difficulties (Gersten et al., 2020; Lovett et al., 2017). The most recent meta-analyses indicate intervention effects between 0.16 (grade 3–5) and 0.36 (grade K-2) for children who are identified with or at risk of word reading difficulties (Hall et al., 2023) with a mean effect size of 0.31 in decoding for K-6 readers (Dietrichson et al., 2021). In trying to refine intervention effects across studies accounting for design, intervention content, intensity, teacher-child setting, sample size, and orthographic structures of different alphabetic writing systems, this field of research needs more research to pinpoint the best educational practice for children at risk for early reading difficulties. Contrary to most research using experimental designs in English-speaking countries (Share, 2021), randomized controlled studies of early literacy intervention are scarce in European languages and more so in a Scandinavian context (Pontoppidan et al., 2018).

The overall purpose of this randomized controlled trial (RCT) study was to optimize the acquisition of early word decoding, recognition, and comprehension skills in the Swedish semi-transparent orthography by evaluating a program aimed at students with low reading skills after the initial year of school instruction. The intervention emphasized structured phonics, one-to-one teaching, and high intensity.

1 | BACKGROUND

Children typically establish basic phonological decoding skills through a gradual development in recognition of short, regular, and high-frequency words within their first year of literacy instruction (Seymour et al., 2003; Snow et al., 1998). Before formal reading instruction, phonological awareness, in particular phoneme awareness (PA), is a strong predictor of future reading ability. PA is part of a triad of language and cognitive skills including letter knowledge and rapid automatized naming, Rapid Automatized Naming (RAN), considered to underlie individual differences in reading development (Melby-Lervåg et al., 2012). Preschool PA explains significant variance in subsequent reading performance, and it asserts its strongest influence during the initial phase of learning to read (Caravolas et al., 2013; Furnes & Samuelsson, 2011; Lervåg et al., 2009). Individual differences in phoneme awareness are also associated with children establishing early reading difficulties (Melby-Lervåg et al., 2012). Longitudinal and experimental research covering a range of disciplines and methods state that early reading difficulties in children are most likely related to, and even caused by, phonological deficits (for a review see Perfetti et al., 2019). It is argued that phonological problems obstruct the development of explicit grapheme and phoneme correspondences (Vellutino et al., 2004) with subsequent difficulties to map larger orthographic units onto phonology (Ehri, 2014). Though early intervention is crucial for children with weak pre-reading skills, students at risk for early reading difficulties are more reliably identified after the onset of formal reading instruction at school, rather than at preschool or kindergarten ages (Byrne et al., 2009; Olson et al., 2014; Poulsen et al., 2017). Given the causal impact of phoneme awareness on early reading development, it is a critical component of remedial interventions for struggling readers (Suggate, 2016).

1.1 | Phonics as remedial intervention

Following the predictive and supportive role of phonemic awareness in learning to read, systematic instruction, that is phonics-based letter-sound manipulation, segmenting, and blending, should be applied to promote decoding skills in children with reading difficulties (Duff & Clarke, 2011; Foorman et al., 2003). The success of the phonics instructional approach lies in concretizing the relations of spoken language segments to print (McArthur et al., 2018). Consequently, phonics teaching should include letter-sound instruction (Bus & van Ijzendoorn, 1999; Ehri et al., 2001;
Suggate, 2010) and an element of sight word reading (McArthur, Castles, et al., 2015; McArthur, Kohnen, et al., 2015) to maximize reading development. The learning content should be presented explicitly, intentionally, and with plenty of opportunities for the student to respond (see Fletcher et al., 2021 for a review).

Features of explicit instruction include direct teaching of how letters represent speech sounds in print, blending learned letter-sound correspondences, sounding out unfamiliar words, and providing students with decodable texts (Ehri et al., 2001; Stein et al., 1999). Explicit phonics instruction is known to benefit both reading and spelling in poor readers and writers (Graham et al., 2002; Moats, 2009; Rupley et al., 2009). Some studies have even tried to quantify and diverge intervention effects by the degree of explicitness in phonics programs implemented in at-risk elementary students (Foorman et al., 1998). Their findings suggest that the degree of explicitness in phonics programs implemented in at-risk elementary students is important. The more explicit instructional approach yielded gains in students’ word-reading ability that exceeded gains from phonologically implicit or embedded instruction, pointing to the necessity of explicitly showing students the what and how of phonics during instruction.

Systematic teaching is another key factor in phonics instruction (Ehri et al., 2001; Galuschka et al., 2014; McArthur et al., 2018; NICHD, 2000; Torgerson et al., 2019). The approach consists of clear explanations and modeling of desired concepts and skills, as well as feedback to the student during guided and independent practice. The scope and sequence, which characterize systematized interventions, define which phonics content is included and a predetermined order of teaching letter-sound correspondences (Mesmer & Griffith, 2005).

1.2 Phonics in semi-transparent orthographies

The rate of literacy acquisition is related to orthographic consistency. The more transparent the orthography, the more direct is the correspondence between letters and phonemes, making accurate word decoding faster and more accessible compared with learning in opaque orthographies (Seymour et al., 2003). In semi-transparent grapheme–phoneme systems, such as Swedish, the impact of PA in reading decreases after the first year of formal reading instruction (Furnes & Samuelsson, 2010), and reading difficulties are commonly reflected in limited reading speed, rather than accuracy (Landerl & Wimmer, 2008).

Code-based instruction has traditionally been taught in Scandinavian/Nordic elementary classrooms, but teaching phonics has rarely been evaluated for effectiveness in Swedish educational research (Levlin & Nakeva von Mentzer, 2020; Pontoppidan et al., 2018; Taube et al., 2015). To date, only three RCTs of reading interventions representing semi-transparent Swedish, Norwegian, and transparent Finnish orthography have been reported (Saine et al., 2011; Solheim et al., 2018; Wolff, 2011, 2016). First, at-risk Norwegian first graders participated in a multi-component literacy intervention conducted by Solheim et al. (2018). Their semi-computerized program included four systematic phonics training lessons a week over 25 weeks administrated in small groups of 3–7 children and resulted in improved word reading. Second, Saine et al. (2011) implemented an intensive computerized phonics intervention in Finnish first graders four times a week for 28 weeks. The training was conducted in small groups of five children. Positive outcomes were found on reading, fluency, and spelling measures compared with controls in regular special education. Third, Wolff (2011, 2016) found that a one-to-one, explicit, and extensive phonics intervention had a superior and lasting effect on literacy outcomes, including reading speed, in Swedish third-grade students with reading difficulties. The intervention spanned five weekly lessons for 12 weeks.

Further evidence of the training effects of systematic phonics programs in semi-transparent European orthographies includes Volkmer et al. (2019) who evaluated an intervention with German first graders. This study included at-risk students based on reading speed. Participants improved more in reading accuracy compared with the controls; however, reading speed was not improved. An unselected sample of Dutch kindergarteners was randomized to computerized systematic, unsystematic, or no phonics training (de Graaff et al., 2009). The authors concluded that systematic phonics was superior in improving phonemic awareness, word reading, and spelling.
1.3 | The organizational arrangement of reading interventions

Theory and research reports indicate that interventions in schools should be organized as part of the response to intervention framework (RTI), a three-tiered system implementing increasing levels of support. Increasing intensity and individual tutoring by qualified teachers are associated with Tier 3 support (Al Otaiba et al., 2016; Gersten et al., 2008). With an RTI approach, schools should direct successively more focused attention to students struggling with the acquisition of fundamental literacy skills. These tiered approaches are recommended if applied systematically, but there is also research suggesting directly offering intense, systematic interventions to elementary-grade students at risk of reading difficulties to avoid the wait-to-fail approach (Al Otaiba et al., 2014).

For early grades, one-to-one or small-group interventions are associated with positive effects on reading outcomes (Schwartz et al., 2012; Vaughn et al., 2010). The one-teacher-one-student format allows for extended opportunities to practice with timely teacher feedback (Vaughn et al., 2010). Moreover, the high intensity may offer an opportunity for the teacher and student to get more familiar, to establish daily routines, and promote student motivation during training (Neitzel et al., 2022; Warren et al., 2007). Research syntheses report the one-to-one approach as superior to small group instruction in first-grade reading interventions (Gersten et al., 2020). Tier 3 interventions for struggling readers (Neitzel et al., 2022), and for students up to 10 years (Slavin et al., 2011). In a meta-analysis of 20 reading interventions, Gersten et al. (2017) found supplemental reading interventions for at-risk students in grades 1–3 to generate higher mean effect sizes for word reading and reading comprehension outcomes when implemented one-to-one or in a small group with higher mean effect sizes for the individually administered interventions.

1.4 | Rationale

The benefits of phonics as remedial instruction for poor word reading are considered indisputable (Castles et al., 2018; Ehri et al., 2001; Fletcher et al., 2021). Only a few RCT studies evaluating phonics-oriented programs have been performed in a Scandinavian/Nordic context (Saine et al., 2011; Solheim et al., 2018; Wolff, 2011, 2016). These studies vary in their scope of interventions, age of participants, and selected groups. To be more specific, our study should extend our understanding of phonics to promote early reading development in a European context. First, two out of three previous RCT-studies focusing on phonics included a randomized procedure within stratified samples. Apparently not a complete RCT study (cf. Saine et al., 2011; Solheim et al., 2018). In addition, these two studies implemented phonics intervention using computerized teaching of phonics. Saine et al. study from Finland is an RCT-study performed within a fully transparent orthography. Finally, the only complete RCT-designed study on phonics in a semi-transparent European orthography included Swedish young students (Wolff, 2011) in grade 3 (students aged 9 years with more than 2 years of formal literacy instruction) and emphasized reading fluency more than accuracy difficulties facing children at risk for literacy problems at grade 1. Therefore, it is important to establish what findings from interventions in an English-speaking context that can transfer to other alphabetic orthographies. Swedish orthography is not fully phonemic but presents a high correspondence between phonology and orthography (Riad, 2014). Moreover, it is characterized by numerous vowels and marking of vowel length, a rich inflectional and derivational morphology, and a high frequency of compound words (Guron & Lundberg, 2004; Hedlund et al., 2001; Riad, 2014).

In the present study, we implement an RCT in a Swedish context at the beginning of second grade, that is, after 1 year of formal reading instruction. The present program corresponds to a Tier 3 intervention in intensity and tutoring. The intervention program includes the combination of phonics and extensive decoding practice aligned with the overall targets of similar explicit phonics programs (Levin & Nakeva von Mentzer, 2020; McArthur, Castles, et al., 2015; McArthur, Kohnen, et al., 2015). Instructions are manualized to ensure teachers deliver instructions that are highly explicit and provide content in a systematic manner (Gersten et al., 2008). The program is considered extensive compared with some literacy interventions in semi-transparent orthographies (Saine et al., 2011;
Solheim et al., 2018) as it spans 36 individual lessons. The program incorporates additional literacy components over and above mapping of graphemes to phones and decoding, addressing fluency training and spelling (Fletcher et al., 2021; Möller et al., 2021; Torgesen et al., 2001; van Rijthoven et al., 2021; Weiser & Mathes, 2011). Fluency exercises in the format of repeated reading are considered important due to remaining difficulties with reading fluency, rather than accuracy in struggling learners of more transparent orthographies (Seymour et al., 2003).

The aim of the study was to investigate the effects of an intensive phonics reading intervention on decoding, word reading, and comprehension in a group of second-grade Swedish students at risk for reading difficulties.

2 | METHOD

2.1 | Intervention and control groups

Young students at risk for early reading difficulties are more reliably identified after the onset of formal reading instruction at school, rather than using precursors of early reading acquisition such as letter knowledge, phonological awareness, and RAN observed in preschool or kindergarten ages (Byrne et al., 2009; Olson et al., 2014; Poulsen et al., 2017). Two cohorts of grade 1 children were recruited from 21 small to medium sized mainstream Swedish public elementary schools during two consecutive school years: February 2020 and 2021 (6 months after entering formal schooling in August; 2019 and 2020, respectively). Note that Sweden decided to have all elementary schools open as usual during the COVID-19 pandemic hampering primary schools in Europe these 2 years (OECD & OCDE, 2021). Recently, with a data set comprising 97,073 primary school students, Hallin et al. (2022) reported no decline in literacy skills (word decoding and reading comprehension) during the COVID-19 pandemic in Sweden.

In total, six municipalities in rural and central parts of the southern-east region of Sweden were engaged with a participation rate of around 60%. We used two main exclusion criteria. As the intervention procedure (described below) demands attention and the ability to focus on tasks as well as certain levels of skills in Swedish, we asked teachers to exclude children with behavioral difficulties covering all neuropsychiatric disorders and severe sensory deficits that most likely would impede participation. We also excluded students who had attended less than 1 year of kindergarten schooling in Sweden before formal schooling starting at grade 1. Only eight children were excluded based on these exclusion criteria. Parental reporting of symptoms of attention deficit hyperactivity disorder (ADHD) in Sweden is lower compared with many other countries, influencing clinical practices, and imposing a more restrictive approach to diagnosing children at this age (MacDonald et al., 2019). The final sample consisted of 753 students. The parents were informed in written form and formally consented to participate.

The main strategy to identify children at risk for early reading difficulties was to perform two identifying tests (word recognition and spelling), when all children had received formal teaching in reading for 6 months in grade 1 (February 2020), repeat the same test 3 months later (May 2020). We used the word-chain test measuring word recognition (Jacobson, 2014) where children identify and mark the boundaries separating three individual words in letter strings for 2 min. A total of 80 letter strings were included in the test. Reported test–retest reliability was 0.85–0.90. A spelling test included in a standardized test battery (Elwär et al., 2016) was employed as a second identification test. In this task, children spell single words from dictation. The 25 first words from the test were administered. Test–retest reliability was 0.45–0.87. Both identifying tests were administered in whole classes or groups of approximately 10 children.

Children who performed at or below the 25th percentile in both word recognition and spelling at both identifying test sessions (i.e., February and May in grade 1) were considered at risk for reading difficulties. A total of 68 children (approximately 9% of the total sample) were then randomly allocated to an intervention (n = 34) or control group (n = 34). With this fully randomized procedure, 10 schools had three or more children in the intervention or control groups (varied between 3 and 9 students, 53 children in total) and the remaining schools had only 1–2 children included in one or both groups (15 children). Descriptives for the word chain and the spelling test at screening 1 and screening 2 are presented in Table 1.
### TABLE 1
Mean, SD, min, max, skewness, and kurtosis for the reading tests at screening 1 and 2 for the later allocated groups.

<table>
<thead>
<tr>
<th>Test</th>
<th>Group</th>
<th>Screening 1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Screening 2</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>Min</td>
<td>Max</td>
<td>Skewness</td>
<td>Kurtosis</td>
<td>M</td>
<td>SD</td>
<td>Min</td>
<td>Max</td>
<td>Skewness</td>
<td>Kurtosis</td>
</tr>
<tr>
<td>Word chain</td>
<td>Controls</td>
<td>1.68</td>
<td>1.27</td>
<td>0</td>
<td>4</td>
<td>0.18</td>
<td>2.03</td>
<td>3.65</td>
<td>1.91</td>
<td>0</td>
<td>7</td>
<td>−0.07</td>
<td>2.36</td>
</tr>
<tr>
<td>Word chain</td>
<td>Intervention</td>
<td>2.26</td>
<td>1.38</td>
<td>0</td>
<td>7</td>
<td>1.07</td>
<td>5.33</td>
<td>3.94</td>
<td>1.84</td>
<td>0</td>
<td>7</td>
<td>−0.15</td>
<td>2.10</td>
</tr>
<tr>
<td>Spelling</td>
<td>Controls</td>
<td>4.09</td>
<td>2.04</td>
<td>0</td>
<td>9</td>
<td>0.23</td>
<td>2.95</td>
<td>6.21</td>
<td>3.17</td>
<td>0</td>
<td>11</td>
<td>−0.03</td>
<td>1.88</td>
</tr>
<tr>
<td>Spelling</td>
<td>Intervention</td>
<td>4.56</td>
<td>2.86</td>
<td>0</td>
<td>9</td>
<td>−0.11</td>
<td>1.75</td>
<td>7.35</td>
<td>3.14</td>
<td>1</td>
<td>11</td>
<td>−0.40</td>
<td>1.93</td>
</tr>
</tbody>
</table>
The average age was 96.8 (4.0 SD) months in the intervention group and 96.5 (5.0 SD) in the control group with 56% male students in each group. WISC vocabulary and block design tests were performed on all children in the intervention and control groups to estimate and compare general cognitive abilities (Wechsler, 2014). The mean scaled score on WISC block design was 8.4 (2.7 SD) in the intervention group and 8.8 (2.5 SD) for the control group, whereas scaled WISC vocabulary was estimated to be 6.4 (2.9 SD) and 6.8 (3.3 SD), respectively. Importantly, this description shows no differences across groups, but indicates significantly lower intellectual capabilities according to Swedish norms on WISC.

2.2 | Pre- and post-tests

Pre-tests were administrated in early September (grade 2) 2 weeks before intervention and the same tests were used as post-tests in November/December later the same year. All pre- and post-tests were administrated individually for both groups at each school by special education teachers. Word recognition was measured by a standardized word reading test (Elwér et al., 2016). The children were asked to read aloud word lists (A and B) consisting of 100 words each as quickly as possible within 45 s. The sum (A + B) of words read correctly was registered. Reported test–retest reliability (1 month between test administration) is 0.93 and criterion validity 0.88. Word decoding, also from Elwér et al. (2016), was measured by two lists of nonwords (A and B; 63 nonwords in each list) and the total number of correctly read nonwords from each list within 45 s was recorded. Test–retest reliability (1 month) taken from the manual is 0.78 and criterion validity is 0.75. Note that we were not able to include standardized scores addressing percentiles, reading age, nor grade level reading skills evaluating intervention effects as most children were at least 1 year behind typical readers aged 7.9 years. Reading passages comprehension (Woodcock, 1987) was adapted to Swedish as a measure of reading comprehension. In the test, children are asked to read 68 text passages and to fill in a missing word in sentences or short paragraphs. Testing is terminated when the child has made errors on six consecutive items.

2.3 | Intervention content and procedure

All students had received reading and writing instruction for a full school year before the intervention and were aware of letter–sound correspondences. However, their age-equivalent skills in word reading and spelling were on average at least 1 year behind. The intervention was administered by 13 special education teachers working at the students’ schools. In addition, there were three newly retired, experienced elementary teachers and one member of the research team who delivered the intervention. Most special education teachers were well known to the students since they are often engaged in classroom work in the elementary grades. Due to the nature of the intervention, it was not possible to keep teachers blind to the conditions of participants in their own schools.

The control groups received teaching as usual attending regular classroom instruction and possible supplemental special education or other arrangements (small group, teaching assistant support etc.). For ethical approval and to ensure that those children at risk for reading difficulties assigned to the control group received adequate and planned educational support during the intervention period, each school was financially supported by covering 1 month’s teacher salary for each child enrolled in the intervention group. We asked participating schools about children in the control group and other students (not identified by our screening procedure) in need of special education during the 4-month period implementing the intervention. All schools with three or more students included in the intervention (10 schools) secured extra special teaching resources by either engaging newly retired special education teachers or locum tenens. The aim was to maintain support as planned for each child. Some schools report that this effort might have been undermined not being performed by their regular teacher, but several schools also report that the project helped children in need of educational support and, not least, that the extra financial support from the project helped step up the ambition to reach as many as possible with special needs.
Classroom teachers, principals, and special education teachers took part in separate pre-intervention meetings. For special education teachers, 2 × 2 h preparation consisted of instruction in program and lesson content, intervention, and test administration. Students in the intervention group received 36 detailed structured intervention 35-min lessons, 4 days a week for 10 weeks (21 h of instruction, 1-week autumn break), administered one-to-one. Due to temporary teacher and child absence (mainly caused by official rules to recover from COVID-19), the intervention period was extended to a maximum of 12 weeks for a minority of the intervention group. Missed lessons were compensated for within the timeframe of 12 weeks. The schools were free to schedule the intervention at their convenience during school hours.

The teachers monitored the implementation of each lesson by registering deviations depending on student factors or the content and delivery of the intervention in an intervention protocol. During the implementation period, a couple of shared, digital meetings were scheduled with all interventionists for fidelity checks and support from the research team. Additional contact and practical support from the team were offered to the teachers on demand in the weeks of the intervention. A brief written questionnaire where teachers were asked about opinions on content and procedures was administered when the intervention was finished. Due to COVID-19 restrictions at the time of this study’s data collection, videotaped lessons, which were planned as a fidelity measure to control for program adherence, had to be cancelled.

Swedish is an orthographically semi-transparent language with few irregularly spelled words. However, Swedish is characterized by dense multi-consonant clusters and a high incidence of compound words and inflectional morphemes that can alter the pronunciation of words. Additional features are nine vowel graphemes, and 18 vowel sounds where some have alternative spellings within words (Guron & Lundberg, 2004). To emphasize phonics in the intervention, the following components represent the lesson content (see Table A1, Appendix).

### 2.3.1 | Letter knowledge practice

Isolated phonemes were coupled with corresponding graphemes in a sequence: the teacher orally introduced letter names, corresponding phonemes, and presented letter cards, after which the student identified and produced the phonemes and matched them to graphemes. These exercises were part of the first half of the program (i.e., 18 lessons), after which all letters of the alphabet had been processed individually. Several letters were introduced each week, which was considered more efficient than at a slower pace (Sunde & Lundetrae, 2019).

### 2.3.2 | Phoneme synthesis and segmentation

Students practiced blending and segmenting of orally presented words while simultaneously manipulating corresponding letter cards. Students were instructed to identify and position single phonemes in words scaffolded by empty lines that signaled the correct number of phonemes in the word. These trials were presented as an oral exercise (the first seven lessons), and the teacher wrote the whole word when the task was completed. Later in the intervention program, the students identified the phonemes and wrote the words in this exercise.

### 2.3.3 | Word recognition

Smooth blending was taught through connected phonation of phonemes during word decoding, which facilitates the transfer from decoding of words beginning with continuant consonants to words beginning with stop consonants,
for example, /p/, /t/, /g/ (Gonzalez-Frey & Ehri, 2020). To the novice reader, inconsistent orthographic-phonemic patterns could be problematic (Caravolas, 2018), hence trained words represented high to full grapheme–phoneme correspondence. Words were chosen by consulting contemporary, mainstream early literacy instruction textbooks. The sequence of word structure went from plain to more complex orthographic structure: \((V = \text{vowel, } C = \text{consonant})\) VC, CV, CVC, CVCV, CCVC, CCVCV, CCVCVC. Inflectional and derivational morphemes were introduced early in the intervention and the change of meaning was illustrated to highlight how a small orthographic segment added to a word could alter grammatical or lexical meaning. In the last stage of the program, compound words were introduced and students practiced reading and compounding.

### 2.3.4 | Sight words/high frequency words

High-frequency, short words were introduced on flashcards. The teacher read all words aloud, asked the student to read in chorus and finally to identify words. An oral explanation about how learning to read these words by recognition can facilitate fluent word reading in texts was given on each occasion. Trained sight words were successively introduced in sentence reading and writing.

### 2.3.5 | Writing/encoding

Words were dictated to the student as a repetition of words previously read in the current lesson. From mid-intervention, the student wrote words and sentences by hand whereas they previously were composed of letter cards.

### 2.3.6 | Word list reading and sentence reading

Reading aloud was practiced by wordlists and with gradually longer sentences that contained fully decodable, trained words (except sight words). Repeated reading was systematically embedded in the program and all practice words were always read at least twice. The words were compiled from a Swedish word metric database and represented fully decodable VC \((V = \text{vowel, } C = \text{consonant})\), CV, CVC, CVCV, and CCVC word structure (Witte & Köbler, 2019).

### 2.3.7 | Listening to a short story

During each lesson, a short episode of a coherent story developed for the study introduced some central piece of the learning content, that is the grapheme–phoneme correspondences or words of the day. The story was read aloud to the student. It included high-frequency, everyday vocabulary as well as content-specific words. In general, the teacher posed a question after the story tapping into story content and inferences possible to make from the text or relating to the student's world.

### 2.3.8 | Games

Short lotto, memory, or word card games were played as a final daily activity. Games were used to further reinforce the repetition of words read and/or spelled during each lesson.
2.4 | Ethics statement

The study plan was approved by the Swedish National Research Ethics Committee (ref. no 2019-04084). Students and parents of first graders received printed information and gave written consent to participate. To ensure local special education resources, financial compensation was granted by the project to municipalities for each student in the intervention.

2.5 | Data analysis

All data analyses were conducted in r (R Core Team, 2021) and RStudio (RStudio Team, 2022) and the manuscript was formatted with the papaja package (Aust & Barth, 2020). The following r packages were used: tidyverse (Wickham et al., 2019) for data wrangling and more, mice (van Buuren & Groothuis-Oudshoorn, 2011) for imputation of missing value, emmeans (Lenth, 2021) for calculation of planned contrasts, afex (Singmann et al., 2021) for plotting. There was only one missing datapoint for the reading tests, one missing value at the post-test for word decoding. That value was imputed with the mice package based on the other test scores after testing with the default method (pmm) to create one complete data set that was used in the inference statistics.

It was expected that both the controls and the intervention group would improve, which was confirmed with a t-test. The effect size (Cohen’s d) for improvement was calculated as the mean improvement from the planned contrast divided by the SD from the pre-test (pooled over both groups).

The analysis of most interest was the effect of the intervention, which was calculated as one regression analysis per reading test where the dependent variable was the post-test score for the reading test, the independent variables group, and before test score. We are not interested in the effect of the pre-test per se, but it is included in the analysis to adjust for group differences in the pre-test scores. Based on the regression analysis, planned contrasts were calculated as the intervention group score—the control group score on the post-test scores. The effect size (Cohen’s d) for improvement was calculated with the eff_size function from the emmeans package with the mean improvement from the planned contrast and pooled SD based on the regression analysis. The assumptions for regression were tested with the gvlma (Pena & Slate, 2019) package. The assumption on heteroscedasticity was not met for reading comprehension. In addition, even if it is not formally an assumption for a regression analysis, gvlma indicated that there were problems with skewness and kurtosis for the regression on word decoding. Therefore, a robust regression with the rlm function from the MASS package (Venables & Ripley, 2002) was used instead for all three tasks.

In all analyses, a directional hypothesis (post-test score is higher than the pre-test score or intervention improvement is larger than control improvement) has been used (but the 95% confidence intervals presented are based on a two-sided test to avoid that the upper confidence interval is infinity, which is uninformative).

3 | RESULTS

Table 2 presents descriptives for reading tests, divided by timepoint (before and after intervention) and group (control group and intervention group). As can be seen, kurtosis was a bit high for most tests and groups. Table 2 also presents the improvement, as tested with t-tests. As expected, there were significant improvements on all tests for all groups. The effect sizes for the intervention group (1.56; 1.54; 1.38) were larger than those for the control group (0.96; 1.16; 1.05).

The focus of the intervention was to improve the student’s ability to read words by emphasizing the connection of phonemes and graphemes and practicing word recognition.
TABLE 2  Descriptives (median, mean, SD, skewness, and kurtosis) before and after the intervention for the reading tests for both groups (in addition, the improvement for each test and group is presented as mean improvement, statistics from a t-test, and effect size as Cohen's $d$).

<table>
<thead>
<tr>
<th>Test</th>
<th>Before</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Improvement</th>
<th>Cohen's $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group</td>
<td>$n$</td>
<td>$Mdn$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$Skewness$</td>
<td>$Kurtosis$</td>
<td>$\Delta M$</td>
</tr>
<tr>
<td>Word recognition</td>
<td>Controls</td>
<td>34</td>
<td>33.00</td>
<td>37.09</td>
<td>19.71</td>
<td>0.63</td>
<td>2.72</td>
<td>16.85</td>
</tr>
<tr>
<td></td>
<td>Intervention</td>
<td>34</td>
<td>37.00</td>
<td>37.65</td>
<td>15.06</td>
<td>0.11</td>
<td>2.68</td>
<td>9.56</td>
</tr>
<tr>
<td>Word decoding</td>
<td>Controls</td>
<td>34</td>
<td>24.00</td>
<td>24.44</td>
<td>10.16</td>
<td>0.22</td>
<td>2.72</td>
<td>11.76</td>
</tr>
<tr>
<td></td>
<td>Intervention</td>
<td>34</td>
<td>25.50</td>
<td>26.94</td>
<td>10.06</td>
<td>0.07</td>
<td>2.75</td>
<td>15.59</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>Controls</td>
<td>34</td>
<td>6.00</td>
<td>7.56</td>
<td>5.27</td>
<td>0.88</td>
<td>2.87</td>
<td>5.56</td>
</tr>
<tr>
<td></td>
<td>Intervention</td>
<td>34</td>
<td>7.50</td>
<td>8.35</td>
<td>5.31</td>
<td>0.37</td>
<td>1.82</td>
<td>7.29</td>
</tr>
</tbody>
</table>
The planned contrast with the pre-test score as covariate in a robust regression showed a large difference between groups after the intervention ($\Delta M_{\text{word recognition}} = 10.49$, 95% confidence interval, CI [5.46, 15.53], $t(65) = 4.08$, $p < 0.001$, Cohen’s $d = 1.04$ [0.54, 1.53]). A plot of the improvement in word recognition can be found in Figure 1.

The word decoding task was used at pre- and post-test to examine the degree of establishment of the alphabetic principle in the intervention and control group. As phoneme–grapheme correspondence as well as blending and segmenting of phonemes and words were key aspects of the intervention; progress in word decoding was expected in the intervention group even though word decoding per se was not included in the intervention program.

The planned contrast with the before-score as covariate in a robust regression showed a difference between groups after the intervention. The intervention group showed greater progress between pre- and post-test in word decoding compared with the control group. The difference between the groups was of medium size ($\Delta M_{\text{word decoding}} = 4.54$, 95% CI [0.87, 8.22], $t(65) = 2.42$, $p = 0.008$, Cohen’s $d = 0.54$ [0.10, 0.97]). The intervention appears to have accelerated the establishment of the alphabetic principle in the intervention group. A plot of the improvement in word decoding can be found in Figure 2.

Reading comprehension was not explicitly addressed in the intervention, but the program included aspects of listening comprehension such as discussions about a story. However, as decoding is an important prerequisite of reading comprehension in the early grades, intervention effects were expected.

The planned contrast with the pre-test score as a covariate in a robust regression showed a difference between groups after the intervention. Across the 9-week period, the reading comprehension results of the intervention group improved more than those of the control group. The effects medium were significant ($\Delta M_{\text{reading comprehension}} = 2.59$, 95% CI [-0.10, 5.28], $t(65) = 1.88$, $p = 0.030$, Cohen’s $d = 0.50$ [-0.02, 1.02]). The reading comprehension improvement is visualized in Figure 3.
FIGURE 2  Development of word decoding from pre- to post-test.

FIGURE 3  Development of reading comprehension from pre- to post-test.
4 | DISCUSSION

Intervention research addressing children with delayed capacities to develop early age-appropriate literacy is extensive. However, there are few RCT’s studies addressing phonics for children at risk for early reading difficulties. Some more but still rare in English orthographies and even more so in more transparent European orthographies transparent languages. To address similarities and dissimilarities across those RCT studies (both in English and other orthographies) is manageable, but to disentangle these differences scrutinizing methodological differences, intervention procedures and teaching materials, not least its potential impact on intervention effects, seem too early for a more detailed meta-analysis. We argue that it is important to empirically secure and generalize the value of phonics in intervention programs addressing young students at risk for reading difficulties across alphabetic orthographies.

Recent review articles in this field embracing high standards for inclusion, design, and fidelity reduce this research to about 50 original reports (Gersten et al., 2020; Hall et al., 2023; Neitzel et al., 2022). In a world with different school systems, resources, and orthographies, along with more settled but important moderating factors such as intervention dosage and group size (among others) influencing intervention effects, these reviews indicate that average reading intervention effects addressing phonics vary between 0.23 and 0.39. Our intervention study outlined in Swedish school context tries to replicate, qualify, and address crucial methodological issues in intervention research for children at risk for early reading difficulties.

The present study evaluated the effect of an intensive, code-based reading intervention on three reading measures with second graders who were identified as at-risk for reading failure following two screening sessions in first grade. Phonics was the main ingredient in the intervention program. The program comprised exercises in phonemic segmenting and blending, decoding using a sounding-out strategy with simple and complex word structures, sight word reading, repeated reading of words and sentences, spelling, and morphology practice of infections and compounds, the latter known to present most challenges in learning to read Swedish orthography (Guron & Lundberg, 2004). The intervention effects reported in our study (range 0.50–1.04) are well above average phonics intervention effects measured by standardized outcome measures (Gersten et al., 2020), and with a significant and positive transfer effect to reading comprehension.

4.1 | Decoding and word recognition

The intervention group made most progress in word recognition (Cohen’s d 1.04), with twice the significant effect for non-word decoding (d 0.54). These effects are of educational importance after completing intervention but still not evaluated for long-term effects (Bloom et al., 2008; Kraft, 2020). Our measures of word decoding and recognition both addressing accuracy within time limits do not allow separating decoding accuracy and word recognition speed. However, longitudinal studies mostly comparing English and European orthographies provide convincing and systematic support that decoding accuracy per se is not the primary problem for children at risk for early reading difficulties in European semi-transparent languages (Furnes & Samuelsson, 2011; Landerl & Wimmer, 2008). Instead, the main challenge seems related to increasing word reading fluency.

Our intervention effects on word reading accuracy and fluency are expected, most encouraging for word recognition accelerating by an effect size of 1.04 (d) in 3 months. It was beyond the limit of the present study to investigate the students’ level of orthographic processing, often measured by presenting a real word together with a homophone, non-word, in print and asking which one is correctly spelled (Georgiou et al., 2008). Such information could be informative considering the intervention group’s substantial improvement in word recognition, since it reflects the quality of and access to orthographic representations of lexically stored words (Cunningham et al., 2001). Recently, in a Swedish quasi-experimental second-grade study Levlin and Nakeva von Mentzer (2020) suggested systematic phonics should be made a mandatory practice in early elementary school. We consider it well
justified to recommend explicit phonics as a supplemental special education intervention to accelerate word decoding and word recognition in young Swedish students.

4.2 | Reading comprehension

Reading comprehension was not explicitly targeted in the intervention, nevertheless, an effect size of 0.50 (Cohen’s d) was obtained for the intervention group compared with the controls. However, reading interventions enhancing decoding and word recognition in early grades are likely to show a transfer to reading comprehension (Ehri, 2014; Perfetti, 1985). Students were given abundant opportunities to practice grapheme–phoneme blending, word and sentence reading was practiced to mastery, and fluency was emphasized through repeated reading. Thus, several intervention components integrated words into sentences with fluency (Torgesen et al., 2001) and enhanced the ability to derive meaning from text (Nation, 2019).

Listening to short read-aloud and conversations related to content was also part of the program. These aspects are highly beneficial for improving early reading comprehension. However, more advanced reading comprehension is a multidimensional process, which was not comprehensively covered in our intervention. Note also that the Woodcock-Johnson passage comprehension test is highly correlated with word decoding level in young children (Keenan et al., 2008). To some extent, text was orally and explicitly elaborated as teachers asked content or inference questions about the story that connected individual lessons. To speculate, listening to and talking briefly about daily text segments paired with intensive code-work proved positive for the post-test reading comprehension scores. Such literacy-stimulating activities were incorporated as we expected students would benefit from a framing story to keep on track during each lesson and get opportunities to discuss text and vocabulary in context (Baker et al., 2013).

Our results indicate that a supplemental intervention specifically targeting decoding and word recognition skills can indeed increase reading comprehension.

4.3 | Program design aspects

Intervention dosage and intensity in terms of frequency, group size, duration, and time management within every session (Vaughn et al., 2012) were considered in designing the intervention package. First, the one-to-one format has been associated with successful student learning, but results are mixed suggesting small groups as an equally effective practice (Wanzek & Vaughn, 2007). Our study adds support to intervention programs with a one-to-one student format (Schwartz et al., 2012; Vaughn et al., 2010; Wanzek et al., 2018).

Second, our intervention program addressed children at risk for reading difficulties after 1 year of formal reading instruction, avoiding a wait-and-see policy, to address the importance for young students to be included in intensified training of functional literacy skills in early grades (cf. Al Otaiba et al., 2014; Vaughn et al., 2010). Third, phonics instruction should emphasize direct correspondences between graphemes and phonemes (Ehri, 2001; Fletcher et al., 2021). This was a principle strictly employed in our intervention program. By modeling content, a clear sequence and progression in exercises, explicit use of letters while blending and segmenting, immediate feedback, and ample opportunities to repeat words embedded in each lesson. In all, the chosen intensity aspects of individual tutoring, high standardization of content, and delivery and the frequency of lessons seem to have benefitted the program.

4.4 | Interventions in transparent orthographies

Our program, above targeting phonics, shared some salient instructional features with previous RCTs in the Nordic semi-transparent to fully transparent orthographies of Swedish, Norwegian, and Finnish (Saine et al., 2011; Solheim...
et al., 2018; Wolff, 2011), namely an explicit and intensive, phonics approach. However, the current intervention differed on some important points. We implemented reading training in second grade, whereas others (Saine et al., 2011; Solheim et al., 2018) targeted pre-reading skills along with reading in their first-grade interventions. Consequently, we also identified at-risk students differently from these authors who employed reading precursors as identifying measures where we used manifested reading and spelling levels found to be the more reliable sources of a child’s reading risk status (Olson et al., 2014; Poulsen et al., 2017). The choice to include children after nearly 1 year of formal literacy teaching made it possible to repeat pre-test measures as post-test measures contrary to Saine et al. (2011) and Solheim et al. (2018). Our results indicate the importance of immediate intensified training when students lag behind their peers after initial literacy instruction. While our program was delivered by teachers, Saine et al. (2011) provided a fully computerized training in one intervention condition, and Solheim and her team (Solheim et al., 2018) used a computerized component as part of daily sessions in two of their intervention groups. Clearly, instructional means and mode differed somewhat. Furthermore, this intervention more distinctly singled out phonics as the active ingredient, including repeated word and sentence fluency training, as the program did not contain free spelling, shared reading, vocabulary or pronounced reading comprehension segments found in other programs (Saine et al., 2011; Solheim et al., 2018; Wolff, 2011). Notably, our program found medium to large effects of training after 21 h with approximately the equivalent weekly intensity compared with the 40–75 h of training in the previous program evaluations (Saine et al., 2011; Solheim et al., 2018; Wolff, 2011).

4.5 | Limitations

Students’ motivation to participate in an intensive skill-based intervention and teachers’ feelings about adapting their teaching to a one-to-one format were possible obstacles in initiating this study. However, only one child dropped out from the intervention, and teacher reports of students’ experiences were overall positive. The lack of insights into the control group’s reading activities limits our understanding of possible differences in the amount and quality of teaching between control and intervention group participants (Cheung & Slavin, 2016). Preferably the intervention should have been monitored by video-recordings of lessons as initially planned; however, this fidelity measure was ultimately not possible to implement during the Covid-19 pandemic. Finally, our results are limited to positive immediate post-intervention outcomes.

5 | Conclusions

Our early reading intervention study shows improvements in word decoding and word recognition measured by independent standardized tests and a convincing transfer effect to reading comprehension. By implementing high experimental standards, intensity, and fidelity, our results suggest that a one-to-one phonics approach addressing children at risk for early reading difficulties is successful and works for semi-transparent European orthographies. Our and many other researchers’ approach to applying RCT-designs evaluating teaching methods are important and theoretically well-motivated but inevitably limited for two main reasons: can we argue that schools should provide one-to-one teaching, and can we persuade teachers to provide word-by-word lessons page-by-page regularly on a weekly basis?

Acknowledgements

The project was funded by the Swedish Research Council (2019-03873) and received ethical approval by the Swedish National Research Ethics Committee (ref. no 2019-04084). The authors wish to thank all participating schools, students, and teachers for participating in this project.
CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Data from the extended testing before and after the intervention are available at https://osf.io/p94bc/.

ORCID
Hanna Lindström-Sandahl https://orcid.org/0000-0001-7495-1384
Åsa Elwér https://orcid.org/0000-0001-7824-1750
Stefan Samuelsson https://orcid.org/0000-0002-5234-5495
Henrik Danielsson https://orcid.org/0000-0002-0446-0827

REFERENCES
LINDSTRÖM-SANDAHL ET AL.


### APPENDIX

#### TABLE A1  Lesson example.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Content</th>
<th>Skills</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start up</td>
<td>Introduction of lesson</td>
<td></td>
<td>1 min</td>
</tr>
<tr>
<td>Repetition</td>
<td>Repetition of selected material from previous lesson</td>
<td></td>
<td>2 min</td>
</tr>
<tr>
<td>Segmenting and blending</td>
<td>Listening exercises, reading, and writing words</td>
<td>Phonemic awareness, phonics</td>
<td>3 min</td>
</tr>
<tr>
<td>Story listening and conversation</td>
<td>Listening to a short episode of a story and discussing the content, presentation of today's central content (GPC and/or reading strategy)</td>
<td>Text comprehension, vocabulary</td>
<td>4 min</td>
</tr>
<tr>
<td>Grapheme–phoneme work</td>
<td>Listening to words and finding the correct position of a phoneme, writing, and reading words with practiced GPCs</td>
<td>Phoneme identification, position identification, sound-blending</td>
<td>5 min</td>
</tr>
<tr>
<td>Dictation</td>
<td>Writing words with practiced GPCs, read the words aloud</td>
<td>Phoneme analysis, encoding</td>
<td>4 min</td>
</tr>
<tr>
<td>Words lists and/or sentences</td>
<td>Repeated reading of word lists and/or sentences using practiced words, sounding-out strategy</td>
<td>Fluency</td>
<td>6 min</td>
</tr>
<tr>
<td>Games</td>
<td>Playing simple games to promote fluent reading highlighting the central content</td>
<td>Fluency, repetition</td>
<td>8 min</td>
</tr>
<tr>
<td>Summary</td>
<td>Repetition of central content from today's lesson, marking practiced letters in a schedule</td>
<td>Repetition</td>
<td>2 min</td>
</tr>
<tr>
<td>Speed reading</td>
<td>Reading of a word list</td>
<td>Fluency</td>
<td>1 min</td>
</tr>
</tbody>
</table>

Abbreviation: GPC, grapheme–phoneme correspondence.