American Psychological Association

Journal of Educational Psychology

ISSN: 0022-0663

2021, Vol. 113, No. 2, 248–267
http://dx.doi.org/10.1037/edu0000489

Prevention of Reading Difficulties in Children With and Without Familial Risk: Short- and Long-Term Effects of an Early Intervention

Haytske Zijlstra
University of Amsterdam

Elsje van Bergen
Vrije Universiteit Amsterdam

Anne Regtvoort, Peter F. de Jong, and Aryan van der Leij
University of Amsterdam

In a randomized-controlled trial we tested a computer-assisted intervention for the prevention of reading difficulties, delivered by nonprofessional tutors, running from kindergarten to halfway Grade 2. The full sample included 123 prereaders ($M = 5$; 6 years; 56 intervention; 67 controls) with low preliteracy skills. Parents were sent a questionnaire to assess family risk (FR) for reading difficulties. There was no intervention effect in the full sample, but, unexpectedly, the effect differed between subsamples that did and did not return the questionnaire. The intervention did not affect reading acquisition in the subsample ($N = 49$) without FR-data, mostly children from immigrant, non-Dutch speaking, low-socioeconomic status (SES) families, but had large effects in the subsample of Dutch-speaking, middle and high SES-parents with FR-data ($N = 74$). The latter subsample was followed until Grade 6, 4 years after the intervention, and included 36 intervention children and 38 controls. Long-lasting improvements were found in word-reading fluency, which transferred to reading fluency for pseudowords, English words: and texts, and to spelling. The intervention substantially reduced the need for remedial teaching and grade retention. On all measures, children with FR performed worse than children without FR. The intervention had similar effects on the progress of both groups, but the FR children needed more sessions. This study shows that a 2-year cost-effective early intervention can reduce the incidence of reading difficulties. However, it remains a challenge to make the intervention suited for children in which a lack of preliteracy skills merely seems to reflect a lack of learning opportunities.

Educational Impact and Implications Statement

Is it possible to prevent reading difficulties? Children with poor preliteracy skills in Kindergarten received a computer-based intervention delivered by parents and volunteers, ~3x/week, 10–15 min/session, for 2 years (kindergarten to second Grade). Children from immigrant, non-Dutch speaking, low-SES families did not benefit from the intervention. In the other children, the intervention lastingly improved reading and spelling up till sixth grade, and reduced grade retention. These findings indicate that providing early intensive supplementary reading intervention is a cost-effective way to prevent reading difficulties. However, the intervention needs to be improved to make it successful for children in which a lack of preliteracy skills reflects a lack of learning opportunities.

Keywords: intervention, randomized-controlled trial, reading, poor preliteracy skills, familial risk of reading difficulties

This article was published Online First July 13, 2020.

Haytske Zijlstra, Research Institute of Child Development and Education, University of Amsterdam; Elsje van Bergen, Biological Psychology and Research Institute LEARN! Vrije Universiteit Amsterdam, the Netherlands; Anne Regtvoort, Peter F. de Jong, and Aryan van der Leij, Research Institute of Child Development and Education, University of Amsterdam, The Netherlands.

The study has been funded by the Ministry of Education, Culture and Science, project ODB08052 to Aryan van der Leij. Elsje van Bergen was supported by an NWO Rubicon Fellowship (Grant 446-12-005) and an NWO VENI Fellowship (Grant 451-15-017). After the intervention was finished, the online program Bouw! has been adopted by Lexima, a company for educational software. Since then, the author Aryan van der Leij receives a small percentage of the revenues. Peter F. de Jong and Aryan van der Leij share the last authorship.

Correspondence concerning this article should be addressed to Peter F. de Jong, Research Institute of Child Development and Education, University of Amsterdam, P.O. Box 15780, 1001 NG Amsterdam, The Netherlands. E-mail: P.F.deJong@uva.nl

About 10–15% of children suffer from difficulties in the acquisition of reading accuracy and fluency at the word level (Fluss et al., 2009; Katusic, Colligan, Barbaresi, Schaid, & Jacobsen, 2001). Such reading difficulties increase the chance of academic failure, decreased motivation, lowered self-esteem, grade retention, and school dropout (Poskiparta, Niemi, Lepola, Ahtola, & Laine, 2003; Stanovich, 1986). Therefore interventions that prevent word reading difficulties are of great importance. In the present study we examined the effects of a two year computer-based prevention program starting in kindergarten for children at risk for reading difficulties.

Current evidence suggests that an early start of the intervention increases the chance of preventing reading difficulties in comparison to a later start (Connor et al., 2013; Lovett et al., 2017; see also Torgesen, 2005). However, focusing solely on the period before the start of formal reading in first grade may not be sufficient. Meta-analyses indicate that kindergarten interventions on average have moderate effects (Bus & van IJzendoorn, 1999; Ehri et al., 2001; Suggate, 2010). The effects, however, vary among studies and tend to fade out. Moreover, interventions reveal only small or no transfer effects to later reading, spelling, or reading comprehension (e.g., Byrne, Fielding-Barnsley, & Ashley, 2000; Foorman, Breier, & Fletcher, 2003; Schneider, Roth, & Ennemoser, 2000; Suggate, 2016; van der Leij, 2013). In addition, reading intervention studies typically only consider short-term effects. According to a meta-analysis by Suggate (2016), the mean time between posttest and follow-up was less than 12 months and the majority of follow-up tests were within 2 years. An exception is the study of Blachman et al. (2014), but this study targeted second and third grade poor readers. For interventions in kindergarten, follow-ups of 1 to 2 years after completion imply that children's reading development was usually followed no longer than until the end of second grade. However, a fair amount of reading difficulties emerge after second grade (Catts, Compton, Tomblin, & Bridges, 2012; Torppa, Eklund, van Bergen, & Lyytinen, 2015). In all, very little is known about the effects in the longer run of interventions starting in kindergarten, aiming at the prevention of reading difficulties.

A focus on the precursors or foundational skills of reading, that is phonological awareness and letter knowledge, might not be sufficient to prevent reading problems in the long run. Precursors do not fully account for the development of word reading (e.g., de Jong & van der Leij, 1999; Lervåg, Bråten, & Hulme, 2009). Moreover, behavior genetic studies suggest that the genes involved in the precursors of reading only partly overlap with the genes underlying word reading ability (Byrne et al., 2009, 2006).

Interventions that continue after kindergarten and also involve word reading itself, might produce larger and sustainable effects, which also transfer to skills related to word reading, such as spelling and reading comprehension. Usually, classroom reading instruction aligns with the development of reading of the average child shifting from a focus on foundational skills in kindergarten to cracking the alphabetic code and the accurate decoding of single words in first grade toward the fluent reading of word lists and connected text (e.g., Altani, Protopapas, Katopodi, & Georgiou, 2020; Castles, Rastle, & Nation, 2018). A change in the focus of instruction usually puts extra demands on children, increasing the risk that children lag behind. At such a moment in time, it is probably very important that the intervention continues thereby providing extra support in a relatively difficult period (Bailey, Duncan, Odgers, & Yu, 2017). Longer interventions may also be necessary to equip children with the building blocks necessary for their further independent development of reading (Bailey et al., 2017). Obviously one important building block is letter-sound knowledge. But another indispensable building block is the ability to decode words independently, a self-teaching mechanism enabling the acquisition of orthographic knowledge (e.g., Castles et al., 2018; Share, 2008). It may be argued that interventions for the prevention of reading problems should at least continue to the time that reading speed starts to accelerate. A related reason to prolong the duration of interventions beyond kindergarten is that positive effects often tend to fade out because children are inclined to revert to their original developmental trajectory (Bailey et al., 2016). The risk for such a change might decrease when the reading ability of children has reached a level at which they are able to improve their reading independently.

About 70% of the kindergarten studies enrolled children at risk for reading difficulties (see Suggate, 2016). Risk status before the start of reading instruction is usually based on the cognitive precursors of reading (e.g., Catts, 2017), but not on more fundamental preexisting differences within the risk group. Because preliteracy skills are roughly 50/50 influenced by environmental and genetic factors (e.g., Olson, Keenan, Byrne, & Samuelsson, 2014), children may exhibit a cognitive risk profile for different reasons. Poor preliteracy skills may stem from environmental factors such as low home literacy related to educational level of the parents (Fluss et al., 2009; Mol & Bus, 2011). For example, children from lower socioeconomic status (SES) families tend to have lower phonological awareness than children from higher SES families, probably because of a lower exposure to print activities (e.g., Lonigan, Burgess, Anthony, & Barker, 1998). If early, targeted and intensive intervention at school does not compensate for these disadvantages, these children are most likely to develop reading failure (Torgesen et al., 1999). Alternatively, children may seek and profit less from code related activities (Sénéchal, Whissel, & Bildfell, 2017) due to a genetic predisposition, that is, low sensitivity for learning the alphabetic principle. Familial risk (FR) for reading difficulties is a good proxy for such a disposition as familial transmission of reading skills is largely genetic (Swagerman et al., 2017; van Bergen, Bishop, van Zuijen, & de Jong, 2015; Wadsworth, Corley, Hewitt, Plomin, & DeFries, 2002). FR children are more likely to lag behind on preliteracy skills (e.g., Snowling & Melby-Lervåg, 2016; van Viersen, et al., 2018). With regard to preexisting differences, it may be hypothesized that, even when an intensive and long-term intervention is provided, it is harder for children with a familial risk for reading difficulties to enhance fluent reading skills in comparison to children who suffer from environmental deprivation.

A number of studies trained kindergartners with an FR (e.g., Elbro & Petersen, 2004; Hindson et al., 2005; Regtvoort & van der Leij, 2007; van Otterloo & van der Leij, 2009). Overall, prereading interventions with FR children were effective after finishing the intervention in kindergarten, but did not result in better reading in first grade. Compared with typically developing kindergartners, Hindson et al. (2005) found that FR kindergartners needed more sessions to finish the intervention program, indicating lower response to training. Only one kindergarten intervention study focusing on FR children reported long-term effects, with trained FR children performing better than untrained FR but worse than typically developing children on reading in second grade (Elbro & Petersen, 2004). Still, the gap with the typically developing children tended to increase with age. The question whether intervention effects are smaller for FR children, or put differently, whether they are less susceptible to intervention than children without FR, cannot be answered because these studies included only FR children, rendering a direct comparison impossible.

In the current study we evaluated a 2-year intervention, Bouw! (Build!), starting in kindergarten. The intervention encompassed the foundations of reading, phonological awareness and letter knowledge, followed by word decoding, and gradually progressed toward the fluency of reading mono- and multisyllabic isolated words and lists of words. The intervention was delivered individually and computer-based. An individualized delivery mode is especially recommended for children who lag behind and are very dependent on repetition, direct feedback, and positive reinforcement (Al Otaiba et al., 2014). However, the children did not work on their own, but each child was guided by a tutor. Although the advantages of a child working independently with a computer-based intervention have been advocated (e.g., by Saine, Lerkkanen, Ahonen, Tolvanen, & Lyytinen, 2011), a personal tutor can provide something extra. A tutor is capable of adapting the instruction and respond to a child's needs, in particular with respect to encouragement and motivation (e.g., Azevedo, Moos, Greene, Winters, & Cromley, 2008). As the principles of instruction and learning are included in the program, the tutor does not have to be a professional teacher. The tutor is guided by directions presented on the screen. Volunteers or older children at school, as well as parents at home, can act as a tutor.

In an earlier study, we conducted an randomized-controlled-trial (RCT) with a primary-school version of the program, encompassing first grade to halfway second grade (Regtvoort, Zijlstra, & van der Leij, 2013). Results showed that volunteers, parents, educational assistants and older children were able to deliver the program, supplementary to classroom instruction. The intervention successfully improved the reading development of children with a cognitive risk profile (i.e., poor letter knowledge and phonological awareness), including fluency of word reading and transfer to fluency of pseudoword and text reading, and reading comprehension. The gain was maintained in third grade, 1 year after finishing the program. However, completion of the program proved to be essential to produce effects (see for a similar finding Elbaum, Vaughn, Tejero Hughes, & Watson Moody, 2000).

The present study had a number of novel features. The duration of the intervention was extended to 2 years, starting in the second half of the last kindergarten year until halfway second grade. Participants were children at risk for reading difficulties. We assessed long-term effects of the intervention in a follow-up measurement at the end of second grade (half a year after the intervention). Our main question was whether the children in the intervention group would read more fluently at posttest and follow-up than the group of children that only followed the regular (pre)reading instruction during kindergarten through second grade. We also examined the effects of the intervention on the amount of remedial teaching for reading in and outside the classroom provided by the school, and the transfer to text reading fluency at the end Grade 2 follow-up.

For a subsample of the children, we also had information about their family risk status for reading problems, the FR-information subsample. In this subsample, we were especially interested in the differential effects of the intervention for children with or without a family risk for reading difficulties. Moreover, in this subsample the effects of the intervention were examined in more detail. The children were further followed and assessed mid-Grade 3 (1 year after the intervention) and mid-Grade 6, that is 4 years after completing the intervention. We also investigated the effect of the intervention on the proportion of children with reading difficulties, as well as transfer effects to text reading, spelling, word reading ability in English, and reading comprehension. Finally, we examined in this subsample the impact on noncognitive outcomes, in particular the amount of remedial teaching for reading in and outside the classroom, and grade retention.

Method

Participants

Children came from 48 classes of 13 schools in the area of Amsterdam. We selected children at risk for reading problems in December of the final kindergarten year at approximately the age of 5; 6 years old. In the Netherlands, formal reading instruction starts in first grade. Figure 1 shows the selection procedure and the assignment to the groups. First, 362 children were selected who, according to their teacher, belonged to the bottom 50% on preliteracy skills. Next, these below average children were tested on productive letter knowledge and phonological awareness (see Instruments section). A child was considered to have poor literacy skill if its score on letter knowledge and/or phonological awareness was equal to or lower than the 25th percentile of the full sample. From the children denoted by their teacher as below average preliteracy skills 162 fulfilled these criteria of poor preliteracy skills. All children with low productive letter knowledge had also low receptive letter knowledge and 75% of the selected children scored poor on *both* phonological awareness and letter knowledge.

Selection took place halfway the final year of kindergarten (i.e., half a year before the start of formal reading instruction). At the time of selection the children included in the study could on average only name three letters correctly. Their mean score on phonological awareness was poor, corresponding to a standard score of eight (<25th percentile). For means of comparison, we also tested three to five children per class that according to their teachers belonged to the top 50% on preliteracy skills in their class ($N = 121$). As expected the children with poor preliteracy skills included in the current study performed lower than these children with (above) average preliteracy skills on phonological awareness (mean of 12.29; $SD = 6.89$ vs. 26.74; $SD = 9.72$, $t(280) = -14.595$, $p < .001$) and on productive letter knowledge (3.94; $SD = 4.23$ vs. 16.75; $SD = 8.01$, $t(280) = -17.338$, $p < .001$). Also on another precursor of reading, rapid naming of objects, the children with poor literacy skills lagged behind the above-average children (93.15; $SD = 32.20$ vs. 77.20; $SD = 26.78$, $t(261) = 4.425$, $p < .01$.

The 162 children were randomly assigned (per school and classroom) to the intervention and control condition. Of the 162 children, 39 children (25%) dropped out of the study, 25 children

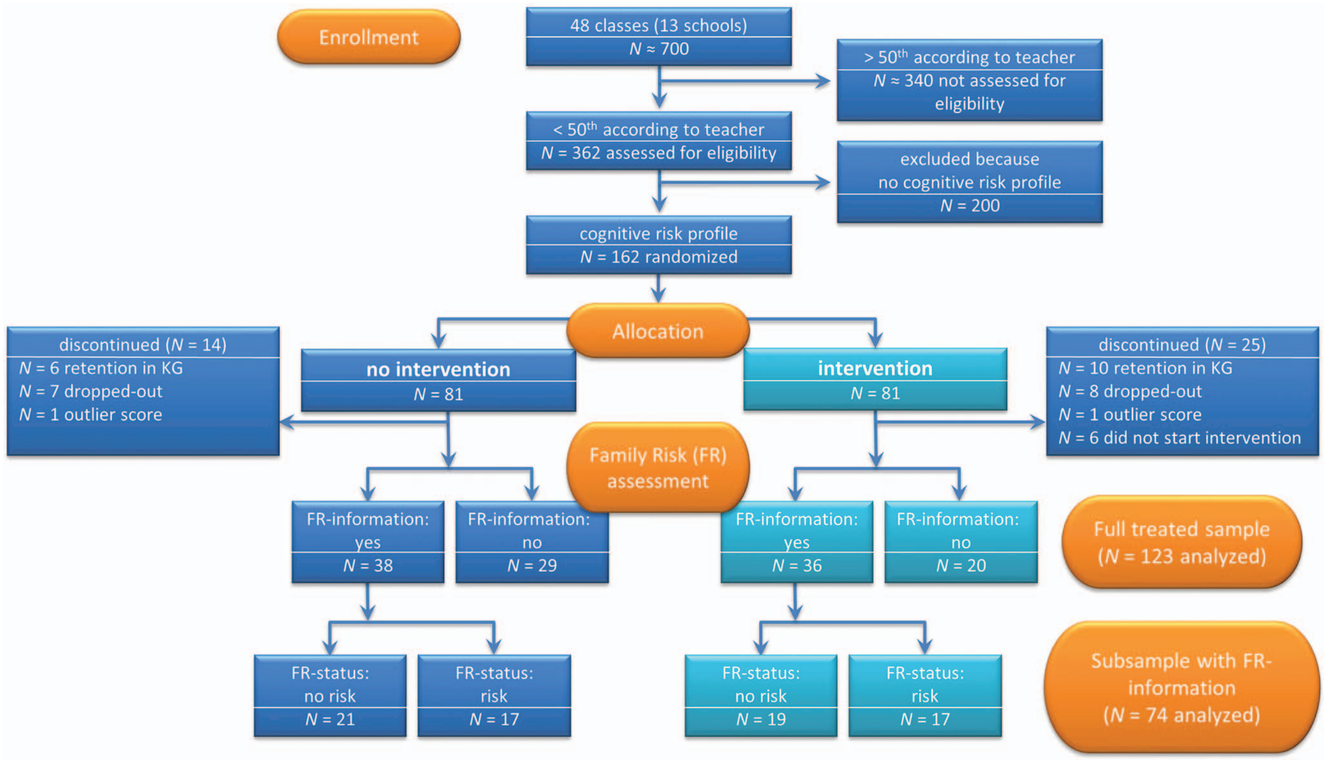

Figure 1. Diagram showing flow of participants through the study. See the online article for the color version of this figure.

in the intervention and 14 children in the control condition. Of these 39 children 16 were omitted because they repeated kindergarten (10 intervention children and six controls). Another 15 children (eight intervention children and seven controls) dropped out for various reasons: Six children discontinued because they moved to another school outside the area of Amsterdam (four intervention children and two controls), three children withdrew for medical reasons (one intervention child, two controls), six parents withdrew their consent after the selection was made (three intervention children and three controls), and two children (one intervention and one control) were omitted because of outlier scores on most outcome variables (larger than 3 *SD* from the mean). Finally, six children did not start with the intervention (four did not start at all and two stopped within eight sessions) and did not have scores on the outcome variables. In all, the full sample comprised 123 children with poor preliteracy skills in kindergarten (mean age of 5;8 years; $SD = 4.65$ months; 63% boys). Of these children, 56 were enrolled in the intervention program and 67 in the no-intervention control group.

Chi-square tests did not show any differences between the group of children that remained in the study and the group that dropped out with respect to percentages of boys, use of minority language at home or non-Dutch ethnic origin. *T* tests showed that the groups did not differ in active and passive letter knowledge, but the group who dropped out of the study was on average 3 months younger, $t(159) = -4.076$, $p < .001$, and performed lower on the phonological awareness test, $t(159) = -2.578$, $p = .011$. Differences between the groups on vocabulary, nonverbal intelligence, rapid naming and SES could not be examined as this information was mostly gathered during the study at a time that most of the children who dropped out had already stopped to participate in the study.

At the beginning of first grade, the child's FR-status was determined. We assessed risk by parent's self-report of literacy difficulties (e.g., Snowling, Dawes, Nash, & Hulme, 2012; van Bergen, de Jong, Maassen, & van der Leij, 2014). Of the 123 children, the FR-status of 49 children could not be determined because parents did not return the questionnaire. Many of these children came from immigrant families that did not speak Dutch at home and/or from disadvantaged families, in which the parents had a low educational level (e.g., primary school only). It was very difficult to examine FR for reading problems in these families.

In Table 1 descriptive information about the full sample broken down by intervention (yes or no) and FR-risk information (yes or no) is given. Family SES was based on the mean educational level of the parents on a scale from *only primary school* (1) to *university degree* (5). The selection and control measures are described below (see section on Measures). The background characteristics of the intervention and the control condition were very similar. Differences in percentages of boys, non-Dutch ethnic origin and non-Dutch home language between the intervention and control condition were negligible. *T* tests indicated that differences between the conditions on the selection and the control variables were not significant. However, the groups of children with and without information about FR differed considerably. The percentage of boys (70%) was higher in the group with FR-information than in the group without information (50%), $\chi^2(1) = 4.666$, $p =$

Table 1
Full Sample (N = 123): Characteristic of the Four Groups (Intervention Condition by FR-Information)

Measure	No intervention		Intervention	
	No FR-info	FR-info	No FR-info	FR-info
Background information				
Age in years (selection)	5.63 (0.36)	5.63 (0.42)	5.70 (0.49)	5.67 (0.31)
Sex (% boys)	59%	66%	40%	75%
Ethnicity (Dutch[a])	3%	66%	5%	61%
Home language Dutch (%)	34%	100%	40%	100%
Family SES	2.18 (1.16)	3.35 (0.98)	2.17 (0.87)	3.46 (1.12)
Selection measures				
Phonological awareness	11.31 (6.52)	15.08 (6.59)	10.50 (8.15)	13.75 (6.77)
Receptive letter knowledge	8.52 (4.96)	8.05 (3.53)	9.30 (5.16)	8.42 (4.51)
Productive letter knowledge	4.48 (4.27)	3.37 (2.57)	5.45 (5.63)	3.44 (4.07)
Control measures				
Receptive vocabulary	56.72 (15.17)	71.32 (8.39)	50.05 (11.51)	67.11 (11.53)
Nonverbal IQ (stand. score)	5.47 (1.98)	5.86 (1.77)	5.25 (2.14)	6.29 (1.78)
Rapid automatized naming	100.79 (45.23)	86.92 (24.52)	91.10 (23.71)	86.56 (21.71)

Note. Standard deviations are in parentheses.

.031, and clearly the percentage of families of non-Dutch origin (96%) who did not use the Dutch language at home (60–65%), was far larger in the No FR-information group. Furthermore, *t* test showed that children in the no FR-information group came from families with a considerably lower SES, $t(115) = 6.133$, $p < .001$. In the no FR-information group 60% did not continue their education after high school whereas this was only 12% in the FR-information group. The children in the no FR-information had less vocabulary knowledge, $t(122) = 6.773$, $p < .001$, and a somewhat lower phonological awareness, $t(121) = 2.731$, $p < .01$, and nonverbal intelligence, $t(122) = 1.996$, $p = .048$. Surprisingly, the active letter knowledge of the no FR-information group was higher, $t(121) = 1.992$, $p = .049$.

We computed the power to obtain an effect of the intervention. Sensitivity analyses were conducted in G*Power (Version 3.1.9.4; Faul, Erdfelder, Lang, & Buchner, 2007) to calculate the minimum detectable effects (or: required effect size), given a power (1-β) of 0.80, an α-level of .05, and our obtained sample sizes. For the full sample ($N = 123$), we were interested in the minimum detectable effects for two main effects and their interaction effect. The two between-subjects factors were Intervention (i.e., yes or no) and FR-information (i.e., whether information on FR-status could be obtained). We used G*Power's "ANOVA: Fixed effects, special, main effects and interactions," with for all three effects: four groups, and $df_{\text{numerator}} = 1$ (i.e., for a main effect $df_{\text{numerator}} = [2 \text{ levels} - 1] = 1$ and for the interaction $df_{\text{numerator}} = [2 \text{ levels} - 1] * [2 \text{ levels} - 1] = 1$). This yielded for all three effects a minimum detectable effect size *f* of 0.25. For the main effects, this equals a minimum detectable difference between the two groups of 0.50 *SD* (i.e., a medium Cohen's *d*; Cohen, 1988). We gained some power by having repeated measures. For the main effects, the minimum detectable effect size *f* dropped to 0.22, which equals a difference between the two groups of 0.45 *SD* (using G*Power's "ANOVA: Repeated measures, between factor"; two groups; four measurements; correlated with .70).

The effects of FR could only be examined in the subsample in which information on FR-status is available. We will refer to this subsample as the FR-information subsample. The subsample comprised 74 children with poor preliteracy skills in kindergarten (mean age of 5;8 years; $SD = 4.44$ months; 70% boys). Of these children, 36 were enrolled in the intervention program and 38 in the no-intervention control group.

FR-status was based on parent's self-reported literacy problems with two of the three questions used by van Bergen, de Jong, Maassen, and van der Leij (2014): (a) Do you think you are a fast, average or slow reader? and (b) Do you think you have more, average, or less difficulties with spelling than other people? This self-report measure correlates .84 with a composite of word- and pseudoword-reading fluency (van Bergen et al., 2014), showing that it is a valid proxy for identifying adults with reading difficulties. A child was considered as at FR for reading difficulties if at least one of the parents indicated to be a slow reader (instead of average or good) and/or have more spelling difficulties than other people. It is noteworthy that (a) just over half of the FR parents reported to have a first-degree family member with reading difficulties (i.e., parent's sibling or parent) whereas the group without FR did not mention any reading difficulties within the family; and (b) about 50% of our selection of children with poor preliteracy skills had an FR for reading difficulties. In all, of the 36 children in the intervention group there were 17 with and 19 without FR. In the no-intervention control group, 17 did have and 21 did not have FR.

With respect to the power, for the FR-information subsample ($N = 74$), we were interested in the minimum detectable effects for two main effects and their interaction effect. Apart from intervention (yes/no), the second between-subjects factor here was FR-status (i.e., yes/no family history of reading difficulties). Using the same method as above yielded for all three effects a minimum detectable effect size *f* of 0.33, or for the main effects, a difference between the two groups of 0.66 *SD* (i.e., a medium to large Cohen's *d*; Cohen, 1988). For the main effects we could also calculate the minimum detectable effects with repeated measures. Mirroring the above, the effect size *f* dropped to 0.29, or a group difference of 0.58 *SD*

Descriptive information about the four groups is presented in Table 2. As expected, the children without FR came from lower SES families than the children with FR, $t(72) = 2.192$, $p = .032$. More than half of the children had native Dutch parents. Children from native Dutch parents were slightly overrepresented in the FR-groups, $\chi^2(1,74) = 4.557$, $p = .033$. No other differences between the FR-groups nor between the intervention groups were significant (all $ps > .10$). The mean scores for receptive vocabulary indicated average ability.

Design

Children were randomly assigned to the intervention program Bouw! (Build!) or a control group. Schools were free to provide extra reading instruction to all children, according to their school policy. In the full sample within the intervention and control groups a distinction was made between children with and without information about family risk (FR) for reading problems. In the FR-info sample we distinguished children with and without family risk (FR) for reading problems.

Figure 2 shows the intervention phases and the occasions of measurement. The intervention lasted 2 years and included (a) the prereading phase (the final 18 weeks in kindergarten); (b) the beginning reading phase (28 weeks in Grade 1); and (c) the more advanced reading phase (the first 16 weeks of Grade 2). In the full sample, children were tested approximately each half a year during the intervention period: end of kindergarten (end KG), start (October/November) Grade 1, mid (February), and end (June) Grade 1, followed by the posttest mid-Grade 2 (post)-test. A follow-up was conducted half a year after the posttest (end of Grade 2, first follow-up). Two further follow-ups were done only in the FR-information subsample: 1 year (mid-Grade 3, second follow-up) and 4 years (mid-Grade 6, third follow-up) after the posttest.

Intervention

A detailed description of the intervention program Bouw! (Build!) is given by Regtvoort, Zijlstra, and van der Leij (2013). In short, the intervention is delivered via a computer program and the assistance of a tutor. The program covers the preliteracy skills (kindergarten), beginning reading (decoding in first grade), and reading fluency (end of first grade and first half of second grade). The exercises on reading address all the orthographic complexities of Dutch (single letters and digraphs, letter combinations, double consonants, open and closed syllables; see Van der Leij & Van Daal, 1999).

The 2-year program consisted of 12 modules (523 lessons in total, 270 were main lessons and 253 additional lessons for children who need more exercises). One of the main principles of the response-to-intervention approach is to evaluate children's individual progress and adapt accordingly (Greulich et al., 2014; Gersten et al., 2009). Following this principle, after each module a test was given. The child was promoted to the next module if at least 80% of the items were correct. If the test was not passed, the child had to repeat the main lessons of the module. In this way, the content of the program was adapted to the level of the child.

Children practiced with the computer program two to four times a week in sessions of 15 min. On average, children were expected to do 186 sessions during the intervention (62 weeks). In one session several lessons were done. The number of lessons per 15-min session was allowed to vary. The number of lessons completed per session ranged from one to four. Depending on how well a child responded to the intervention (i.e., how soon the child was promoted to the next module), the child could go faster through the program material. However, none of the children were allowed to finish the program before the end of Grade 1. By design, all children had to finish no later than halfway Grade 2.

During practice with the program, children were assisted by nonprofessional tutors (parents, older students, volunteers) who provided feedback and social-emotional support. Instructions and feedback for the tutor were provided in the left part of the screen. In kindergarten, 78% of the intervention children were tutored by one of their parents and 22% by a volunteer at school because their parents were not able to act as tutor. In Grade 1 and 2, 86% of the

Table 2
FR-Information Subsample ($N = 74$): Characteristic of the Four Groups (Intervention Condition × FR Risk-Status)

Measure	No intervention		Intervention	
	No FR	FR	No FR	FR
Background information				
Age in years (selection)	5.64 (0.48)	5.61 (0.33)	5.58 (0.31)	5.78 (0.29)
Sex (% boys)	76%	53%	68%	82%
Ethnicity (Dutch[a])	57%	77%	47%	77%
Home language Dutch (%)	100%	100%	100%	100%
Family SES	3.31 (1.10)	3.40 (0.83)	3.00 (0.97)	3.97 (1.08)
Selection measures				
Phonological awareness	15.33 (7.19)	14.76 (5.97)	13.84 (6.53)	13.65 (7.24)
Receptive letter knowledge	7.29 (3.45)	9.00 (3.50)	8.42 (4.13)	8.41 (5.04)
Productive letter knowledge	2.90 (2.51)	3.99 (2.61)	3.26 (2.90)	3.65 (5.17)
Control measures				
Receptive vocabulary	69.76 (8.23)	71.00 (8.78)	64.84 (8.94)	69.65 (13.71)
Nonverbal IQ (stand. score)	5.73 (1.77)	6.02 (1.82)	6.31 (1.77)	6.27 (1.83)
Rapid automatized naming	90.19 (29.81)	82.88 (15.73)	91.00 (26.06)	81.59 (14.76)

Note. Standard deviations are in parentheses.

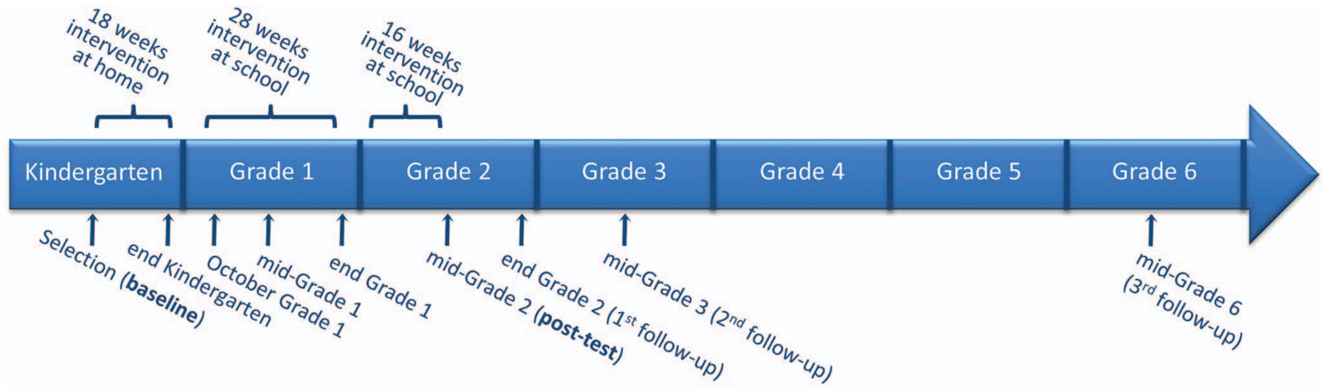

Figure 2. Timeline. The intervention phases are shown at the top and the test occasions at the bottom. See the online article for the color version of this figure.

children were tutored at school by a volunteer, teaching assistant or an older student, and 14% by one of their parents at home.

Intervention Fidelity

Treatment fidelity was promoted through three monthly group meetings of the first author with the special-needs coordinator of the school and a school psychologist. In these meetings intervention reports were handed out with an advice for each child regarding the quantity of sessions and the individual progression (e.g., whether a child had to repeat more lessons or not). The information in the reports was based on the digital log files.

Zijlstra, Koomen, Regtvoort, and van der Leij (2014) reported on the observed quality of instruction and emotional support provided by part of the nonprofessional tutors in the current study. Briefly, Zijlstra et al. (2014) observed 32 tutors and children two times, in kindergarten and in Grade 1. Videotapes were rated by at least two independent raters. The observation scales were an adaptation of the observation instrument of Thijs and Koomen (2008), which is based on the Erickson scales for parent–child interaction (Erickson, Sroufe, & Egeland, 1985). Tutor-support quality was investigated by means of tutor emotional support (i.e., the degree to which the tutor is sensitive to the emotional needs of the child and shows confidence in the child) and tutor adaptive instruction (i.e., to what degree the tutor gives clear instructions and adapts the instruction adequately to the level and needs of the child). On a 7-point scale from very low to very high, it turned out that 87% of the tutors provided sufficient (4) to high levels (6) of support (Zijlstra et al., 2014).

Remedial Teaching In and Outside The Classroom

In the Netherlands most schools have adopted a response-to-intervention model for children who lag behind in reading and spelling. At first, children with reading difficulties get extra classroom instruction in small groups. But if this extra instruction is insufficient, extra help outside the classroom is provided by a specialist in remedial teaching. Special-needs coordinators filled in a questionnaire for each child to indicate whether the child had extra classroom instruction (yes, no) and remedial teaching (yes, no) in the period from Grade 1 and through Grade 2. Children were considered to receive extra classroom instruction and remedial teaching when this type of extra help was provided three or more times per week (≥10 min per session) for at least 6 months.

Instruments

All tests, except for the tests for productive letter knowledge (grapheme test) and word reading accuracy, were norm referenced tests.

Selection Measures

Productive letter knowledge. This was measured with the Grapheme test (Verhoeven, 1993). The child had to name the phonemes corresponding to 34 graphemes, including diagraphs. The maximum score is 34. Cronbach's alpha is .85 (Verhoeven, 2000).

Phonological awareness. We used the phonological awareness subtest of the CELF-4-NL (Clinical Evaluation of Language Fundamentals, Dutch version; Kort, Schittekatte, & Compaan, 2008). This test measures the ability to recognize and manipulate sounds in spoken words. The maximum score is 45. A standard score of 10 reflects the 50th percentile. Cronbach's alpha is .85.

Control Measures

Receptive vocabulary. We administered the TAK (Taaltoets voor Allochtone Kinderen; Verhoeven & Vermeer, 1996) to measure receptive vocabulary knowledge at the end of kindergarten. The child had to choose among four alternatives the picture that best matched a given word. For example "Can you show me the horseshoe?" This test consists of 96 items of increasing difficulty. Administration was stopped when the child failed six out of the last eight items. Cronbach's alpha reliability at the beginning of first grade is .97 (Verhoeven & Vermeer, 1996).

Nonverbal IQ. We used Raven's Colored Progressive Matrices to measure nonverbal IQ. Every item consists of a rectangular pattern in which one part is missing. The child has to look for the missing part from six alternatives. The maximum test score is 36. Cronbach's alpha reliability ranges from 0.80 to 0.90 across languages and age levels (Raven, Raven, & Court, 2004).

Rapid automatized naming. Naming speed for objects was assessed at the time of selection in kindergarten. Each card consisted of a series of 50 symbols. Five objects were repeated randomly 10 times and randomly ordered in five columns of 10 symbols. The score was time (seconds) to completion. According to the manual, the reliability of the test is .81 (van den Bos, Lutje Spelberg, & Eleveld, 2004).

Primary Outcome Measures

The main focus of the intervention was on learning to read words through practice with a large variety of Dutch words. Therefore, the primary outcome of the intervention was word-reading ability. We assessed this ability on each occasion of measurement.

Word-reading accuracy (WRA). At the end of kindergarten and in October of first grade, approximately 8 weeks after the start of reading instruction, children were asked to read five vowel-consonant (VC) and five consonant-vowel-consonant (CVC)-words, such as *in, oom* [uncle], *bus,* and *kar* [cart]. The words were not taught in the intervention program or part of the school curriculum. One point was given if a word was read correctly by sounding out and blending letter sounds and two points were given when the child read a word directly. The maximum score on the test was 20. The test has been successfully used in a previous intervention study (Regtvoort et al., 2013).

Word-reading fluency. This was measured with the Drie Minuten Test (3 minute test, Verhoeven & van Leeuwe, 2003). The tests consists of three cards with lists of words. The score on a card is the number of words read correctly within 1 min. In the current study the second and the third card were used. The second card has a list of 150 monosyllabic words with consonant clusters (CCVC, CVCC, CCVCC, and CCCVC words). The third card comprises 120 polysyllabic words with various orthographic complexities. Reported Cronbach's alpha′s are larger than or equal to 0.95. Word reading fluency of monosyllabic words was administered on the occasions from mid-Grade 1 to mid-Grade 3. Word reading fluency of polysyllabic words was administered from end Grade 1 to mid-Grade 6.

Transfer Measures

We assessed whether the program had effects on outcomes that were not practiced in the program. These involved pseudoword reading, text reading fluency, the reading of English words, spelling, and reading comprehension. The latter three tests were administered only in the FRinfo sample.

Pseudoword reading fluency. We administered the KLEPEL at the first follow-up at the end of Grade 2 (van den Bos, Lutje Spelberg, Scheepstra, & de Vries, 1994). The test consists of a list of 116 pseudowords of increasing difficulty from one to four syllables (e.g., *nim*, *nargel*, and *megoezen*). The child was instructed to read as many pseudowords as possible, with the score being the number correct within 2 min. Reported Cronbach's alpha is 0.95.

Text-reading fluency. This was measured at end Grade 2 with a story text at the level of Grade 2 (AVI; Visser, van Laarhoven, & ter Beek, 1994). The story contains 152 words in total. Children had to read the full text as accurately and quickly as possible. The maximum time was 4 min. The total score was converted to the average number of words accurately read per minute.

English word-reading fluency. In the Netherlands there is a fair amount of exposure to the English language through TV, computer games, and some instruction in school. Therefore we also measured English word-reading fluency mid-Grade 6 with the sight word efficiency subtest (SWE) of the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 2012). The task consists of a list of 108 items, increasing in difficulty. The score was the number of correct words in 1 min (0–108). We used 1 min instead of the prescribed 45 s, because the duration of our Dutch word-reading fluency tests is also 1 min. Strict and clear rules for the correct pronunciation of the words were used (e.g., *find* should be pronounced as /find/ and not as the Dutch word /*vind/*). Dialect was allowed as long as the English word was read correctly (e.g., variations on the word *fast* were allowed, but there was one correct answer for *much*).

Spelling. This was tested end Grade 2 and mid-Grade 6 with word spelling to dictation using the PI-dictee (Geelhoed & Reitsma, 1999). The tested words increase in difficulty. We selected the two blocks of 15 words that are at the Grade 2 level and a selection of seven out of 15 words for each of the blocks suited for Grade 5/6. Accordingly, the maximum score was 30 correct words in Grade 2 and 14 in Grade 6. Reported reliability is above 0.90 (Geelhoed & Reitsma, 1999).

Reading comprehension. This was measured mid-Grade 3 with a norm-referenced test for Grades 1–3 (Aarnoutse & Kapinga, 2006). The test consisted of seven short stories. Each story was followed by several multiple choice questions. In total there were 38 questions. The test was administered to groups of four to five children by a trained assistant. After the instruction, the children had to fill out the test individually. There was no time limit. The maximum score on the test was 38. Reported Cronbach's alpha is 0.93.

Statistical Analyses

The data had a three-level structure: children were nested in classes which were nested in schools. In such data, the scores of children might be dependent, the extent of which is reflected by the intraclass correlation (ICC). If the nested structure is neglected standard errors can be underestimated and alpha levels might be inflated (Snijders & Bosker, 2012). However, in the current study, it was almost impossible to capture the nested structure in our data-analyses because between each occasion of measurement there were children who moved to another school. As a result, we started with 13 schools but at the end of the study 28 schools were involved. However, we did not expect much dependency at the class level, because, with one exception, there were only a few children (one to four) per class. To get an indication of the dependency within schools we did a multilevel analysis on each of the nine measures administered from midway first grade through the end of second grade using the original grouping of the children in the schools. In all of these analyses the between-school variance (Level 2) was nonsignificant. All ICCs were below 0.1 and six out of nine were below 0.05. The results were similar in the analyses with class as the Level 2 variable. These analyses show that between class and school variation is low and insignificant and that there is little need to account for the nested structure of these data. Therefore, the data were analyzed with standard univariate or repeated measures analyses of variance (ANOVA).

Results

The results are presented in three sections. First, we consider intervention fidelity in the full sample and in the subsamples with and without information about the FR-status of the children. Next, we report the results on the effects of the intervention in the full sample. In the last section we present the results on the short-term and long-term effects of family risk for reading difficulties on the outcomes of the intervention. These results are based on the subsample of children whose parents provided the necessary information to determine the risk-status of their children.

Intervention Fidelity

In the full sample ($N = 123$) 66% completed the final part of the program. The percentage of children that completed the program in the subsample with FR-information was higher than in the subsample that did not provide FR-information (72% vs. 55%) but this difference was not significant, $\chi^2(1) = 1.701$, *ns*. As mentioned in the Method section, the content of the program was adapted to the level of the child and lessons were repeated if a child did not reach the level needed to be promoted to the next level. The majority of the children who did not complete the program repeated certain program parts and lessons.

The entire intervention aimed at three sessions per week during 62 weeks, 186 sessions in total. Note that usually several lessons per session were done. The average number of sessions was 110.20 ($SD = 59.55$), implying that 59% of the intended sessions were done. The number of sessions per child ranged from 15 to 369. Two children did more than 250 sessions. Without these children, the percentage of completed sessions dropped from 59% to 55%. On average the children in the FR-information subsample received significantly more sessions than the children in the subsample without FR-information (FR-information: $M = 127.75$, $SD = 63.42$; no FR-information: $M = 78.60$, $SD = 35.01$), $t(54) = 3.197$, $p < .01$ ($d = 1.00$).

Schools were allowed to follow their regular procedures for extra help for children that lagged behind in reading. In Grade 1 similar percentages of the children in the intervention (48.1%) and in the no intervention control group (40.6%) received extra reading instruction in the classroom. However, in Grade 2 significantly more children in the no intervention control group (44.3%) received extra instruction in the classroom than children in the intervention group (25.25%), $\chi^2(1) = 4.263$, $p = .039$. For remedial teaching outside the classroom, there were no differences between the intervention groups, $\chi^2(1) = 1.70$, *ns*, but note that here for about 20% of the children this information was missing. In all, these results suggest that from second grade onward the children in the no intervention control group received somewhat more business-as-usual school services.

Data were checked for missing values and outliers before the analyses. There were no missing scores on any of the measures administered from kindergarten through the posttest halfway second grade. One child in the intervention condition had missing scores at the end of Grade 2 follow-up.

Full Sample: Effects of the Intervention

Our main interest in the full sample concerned the effect of the intervention. However, we also examined whether the intervention effect differed across the subsamples that did or did not provide information about FR-status. Therefore in the full sample we conducted ANOVAs or repeated measures ANOVAs with intervention (yes or no) and FR-information (yes or no) as between-subjects variables.

Main Outcomes: Word-Reading Accuracy and Fluency

The main goal of the intervention concerned the improvement of reading accuracy and fluency. Descriptive statistics for word-reading accuracy and word-reading fluency from end of kindergarten to end Grade 2 are presented in Table 3. In addition, the development of word-reading fluency across groups is depicted in Figure 3. At the end of kindergarten children's scores were at floor; the children could hardly read any words. In October Grade 1, after approximately 8 weeks of reading instruction, there was some increase in word reading accuracy. However, an ANOVA showed no effects of intervention and FR-information at this time.

Next, we conducted repeated measures ANOVA to test for the effects of the intervention on word reading fluency. Separate analyses were done for word reading of monosyllabic and polysyllabic words. The former was measured four times in the period from halfway first grade to the end of second grade, half a year after the intervention had finished. The latter was assessed three times in the period from the end of first grade to the end of second grade.

For reading fluency of monosyllabic words the main effect of intervention was not significant. The effect of FR-information was not significant either, but we did find significant interactions of Occasion $\times$ Intervention, $F(3, 354) = 3.22$, $p = .023$, $\eta_p^2 = .027$, Intervention $\times$ FR-information, $F(1, 118) = 5.96$, $p = .016$ $\eta_p^2 = .048$, and Occasion $\times$ Intervention $\times$ FR-information, $F(3, 354) = 4.02$, $p < .01$, $\eta_p^2 = .033$. Follow-up contrasts showed that the effect of intervention was not significant in the subsample without FR-information, $F < 1$, whereas in the subsample with FR-information the intervention group on average outperformed the no intervention control group, $F(1, 118) = 10.46$, $p < .01$. Further contrasts in the FR-information subsample per occasion of measurement showed that the difference between intervention and control group was not yet significant midway Grade 1, $F < 1$. However, on all subsequent occasions there was a significant difference, *F*s (1,118) of 7.68, 10.30 and 11.61, all with $p < .01$, Cohen's *d*s 0.95, 0.94, and 0.96, respectively, at end Grade 1 to posttest mid-Grade 2.

For the reading fluency of polysyllabic words, the main effect of intervention just missed significance, $F(1, 118) = 3.26$, $p = .073$, $\eta_p^2 = .027$. The main effect of FR-information was not significant. There were significant effects of Occasion $\times$ Intervention, $F(2, 236) = 3.11$, $p = .045$, $\eta_p^2 = .026$, and Intervention $\times$ FR-information, $F(1, 118) = 5.83$, $p = .017$, $\eta_p^2 = .047$. In contrast to monosyllabic words, the three way interaction between occasion, intervention, and FR-information was not significant. But note that polysyllabic word reading fluency was not assessed mid-Grade 1 when the effect of the intervention on monosyllabic word reading fluency could not yet be observed (see Figure 3). Following up the Intervention $\times$ FR-Information interaction showed that, as for

Table 3
Full Sample (N = 123): Means and Standard Deviations for Reading Accuracy and Reading Fluency

	No intervention				Intervention			
	No FR-info		FR-info		No FR-info		FR-info	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Word-reading accuracy								
End kindergarten	2.07	3.21	0.82	2.31	1.16	2.17	2.89	3.79
October Grade 1	7.17	5.41	5.55	3.78	7.90	5.29	6.64	4.44
Reading fluency: Monosyllabic words								
Mid-Grade 1	8.52	6.94	4.50	4.43	8.65	8.17	5.75	5.03
End Grade 1	29.28	19.03	17.87	11.06	26.15	16.96	28.33	17.83
Mid-Grade 2	50.90	25.16	37.89	18.54	49.70	26.08	55.25	24.49
End Grade 2	57.83	25.79	48.21	19.33	55.26	27.72	66.75	22.85
Reading fluency: Polysyllabic words								
End Grade 1	16.66	12.61	8.82	7.34	14.65	10.67	16.44	11.54
Mid-Grade 2	33.17	19.30	23.34	13.58	34.00	19.08	36.56	18.10
End Grade 2	39.76	19.21	32.79	13.73	38.47	19.59	46.28	18.54
Transfer end Grade 2								
Pseudoword reading	38.07	18.99	27.79	11.68	36.21	18.65	39.78	19.55
Text reading fluency	94.35	35.20	86.67	26.17	87.78	36.10	107.41	30.99

monosyllabic words, the difference between intervention and control group was not significant in the subsample without FR-information, $F < 1$, whereas the intervention was significant in the subsample with FR-information. The mean score was significantly higher in the intervention than in the control condition, $F(1, 118) = 11.63$, $p < .01$. Subsequent contrasts confirmed differences in the FR-information subsample on all occasions between intervention and control, with Cohen's *d*s for the intervention effects of 1.04 (end Grade 1), 0.97 (at posttest mid-Grade 2), and 0.98 (at follow-up end Grade 2). This mirrors the intervention effect found for the monosyllabic words.

In all, these analyses indicate that the intervention did not have a significant effect on word reading fluency although the difference between intervention and control group just missed significance for polysyllabic word reading fluency. However, we found a significant difference in the effect of the intervention in the subsamples. In the FR-information subsample performance in reading fluency was significantly higher in the intervention group than in the no intervention control group for the reading of both monosyllabic (top panel of Figure 3, green lines) and polysyllabic words (bottom panel of Figure 3, green lines). Effects of the intervention were observed from the end of first grade until the end of second grade, half a year after the intervention was finished. In contrast, in the subsample without FR-information the differences between the intervention and the no intervention control group were minimal (top and bottom panel of Figure 3, overlapping orange lines).

Short-Term Transfer Effects

We also examined the effects of the intervention on pseudoword and text-reading fluency. An ANOVA with pseudoword reading fluency as outcome variable showed no significant effects of Intervention, $F(1, 118) = 2.47$, *ns*, and FR-information, $F(1, 118) = 1.08$, *ns*, but again revealed a significant Intervention × FR-Information interaction, $F(1, 118) = 4.61$, $p = .034$, $\eta_p^2 = 038$. Subsequent contrasts indicated that the difference between the intervention conditions was not significant in the subsample without FR-information, $F < 1$. In the subsample with FR-information the mean score of the intervention group was significantly higher than in the control group, $F(1, 118) = 9.01$, $p < .01$. Similar results were found for text reading fluency. The main effects of intervention, $F(1, 118) = 1.44$, *ns*, and FR-information, $F(1, 118) = 1.02$, *ns*, were not significant, but the interaction of intervention and FR-information was, $F(1, 118) = 5.32$, $p = .023$, $\eta_p^2 = .043$. Further contrasts indicated that the difference between the intervention conditions was not significant in the subsample without FR-information, $F < 1$, whereas in the subsample with FR-information the mean score of the intervention group was significantly higher, $F(1, 118) = 8.02$, $p < .01$.

Overall, these results show that the intervention had significant effects on word, pseudoword, and text reading fluency, even at follow-up half a year after the intervention had finished, but *only* in the subsample with FR-information. The intervention effect was not significant in the subsample without FR-information. We did two additional analyses to examine the differences in the effects of the intervention between the two subsamples. First, we noted that the two subsamples differed significantly in the number of sessions that were done during the intervention. Therefore, we conducted a repeated measures ANOVA on word reading fluency in the intervention group with FR-information as a between-subjects variable and number of sessions as a covariate. In these analyses we omitted two children from the FR-information sample who had done more than 250 sessions and were considered as outliers (more than 3 *SD* from the mean). Irrespective of the inclusion of these outliers, the most important results of these analyses were as follows. The effects of number of sessions on the monosyllabic and polysyllabic words were significant, $F(1, 50) = 7.93$, $p < .01$, $\eta_p^2 = .137$ for monosyllabic and $F(1, 50) = 6.55$, $p = .014$, $\eta_p^2 = .116$ for polysyllabic words, but the effects of FR-information remained significant, $F(1, 50) = 4.80$, $p = .033$, $\eta_p^2 = .088$ for monosyllabic and $F(1, 50) = 4.59$, $p = .037$, $\eta_p^2 = .084$ for polysyllabic word reading fluency. However, the signif-

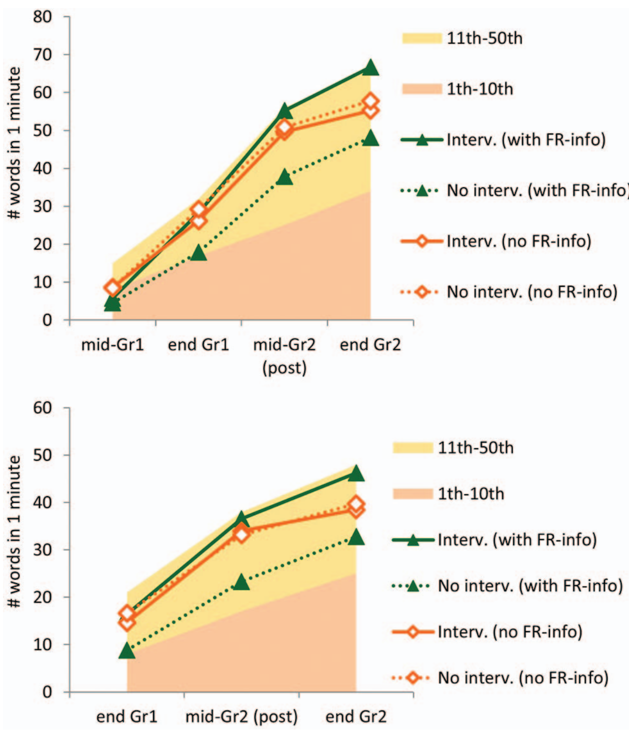

Figure 3. Full sample (N = 123): word-reading fluency development in Grade 1, Grade 2 (posttest) and end Grade 2 (follow-up 1). The score represents the number of words read correctly within 1 min. Monosyllabic word reading (top panel) was assessed from mid-Grade 1 onward; polysyllabic word reading (bottom panel) is more demanding and was therefore assessed from end Grade 1 onward. The solid lines represent the means for the intervention groups, the dotted lines the no-intervention groups, the green triangles the groups with information on FR-status, and the orange diamonds the groups without information on FR-status. For reference, the background shades indicate the percentile scores based on national norms. For example, the intervention group with information on FR-status scores below the 10th percentile mid-Grade 1 (so very poor), but scores around the 50th percentile (so average) after intervention. See the online article for the color version of this figure.

icant effect of the number of sessions indicated that the children who received *more* sessions performed worse in reading. Though this result may seem counterintuitive at first sight, it suggests that the poorer readers received more instruction. We return to this issue in the Discussion section.

Second, we also explored whether differences in the effects of the intervention were related to completion of the full intervention program by doing repeated measures ANOVA on word reading fluency in which we included completion (yes or no) as an additional factor. In both analyses, the positive effects of completion of the program turned out to be large and significant, $F(1,51) = 24.46$, $p < .001$, $\eta_p^2 = .324$ for monosyllabic and $F(1, 51) = 24.94$, $p < .001$, $\eta_p^2 = .328$ for polysyllabic word reading fluency. More importantly, the main effects of FR-information and the interactions of Completion × FR-Information and Occasion × Completion × FR-Information were not longer significant, Fs < 1. Therefore, in additional analyses we compared the intervention group that completed the intervention (completers, $N = 37$) with a group that did not (noncompleters, $N = 19$) and the no intervention control group. These analyses showed large effects of the factor intervention for both monosyllabic and polysyllabic word reading fluency, $F(1, 116) = 12.45$, $p < .001$, $\eta_p^2 = .177$ and $F(1, 116) = 13.18$, $p < .001$, $\eta_p^2 = .185$, respectively. In addition, the main effects of FR-information and the interaction effects with FR-information were no longer significant. Subsequently, we specified two orthogonal contrasts. In the first contrast we compared the completers with the noncompleters and the no intervention controls. The second contrast tested for the difference between the latter groups. Tests of the first contrast showed that the completers outperformed the other two groups, $F(1, 118) = 31.12$, $p < .001$, $\eta_p^2 = .209$ for monosyllabic and $F(1, 118) = 33.17$, $p < .001$, $\eta_p^2 = .219$ for polysyllabic word reading fluency. The comparison between the noncompleters and the no intervention control group just missed significance, $F(1, 118) = 3.76$, $p = .055$, $\eta_p^2 = .031$ for monosyllabic and $F(1, 118) = 3.24$, $p = .075$, $\eta_p^2 = .027$ for polysyllabic word reading. There was a weak trend that the noncompleters performed somewhat more poorly than the no intervention control group. Similar results were found with respect to nonword and text reading fluency at the first follow up. Again, the completers significantly outperformed the other groups, $F(1, 118) = 31.18$, $p < .001$, $\eta_p^2 = .209$ for pseudoword reading fluency and $F(1, 118) = 31.75$, $p < .001$, $\eta_p^2 = .212$ for text reading fluency. However, in these analyses the difference between the noncompleters and the no intervention control group reached significance on pseudoword reading fluency: The latter group obtained somewhat higher scores, $F(1, 118) = 5.83$, $p = .017$, $\eta_p^2 = .047$. For text reading fluency, the difference just missed significance: $F(1, 118) = 3.24$, $p = .075$, $\eta_p^2 = .027$.

Subsample With Information on FR: Effects of FR-Status

All the analyses that follow were done in the FR-information subsample ($N = 74$) in which information about the FR-status of the children was received. Our main interest here was whether children with and without an FR for reading difficulties differed in their response to the intervention. Moreover, in this subsample we also examined long-term effects of the intervention as only this subsample was followed until Grade 6. Effects were tested with ANOVA or repeated measures ANOVA with intervention (yes or no) and FR-status (FR or no FR) as between-subjects variables.

Main Outcomes: Word-Reading Accuracy and Fluency

Descriptive statistics for word-reading accuracy and fluency from the end of kindergarten to mid-Grade 3 are presented in Table 4. At the end of kindergarten children's scores were at floor; the children could hardly read any words. In October of first grade, after approximately 8 weeks of reading instruction, there was some increase in word reading accuracy. An ANOVA showed no main effects of intervention and FR-status, but the interaction of Intervention × FR-Status was significant, $F(1, 70) = 6.12$, $p = .016$, $\eta_p^2 = .08$. Follow-up analyses indicated that at this time the intervention had a significant effect in the FR group, $F(1, 70) = 6.93$, $p = .01$, whereas the effect of the intervention was not significant in the group without an FR for reading difficulties, $F < 1$, *ns*.

Table 4
FR-Information Subsample (N = 74): Means and Standard Deviations for Reading Accuracy and Reading Fluency

	No intervention				Intervention			
	No FR		FR		No FR		FR	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Word-reading accuracy								
End kindergarten	1.05	2.99	0.53	1.01	1.74	2.90	4.18	4.32
October Grade 1	7.00	4.01	3.76	2.61	6.00	4.07	7.35	4.85
Reading fluency: Monosyllabic words								
Mid-Grade 1	5.86	5.02	2.82	2.89	6.86	5.26	4.71	4.68
End Grade 1	22.29	11.78	12.41	7.22	34.68	18.45	21.24	14.53
Mid-Grade 2	46.90	15.92	26.59	15.65	67.58	22.69	41.47	18.70
End Grade 2	57.81	16.56	35.94	16.26	79.11	18.28	52.94	19.56
Reading fluency: Polysyllabic words								
End Grade 1	12.00	8.01	4.88	3.82	20.47	12.13	11.94	9.22
Mid-Grade 2	29.29	12.67	16.00	11.06	45.79	16.46	26.24	26.24
End Grade 2	39.67	11.43	24.12	11.79	56.21	15.13	35.18	15.22
Mid-Grade 3	54.00	12.03	36.35	14.23	68.21	19.05	46.29	17.18

The mean performance on word reading fluency for each occasion split by intervention condition and FR-status is displayed in Figure 4. Two findings are noteworthy. First, both top panel (monosyllabic words) and bottom panel (polysyllabic words) show that within each condition the children without FR have higher mean scores than the children with FR. Second, on each occasion the difference in mean scores between intervention and no intervention condition in the FR-group (differences between red lines) seems approximately equal to the difference between the condition in the no FR-group (differences between blue lines), suggesting similar effects of the intervention for children with and without FR.

The interpretation of the pattern of findings in Figure 4 was supported by ANOVAs with repeated measures for monosyllabic word reading fluency and polysyllabic word reading fluency. The analysis of monosyllabic words showed main effects of intervention, $F(1, 70) = 18.50$, $p < .001$, $\eta_p^2 = .21$, and FR-status, $F(1, 70) = 29.52$, $p < .01$, $\eta_p^2 = .30$. These effects were qualified by interactions of intervention and FR-status with occasion, $F(3, 210) = 13.23$, $p < .001$, $\eta_p^2 = .16$, and $F(3, 210) = 20.99$, $p < .001$, $\eta_p^2 = .23$. The first interaction showed that the increase in monosyllabic word reading fluency from mid-Grade 1 to the first follow-up at end Grade 2 was larger in the intervention than in the no-intervention control condition (see top panel of Figure 4). The FR-Status × Occasion interaction indicated that the increase over time in monosyllabic word reading over this period was larger in the no FR-group than in the FR-group. These patterns were further confirmed in univariate ANOVAs per occasion. Mid-Grade 1, the effect of intervention was not yet significant, $p > .10$. Thereafter, the children in the intervention condition performed better, *p*s <. 01, and the effect size η_p^2 gradually increased from .14 to .23 at the first follow-up. Over the full period, FR children performed lower than no FR children, but the difference increased, η_p^2 ranging from .07 to .33. Importantly, the interaction of intervention and FR-status was not significant. Thus, irrespective of intervention condition, children without an FR for reading difficulties had a higher mean monosyllabic word reading fluency score at posttest.

Similar effects were found for the analysis with polysyllabic word reading fluency as outcome. There were main effects of Intervention, $F(1, 70) = 19.40$, $p < .001$, $\eta_p^2 = .209$, and FR-status, $F(1, 70) = 28.78$, $p < .01$, $\eta_p^2 = .295$. We also found a significant effect of Intervention × Occasion, $F(2, 140) = 6.39$, $p < .01$, $\eta_p^2 = .084$, indicating that the difference between the intervention and the control condition increased over time, and of FR-Status × Occasion, $F(2, 140) = 17.66$, $p < .01$, $\eta_p^2 = .210$, suggesting that the difference between the groups with and without FR increased over time (see also Figure 4, lower panel). These interactions were further confirmed in univariate ANOVAs per occasion. On each occasion the difference between intervention and control group was significant, *p*s < 0.1, but the η_p^2 increased from .17 to .22. Over time, the effect of FR increased from .17 to .33. As in the analysis with monosyllabic words, the interaction of Intervention and FR-status was not significant, $F < 1$, indicating that the intervention did not affect the difference in performance between children with and without FR, or put differently, that children with FR profited as much as the children without FR from the intervention.

Although the previous analyses suggest that the groups with and without FR profited equally from the intervention it is possible that it took the FR-children more sessions to attain the same effect of the intervention. Indeed during the period between mid-Grade 1, the first occasion of measurement of (monosyllabic) word reading fluency, and mid-Grade 2, the end of the intervention, the children in the FR-group received more sessions than the children without FR, $M = 118.47$ ($SD = 58.70$) and $M = 79.42$ ($SD = 29.62$), respectively. This difference was significant, $t(34) = 2.56$, $p = .15$. The difference remained significant if we removed one outlier (3 *SD* from the mean). Next we calculated a word-gain score showing how much children progressed on monosyllabic word-reading fluency between mid-Grade 1 and mid-Grade 2. We divided this word-gain score by the number of sessions in Grade 1 and 2 to obtain per child the "gain per session." A *t* test (for unequal variances) showed that the word gain per session in the FR-group, $M = .40$ ($SD = .30$), was significantly lower than in the group without FR, $M = .90$ ($SD = .47$), $t(30.79) = -3.93$, $p < .001$. In sum, although the ANOVAs on the reading outcomes yielded no significant interaction (suggesting equal gain for children with and without FR), correcting for the fact that the FR-

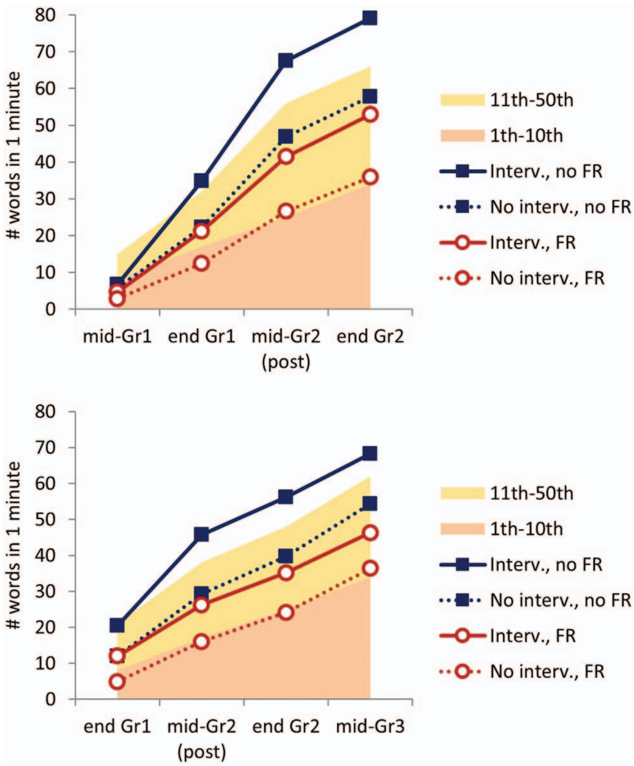

Figure 4. The subsample with FR-information (N = 74): word-reading fluency development in Grade 1, Grade 2 (posttest and follow-up 1) and Grade 3 (follow-up 2). The score represents the number of words read correctly within 1 min. Monosyllabic word reading (top panel) was assessed from mid-Grade 1 onward; polysyllabic word reading (bottom panel) is more demanding and was therefore assessed from end Grade 1 onward. The solid lines represent the means for the intervention groups, the dotted lines the no-intervention groups, the blue squares the no-FR groups, and the red circles the FR-groups. For reference, the background shades indicate the percentile scores based on national norms. For example, the intervention group without FR scores around the 10th percentile mid-Grade 1, but reads above average after the intervention. See the online article for the color version of this figure.

group required more sessions to reach this gain showed that the FR-group benefited less per session.

Finally, we examined the effect of the intervention on the proportion of poor readers (belonging to the 25% weakest readers according to national norms) and the proportion of very poor readers (10% weakest readers) 1 year after the intervention mid-Grade 3. The cumulative percentages of very poor and poor readers broken down by intervention and FR-status are displayed in Figure 5 (top panel). The percentage of poor readers (orange plus yellow) was significantly lower in the intervention group (27.8%) than in the no-intervention control group (60.5%), $\chi^2(1) = 8.02$, $p = .005$. The percentage of very poor readers (orange in the figure) was lower as well, but this difference, 13.9% versus 23.7%, did not reach significance. We also found significant effects of FR-status on the percentages of poor and very poor readers. There were 64.7% poor readers in the FR-group and 27.5% in the group without FR, $\chi^2(1) = 10.30$, $p = .001$. Similarly, in the FR-group 35.3% were very poor readers, whereas this was 5.0% in the group without FR. The percentages of very poor readers in the group without FR were very low (see orange in top panel of Figure 5). In the FR-group the number of poor readers and of very poor readers seems larger in the no-intervention than in the intervention groups (poor readers: 88.2% vs. 41.2%; very poor readers: 47.1% vs. 23.5%) but the numbers of children are too small to test whether these differences are significant.

Short-Term Transfer Effects

Pseudoword reading fluency, text-reading fluency, and spelling were assessed at the end of Grade 2, half a year after the intervention was finished, to test whether the intervention also affected untrained reading-related skills. In addition, reading comprehension was measured mid-Grade 3. The means and standard deviations in the various conditions are presented in Table 5.

An ANOVA with pseudoword reading fluency (end of Grade 2) showed a large effect of Intervention, $F(1, 70) = 14.03$, $p < .001$; $\eta_p^2 = .167$ and FR-status, $F(1, 70) = 25.89$, $p < .001$; $\eta_p^2 = .270$. The Intervention × FR-Status interaction effect was not signifi-

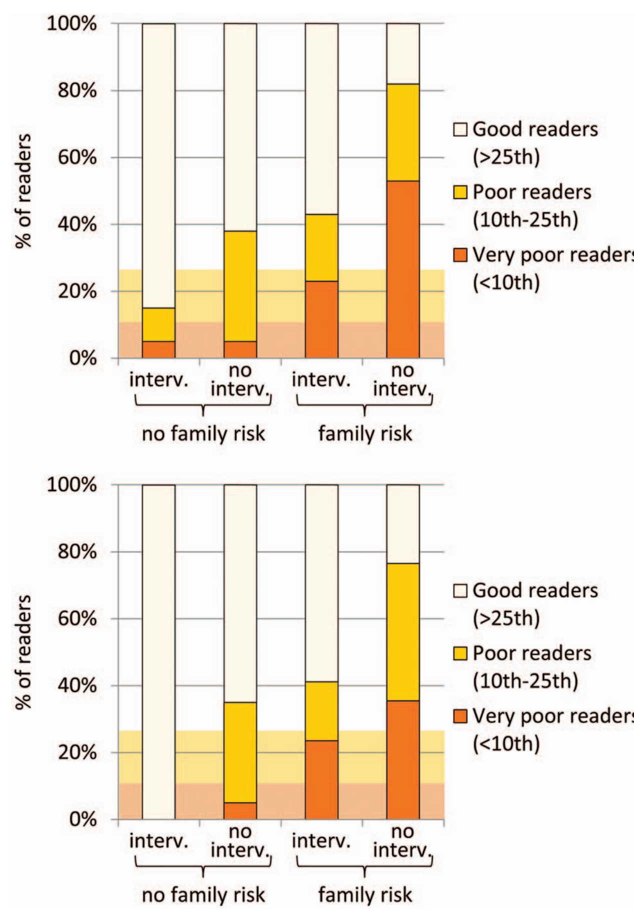

Figure 5. The subsample with FR-information (N = 74): Percentages of good, poor, and very poor readers by family risk and intervention in Grade 3 (top panel) and Grade 6 (bottom panel). Reading fluency of polysyllabic words was used to assess children's reading level. INT = intervention. See the online article for the color version of this figure.

Table 5
FR-Information Subsample (N = 74): Transfer Outcomes at Follow-Up End of Grade 2 and Mid-Grade 3

Measure	No intervention				Intervention			
	No FR		FR		No FR		FR	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pseudoword reading (end Grade 2)	32.76	10.22	21.65	10.60	49.89	18.36	28.47	14.10
Text-reading fluency (end Grade 2)	96.30	18.54	74.99	30.64	120.32	29.62	96.09	23.80
Spelling (end Grade 2)	12.81	7.44	9.00	6.41	18.22	7.75	12.12	7.49
Reading comprehension (mid-Grade 3)	25.70	4.79	22.63	6.63	24.87	4.79	23.59	5.25

cant, $F(1, 70) = 2.60$, $p = .111$; $\eta^2 = .036$. Thus, the groups with and without FR benefited equally from the intervention.

Regarding text-reading fluency (end of Grade 2), the children made very few errors ($M = 3.5\%$; $SD = 7.32$) with almost 90% of the children meeting the national objective to read at a 96% accuracy level. Two outliers with error rates larger than 3 *SD*s from the mean were removed before conducting the statistical analyses. A large effect of intervention and FR-status was found for text-reading fluency, $F(1, 68) = 13.57$, $p < .001$; $\eta_p^2 = .166$, and $F(1, 68) = 13.84$, $p < .001$; $\eta_p^2 = .169$, respectively. There was no significant interaction of intervention and FR-status, $F < 1$.

For spelling (end of Grade 2) also significant main effects of intervention and FR-status were found, $F(1, 70) = 6.22$, $p = .015$; $\eta_p^2 = .082$, and $F(1, 70) = 8.40$, $p < .01$ $\eta_p^2 = .107$. The interaction of intervention and FR-status was not significant, $F < 1$.

Reading comprehension was measured mid-Grade 3. Two children could not participate in this test: One child moved to another school and the other child did not have parental permission for this test. The results for reading comprehension revealed no significant effects of intervention, $F(1, 68) < 1$, and FR-status, $F(1, 68) = 1.90$, $p > .05$; $\eta_p^2 = .027$.

In sum, we found follow-up effects of the intervention on untrained reading-related skills, that is pseudoword and text reading fluency, and spelling. The beneficial effects of the intervention were similar for children with and without FR. The intervention did not have an effect on reading comprehension.

Long-Term Effects: Outcomes in Grade 6

We examined the long-term effects of the intervention (mid-Grade 6, four years after the intervention) on reading fluency, English word-reading fluency, and Dutch spelling. Data of two children without FR (one in the intervention and one in the control group) could not be gathered as these children had moved outside the area of Amsterdam. Both children had above average reading performance halfway Grade 3. In addition, one FR-child in the control condition was removed from the analyses in Grade 6 because he scored more than 3 *SD* below the mean on all reading measures; his Grade 6 reading ability was at the level of halfway Grade 2. The scores of all other variables were approximately normally distributed and did not contain outliers.

The descriptive information on these outcome variables is displayed in Table 6. A 2 × 2 multivariate ANOVA on the two reading fluency outcomes (monosyllabic and polysyllabic words) showed that performance in the intervention condition was significantly higher than in the no intervention group, $F(2, 66) = 4.51$, $p = .015$, $\eta_p^2 = .120$, and children without FR had higher mean scores than children with FR, $F(2, 66) = 12.23$, $p < .001$, $\eta_p^2 = .270$. The interaction of intervention condition and FR was not significant, $F = 2.20$, $p > .10$. Subsequent univariate ANOVAs confirmed these results.

We also found an effect of intervention condition and FR-status on English reading fluency, $F(1, 67) = 7.72$, $p = .007$, $\eta_p^2 = .103$ and $F(1, 67) = 4.01$, $p = .049$, $\eta_p^2 = .057$, respectively. The interaction of intervention condition and FR-status was not significant, $F(1, 67) = 1.43$, $p = .236$, $\eta_p^2 = .021$. For spelling we found a significant effect of FR-status, $F(1, 68) = 4.84$, $p = .031$, $\eta_p^2 = .066$, but the effect for intervention condition just missed significance, $F(1, 68) = 3.95$, $p = .051$, $\eta_p^2 = .055$. The Intervention Condition × FR-Status interaction effect was not significant, $F < 1$.

Regarding the incidence of reading difficulties in Grade 6, we examined the effect of the intervention on the percentage of poor and very poor readers, defined as the weakest 25% and 10%, respectively, according to national norms. The results broken down by intervention condition and FR-status are presented in Figure 5, bottom panel. The percentage poor readers (orange plus yellow) was significantly lower in the intervention condition (22.9% vs. 54.1%), $\chi^2(1) = 7.37$, $p = .007$, whereas there was no significant difference in the percentages of very poor readers (11.4% vs. 16.2%). The percentage of poor readers in the FR-group was significantly higher than in the group of children without FR (61.8% vs. 18.4%), $\chi^2(1) = 14.19$, $p < .001$. We found a similar result for the very poor readers (FR-group: 26.5%, group without FR, 2.6%), $\chi^2(1) = 8.53$, $p = .003$. The percentages of

Table 6
FR-Information Subsample (N = 74): Long-Term Outcomes in Grade 6

Measure	No intervention				Intervention			
	No FR		FR		No FR		FR	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
RF-Mono	101.30	12.01	90.37	11.53	111.39	16.57	95.94	15.13
RF-Poly	83.70	9.74	74.62	11.71	98.33	12.26	77.94	16.00
English reading fluency	52.10	9.88	49.87	11.47	63.05	12.75	54.23	12.18
Spelling	7.13	3.21	6.35	3.35	9.39	2.73	6.97	2.93

Note. RF-Mono = Reading Fluency Monosyllabic words; RF-Poly = Reading Fluency Polysyllabic words.

(very) poor readers in each of the four groups are displayed in Figure 5 (bottom panel). The figure shows that there were hardly any very poor readers in the groups without FR. The effects of the intervention on the percentage of poor readers were similar in the groups with and without FR, a drop of 29.4% and 35%, respectively.

Finally, we examined the percentage of children that were continuously promoted until Grade 6. In the intervention condition this was 91.4%, compared with only 69.4% in the no-intervention condition. This difference was significant, $\chi^2(1) = 5.42$, $p = .02$. There was no significant difference between children with and without FR.

In all, these results show that 4 years after the intervention had finished, the intervention group still consistently outperformed the no intervention control group in reading and reading-related outcomes. The intervention also resulted in a lower number of poor readers and of children repeating a class. The long-term intervention effects, as the short-term effects, were similar for children with and without an FR for reading difficulties.

Discussion

We examined the effects of a 2-year supplementary reading intervention for children at risk for reading difficulties, which started in kindergarten and covered the whole learning process from prereading to the start of fluent reading halfway second grade. Overall, the intervention did not improve the reading skills of children with an early cognitive risk profile (i.e., low preliteracy) in kindergarten as compared with at risk children in the no intervention control group that just received the regular instruction provided by the school. The only effect of the intervention was a reduction in the amount of extra help in the classroom during second grade. However, the intervention was highly successful in a subsample of children whose parents had provided information about the familiar risk for reading difficulties of their children. The intervention was not successful in the smaller subsample of children where family risk status could not be determined. In the subsample with FR-information we found clear benefits of the intervention: The children in the intervention group showed substantially more progress in word-reading fluency during the intervention, as well as 1 and 4 years after the program was completed. The incidence of children with reading difficulties had reduced substantially. Moreover, there were transfer effects to skills that were not specifically trained in the intervention, that is to pseudoword reading fluency, text-reading fluency, and spelling, but there was no transfer to reading comprehension. In addition, the intervention resulted in a lower number of children repeating a class, as well as reduction of the need for extra reading instruction by the teacher. In all the results suggests that the effects of the intervention are dependent on the characteristics of the children.

Effects of the Intervention in the Full Sample

For nearly 40% of the full sample family risk status for reading problems was missing. These parents did not return the questionnaire asking about their literacy difficulties. Although unanticipated, the simple omission to hand in a questionnaire appeared to act as a proxy for a number of background characteristics of these families. The children in this subsample without information about FR-status came from families with a substantially lower SES, were of a non-Dutch origin, and the overall majority used a minority language at home. It may be assumed that the majority of the parents in these families were (very) low in literacy skills, at least in the Dutch language. Probably as a consequence of this unfavorable background, the children in this group were poorer in Dutch vocabulary and phonological awareness than the children in the subsample with FR-information.

The lack of effect of the intervention in this subsample of lower SES children of non-Dutch origin is in accordance with earlier studies showing the absence of effects of early literacy interventions in ethnic minority samples and the small effects in lower SES samples (for a review see Manz, Hughes, Barnabas, Bracaliello, & Ginsburg-Block, 2010). Lack of support of the parents during the current intervention, especially acting as a tutor during kindergarten, is not likely to have been decisive for the absence of an intervention effect. In case parents were not able to give the instructions (in Dutch) when the children were still in kindergarten, a volunteer took over, either at home or at school. Moreover, in Grade 1 and 2 the intervention was executed at school by volunteers or older children. Instead, another hypothesis might explain the lack of effects in the subsample without FR-information. Children from minority and/or lower SES families tend to have less opportunities to acquire the foundations of reading, letter knowledge and phonological awareness (e.g., Lonigan et al., 1998). As a consequence of such a familial background, lags in these precursor skills of reading, which were used to select all "at risk" children for the intervention, might primarily reflect a lack of learning opportunities. According to the school professionals who were in charge of adapting instruction to the needs of the children, most of the no FR-information children were from parents who had only finished elementary school in their home country, suffering themselves from a lack of learning opportunities in their past. When school provides for these opportunities, their children are likely to profit from reading instruction and therefore are less prone to develop reading problems, and less dependent on extra intervention. In contrast, in the subsample with FR-information, consisting of Dutch speaking families with a substantially higher SES, a lag in the development of the precursors skills of reading might reflect genuine deficits and thereby be a better indicator for the risk of reading problems. This difference in the quality of the selection for risk of reading problems might also explain why the reading performance of both intervention and control group in the no FR-information subsample is higher than the performance of the no intervention group in the subsample with FR-information (see Figure 3).

Effects of the Intervention in the Subsample With FR-Information

In the subsample of children with information about FR-status, we found an increasing difference in word reading fluency between the intervention and the control group during the intervention and follow-ups. At the end of the intervention, the effect on word-reading fluency, the target skill, can be qualified as large. Halfway sixth grade, that is 4 years after the intervention was conducted, the effect remained equally strong, and there was no sign of fade-out. Of notice, the children in the intervention condition had received less extra help with reading from their school than the children who did not participate in the intervention. Thus, in the long run, the intervention Bouw! was more effective than the help that schools usually offer to children with reading difficulties. This supports the notion that prevention is more effective than remediation.

Interestingly, the intervention effect on word reading only emerged at the end of first grade, that is 1 and a half year after the start. This finding is in line with earlier studies (e.g., Bus & van IJzendoorn, 1999; Suggate, 2010), showing limited effects of short kindergarten interventions solely focusing on fostering preliteracy skills. It should be acknowledged that we did not examine the effects of the kindergarten part of our intervention, as we did not include measures of letter knowledge and phonological awareness at the end of kindergarten. Thus, in principle, we cannot know whether our intervention in kindergarten was effective. However, given the positive results of our previous study in kindergarten with a predecessor of the current intervention program (Regtvoort & van der Leij, 2007), it seems likely that also the current intervention enhanced preliteracy skills. Therefore, we take our findings to suggest that early interventions to foster word reading in at-risk children should be prolonged, encompassing both preliteracy skills and reading. This contention is further supported by our finding of larger benefits when the intervention was completed, which replicates our earlier study with the same intervention program (Regtvoort et al., 2013). However, we acknowledge that preexisting differences of the children in these groups could account for this finding. Nevertheless, it seems likely that completion is an essential condition to learn all relevant parts of the reading process, from the simplest letters to fluency in recognizing multisyllable words, giving children the necessary base for further independent development of reading. Moreover, if a child drops out, it usually falls back to classroom instruction which may not be adapted to his or her needs (Bailey et al., 2016).

The intervention decreased the incidence of reading difficulties (i.e., scoring in the bottom 25% and 10%) by about 50% in comparison to the control condition. However, not all of these differences reached significance. One reason for this finding could be that our groups of (very) poor readers were rather small, leading to a substantial reduction of power. To examine the effect of the intervention at the very low end of the reading distribution requires implementing the intervention on a far larger scale.

We found transfer effects of the intervention, both short-term and long-term, to skills that are related to word reading, that is pseudoword reading, text reading and English word reading fluency. There were also transfer effects to spelling. Moreover, in the long run the intervention resulted in less grade retention, which is an important finding, as grade retention is known to have negative effects on subsequent school careers and increases the risk of drop out (Hughes, West, Kim, & Bauer, 2018; Jimerson et al., 2006). In addition, preventing grade retention enlarges the cost-effectivity of the intervention.

The effects of the intervention did not generalize to reading comprehension. This result is in accordance with the outcome of Suggate's (2016) meta-analysis, which showed that reading comprehension is only enhanced in mixed interventions targeting both word reading and reading comprehension (see Lovett et al., 2017 for another example). The absence of an effect on reading comprehension seems also understandable. Fluent reading is only one of many skills needed to comprehend texts, and the effect of word reading ability on reading comprehension tends to decrease substantially between first and sixth grade (Garcia & Cain, 2014; van Viersen et al., 2018). Given the complexity of reading comprehension and the overall difficulty to enhance it directly (e.g., Compton, Miller, Elleman, & Steacy, 2014; Miciak et al., 2018; Muijselaar et al., 2018), it seems unlikely that reading comprehension can be improved solely by an intervention for word reading.

Effects of Family Risk for Reading Difficulties

In the study we distinguished between children with and without an FR for reading difficulties. Our main interest here was whether the effects of the intervention differed between children with and without FR. The findings are based on the subsample of children of which information about the FR-status was available.

In line with previous studies (e.g., Snowling & Melby-Lervåg, 2016; van Viersen et al., 2018), FR children performed lower on all outcomes (except for grade retention and reading comprehension) than children without FR. However, children with and without FR for reading difficulties did not respond differently to the intervention. Although FR children needed about 50% more sessions to reach the same progression as the children without FR, the difference in reading outcomes between the intervention and control condition was similar for both FR-groups. The absence of an interaction between familial risk, largely indicative of genetic predisposition, and intervention, an environmental manipulation, suggests that the effect of the intervention did not depend on genotype. Although it may be hypothesized that FR increases offspring's risk of reading difficulties via cultural transmission (e.g., if dyslexic parents would provide a disadvantageous literacy environment), longitudinal familial-risk studies indicate that risk is mainly conferred via genetic transmission (e.g., Torppa, Eklund, van Bergen, & Lyytinen, 2011). The notion that genetic transmission of reading skills outweighs cultural transmission is further supported by the finding that reading skills of parents and offspring are correlated in biological but not in adoptive families (Swagerman et al., 2017; van Bergen et al., 2015; Wadsworth et al., 2002). The current study adds a third source of support, as FR still had a substantial impact after optimizing the environment. Nevertheless, it is encouraging that the intervention positively impacted those with and without a genetic risk similarly. It suggests that individually tailored and prolonged intervention (an environmental impulse) can boost the reading ability of all children at risk for reading difficulties, irrespective of their genetic liability (as indexed by FR), although more effort is needed to reach the same results for children with an FR for reading difficulties.

Limitations

There are several limitations that need to be considered. First, although the implementation of the intervention was under tight control of the researchers, there was a substantial group of children who received fewer practice sessions than was prescribed in the protocol and did not complete the program. It is well established that treatment fidelity is a key element in successful intervention (e.g., Savage et al., 2013; Wolgemuth et al., 2014). Although the researchers, their assistants, and the school teachers tried to get all the parents involved, these attempts were not very successful in the no FR-info group. Most of these children were from immigrant, lower educated, parents who were not able enough in the Dutch language to act as tutor, and hard to get involved. In contrast, in a separate study we investigated treatment fidelity in the FRinfo group, which consisted of children who were all raised with Dutch as the main home language (even when the parents were of non-Dutch ethnicity: about a third). We found that treatment fidelity was high in terms of total training time, program completion, and tailoring the amount and level of practice to the child (Zijlstra et al., 2014). In future research, it is important to focus on the question how to promote parental support and involvement of hard-to-reach parents with immigrant background and limited literacy skills, and evaluate the effects in a large scale study in which schools themselves implement the intervention.

A second limitation is that random assignment was done in the full sample and the division in subsamples based on the availability of information about risk status was executed afterward. In principle assignment within the samples cannot be regarded as random as the ability or willingness to return the parental questionnaire about reading problems might have somehow been related to the intervention condition of the child. However, given the available information, this seems unlikely. The number of families that gave no information did not differ very much between the intervention and the no intervention condition (43% vs. 36%). Moreover, within both subsamples there were hardly any differences between the two intervention conditions at the start of the study. Although we cannot exclude the possibility that the later split in subsamples affected our within sample randomization, we believe the effect, if it exists, is probably small and does not affect the interpretation of our main findings on the intervention effects in the FRinfo sample.

A final limitation concerns the relatively small size of the groups in the FR-information subsample. The absence of an FR × Intervention interaction could be related to a moderate power to detect such an interaction. However, there is little indication that our findings in this subsample are driven by chance rather than true effects. The main effects of intervention condition and FR-status were significant and in the direction expected based on theory and our previous intervention study (Regtvoort et al., 2013). Moreover, the absence of an interaction effect is in line with theoretical arguments (e.g., Graham & Fisher, 2013). Note, though, that the FR-group received more intervention sessions as a result of their slower progress through the program. This was a natural consequence of our guiding principle to tailor the program's pace to a child's need. From a statistical view, this hampered a neat interaction test.

Practical Implications

Regarding practical implications, the findings reveal the importance of assessing the etiology of poor performance in the pre-reading phase. That information should be taken into account when establishing educational needs and setting a feasible reading-level target. Timely and prolonged intervention can, to large extent, prevent reading difficulties and thereby reduce financial and personal costs. Especially a program such as Bouw! seems cost-effective as it is delivered by nonprofessional tutors as compared with more costly interventions that require highly trained teachers. For example, the Reading Recovery program which is also supplementary to classroom instruction, is delivered by a specialized teacher on a one-to-one basis in a period of 12–20 weeks, half an hour per day, which amounts to between 30 and 50 hr of professional help (Sirinides, Gray, & May, 2018). In contrast, although Bouw! includes a period of 2 years, on average it took only 32 hr (on average 128 sessions of 15 min in the FR-information group). Although the involved volunteers have to be trained and supported by teachers, the main instruction and practice effort is relatively cheap. In addition, the reductions in professional remedial instruction and grade retention contribute to the cost-effectiveness of the current intervention. In the Netherlands preventing grade retention saves about $7,500 per child, whereas an online account for Bouw! costs less than $100, including training and support of the volunteers.

Conclusion

In all, the current study shows that a 2-year intervention starting in kindergarten can have long-term positive effects, especially in children from middle and higher SES families in which the dominant language (Dutch) is spoken. Our results suggest that reading difficulties in this group can be prevented in about half of the cases. This might also decrease the risk of subsequent problems in the long run, such as school failure, unemployment, or overall poor well-being, which all impact societal resources. However, the intervention seems not to provide an extra boost to children from lower SES immigrant families. It remains a challenge to adapt the intervention in such a way that reading can also be improved in these children.

References

Aarnoutse, C., & Kapinga, T. (2006). *Begrijpend Lezen 345678* [Reading Comprehension in grades 1–6]. Ridderkerk, the Netherlands: Onderwijs Advisering.

Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K. (2014). To wait in Tier 1 or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children, 81,* 11–27. http://dx.doi.org/10.1177/0014402914532234

Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G. K. (2020). From individual word recognition to word list and text reading fluency. *Journal of Educational Psychology, 112,* 22–39. http://dx.doi.org/10.1037/edu0000359

Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development, 56,* 45–72. http://dx.doi.org/10.1007/s11423-007-9067-0

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and Fadeout in the Impacts of Child and Adolescent Interventions. *Journal of Research on Educational Effectiveness, 10,* 7–39. http://dx.doi.org/10.1080/19345747.2016.1232459

Bailey, D. H., Nguyen, T., Jenkins, J. M., Domina, T., Clements, D. H., & Sarama, J. S. (2016). Fadeout in an early mathematics intervention: Constraining content or preexisting differences? *Developmental Psychology, 52,* 1457–1469. http://dx.doi.org/10.1037/dev0000188

Blachman, B. A., Schatschneider, C., Fletcher, J. M., Murray, M. S., Munger, K. A., & Vaughn, M. G. (2014). Intensive reading remediation in Grade 2 or 3: Are there effects a decade later? *Journal of Educational Psychology, 106,* 46–57. http://dx.doi.org/10.1037/a0033663

Bus, A. G., & Van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal* *of Educational Psychology, 91,* 403–414. http://dx.doi.org/10.1037/0022-0663.91.3.403

Byrne, B., Coventry, W. L., Olson, R. K., Samuelsson, S., Corley, R., Willcutt, E. G., . . . Defries, J. C. (2009). Genetic and environmental influences on aspects of literacy and language in early childhood: Continuity and change from preschool to Grade 2. *Journal of Neurolinguistics, 22,* 219–236. http://dx.doi.org/10.1016/j.jneuroling.2008.09.003

Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of preschool phoneme identity training after six years: Outcome level distinguished from rate of response. *Journal of Educational Psychology, 92,* 659–667. http://dx.doi.org/10.1037/0022-0663.92.4.659

Byrne, B., Olson, R. K., Samuelsson, S., Wadsworth, S., Corley, R., DeFries, J. C., & Willcutt, E. (2006). Genetic and environmental influences on early literacy. *Journal of Research in Reading, 29,* 33–49. http://dx.doi.org/10.1111/j.1467-9817.2006.00291.x

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

Catts, H. W. (2017). Early identification of reading disabilities. In K. Cain, D. L. Compton, & R. K. Parrila (Eds.), *Theories of reading development* (pp. 311–332). Amsterdam, the Netherlands: John Benjamins Publishing Company.

Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology, 104,* 166–181. http://dx.doi.org/10.1037/a0025323

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.

Compton, D. L., Miller, A. C., Elleman, A. M., & Steacy, L. M. (2014). Have we forsaken reading theory in the name of "quick fix" interventions for children with reading disability? *Scientific Studies of Reading, 18,* 55–73. http://dx.doi.org/10.1080/10888438.2013.836200

Connor, C. M., Morrison, F. J., Fishman, B., Crowe, E. C., Al Otaiba, S., & Schatschneider, C. (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. *Psychological Science, 24,* 1408–1419. http://dx.doi.org/10.1177/0956797612472204

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology, 91,* 450–476. http://dx.doi.org/10.1037/0022-0663.91.3.450

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250–287. http://dx.doi.org/10.1598/RRQ.36.3.2

Elbaum, B., Vaughn, S., Tejero Hughes, M., & Watson Moody, S. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology, 92,* 605–619. http://dx.doi.org/10.1037/0022-0663.92.4.605

Elbro, C., & Petersen, D. K. (2004). Long-term effects of phoneme awareness and letter sound training: An intervention study with children at risk for dyslexia. *Journal of Educational Psychology, 96,* 660–670. http://dx.doi.org/10.1037/0022-0663.96.4.660

Erickson, M. F., Sroufe, L. A., & Egeland, B. (1985). The relationship between quality of attachment and behavior problems in preschool in a high-risk sample. *Monographs of the Society for Research in Child Development, 50,* 147–166. http://dx.doi.org/10.2307/3333831

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. http://dx.doi.org/10.3758/BF03193146

Fluss, J., Ziegler, J. C., Warszawski, J., Ducot, B., Richard, G., & Billard, C. (2009). Poor reading in French elementary school: The interplay of cognitive, behavioral, and socioeconomic factors. *Journal of Developmental and Behavioral Pediatrics, 30,* 206–216. http://dx.doi.org/10.1097/DBP.0b013e3181a7ed6c

Foorman, B. R., Breier, J. I., & Fletcher, J. M. (2003). Interventions aimed at improving reading success: An evidence-based approach. *Developmental Neuropsychology, 24,* 613–639. http://dx.doi.org/10.1080/87565641.2003.9651913

Garcia, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research, 84,* 74–111. http://dx.doi.org/10.3102/0034654313499616

Geelhoed, J. W., & Reitsma, P. (1999). *PI-dictee*. Amsterdam, the Netherlands: Pearson.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79,* 1202–1242. http://dx.doi.org/10.3102/0034654309334431

Graham, S. A., & Fisher, S. E. (2013). Decoding the genetics of speech and language. *Current Opinion in Neurobiology, 23,* 43–51. http://dx.doi.org/10.1016/j.conb.2012.11.006

Greulich, L., Al Otaiba, S., Schatschneider, C., Wanzek, J., Ortiz, M., & Wagner, R. (2014). Understanding inadequate response to first-grade multi-tier intervention: Nomothetic and ideographic perspectives. *Learning Disability Quarterly, 37,* 204–217. http://dx.doi.org/10.1177/0731948714526999

Hindson, B., Byrne, B., Fielding-Barnsley, R., Newman, C., Hine, D. W., & Shankweiler, D. (2005). Assessment and early instruction of preschool children at risk for reading disability. *Journal of Educational Psychology, 97,* 687–704. http://dx.doi.org/10.1037/0022-0663.97.4.687

Hughes, J. N., West, S. G., Kim, H., & Bauer, S. S. (2018). Effect of early grade retention on school completion: A prospective study. *Journal of Educational Psychology, 110,* 974–991. http://dx.doi.org/10.1037/edu0000243

Jimerson, S. R., Pletcher, S. M., Graydon, K., Schnurr, B. L., Nickerson, A. B., & Kundert, D. K. (2006). Beyond grade retention and social promotion: Promoting the social and academic competence of students. *Psychology in the Schools, 43,* 85–97. http://dx.doi.org/10.1002/pits.20132

Katusic, S. K., Colligan, R. C., Barbaresi, W. J., Schaid, D. J., & Jacobsen, S. J. (2001). Incidence of reading disability in a population-based birth cohort, 1976–1982, Rochester, Minnesota. *Mayo Clinic Proceedings, 76,* 1081–1092. http://dx.doi.org/10.4065/76.11.1081

Kort, W., Schittekatte, M., & Compaan, E. (2008). *CELF-4-NL: Clinical evaluation of language fundamentals*. Amsterdam, the Netherlands: Pearson Assessment and Information.

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study. *Developmental Psychology, 45,* 764–781. http://dx.doi.org/10.1037/a0014132

Lonigan, C. J., Burgess, S. R., Anthony, J. L., & Barker, T. A. (1998). Development of phonological sensitivity in 2- to 5-year-old children. *Journal of Educational Psychology, 90,* 294–311. http://dx.doi.org/10.1037/0022-0663.90.2.294

Lovett, M. W., Frijters, J. C., Wolf, M., Steinbach, K. A., Sevcik, R. A., & Morris, R. D. (2017). Early intervention for children at risk for reading disabilities: The impact of grade at intervention and individual differences on intervention outcomes. *Journal of Educational Psychology, 109,* 889–914. http://dx.doi.org/10.1037/edu0000181

Manz, P. H., Hughes, C., Barnabas, E., Bracaliello, C., & Ginsburg-Block, M. (2010). A descriptive review and meta-analysis of family-based emergent literacy interventions: To what extent is the research applicable to low-income, ethnic-minority or linguistically-diverse young children? *Early Childhood Research Quarterly, 25,* 409–431. http://dx.doi.org/10.1016/j.ecresq.2010.03.002

Miciak, J., Roberts, G., Pat Taylor, W., Solis, M., Ahmed, Y., Vaughn, S., & Fletcher, J. M. (2018). The effects of one versus two years of intensive reading intervention implemented with late elementary struggling readers. *Learning Disabilities Research & Practice, 33,* 24–36. http://dx.doi.org/10.1111/ldrp.12159

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137,* 267–296. http://dx.doi.org/10.1037/a0021890

Muijselaar, M., Swart, N., Steenbeek-Planting, E., Droop, M., Verhoeven, L., & de Jong, P. F. (2018). The effect of a strategy training on reading comprehension in fourth-grade students. *The Journal of Educational Research, 111,* 690–703. http://dx.doi.org/10.1080/00220671.2017.1396439

Olson, R. K., Keenan, J. M., Byrne, B., & Samuelsson, S. (2014). Why do children differ in their development of reading and related skills? *Scientific Studies of Reading, 18,* 38–54. http://dx.doi.org/10.1080/10888438.2013.800521

Poskiparta, E., Niemi, P., Lepola, J., Ahtola, A., & Laine, P. (2003). Motivational-emotional vulnerability and difficulties in learning to read and spell. *The British Journal of Educational Psychology, 73,* 187–206. http://dx.doi.org/10.1348/00070990360626930

Raven, J., Raven, J. C., & Court, J. H. (2004). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Hartcourt Assessment.

Regtvoort, A., & van der Leij, A. (2007). Early intervention with children of dyslexic parents: Effects of computer-based reading instruction at home on literacy acquisition. *Learning and Individual Differences, 17,* 35–53. http://dx.doi.org/10.1016/j.lindif.2007.01.005

Regtvoort, A., Zijlstra, H., & van der Leij, A. (2013). The effectiveness of a 2-year supplementary tutor-assisted computerized intervention on the reading development of beginning readers at risk for reading difficulties: A randomized controlled trial. *Dyslexia, 19,* 256–280. http://dx.doi.org/10.1002/dys.1465

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

Savage, R., Abrami, P. C., Piquette, N., Wood, E., Deleveaux, G., Sanghera-Sidhu, S., & Burgos, G. (2013). A (Pan-Canadian) cluster randomized control effectiveness trial of the ABRACADABRA web-based literacy program. *Journal of Educational Psychology, 105,* 310–328. http://dx.doi.org/10.1037/a0031025

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

Sénéchal, M., Whissel, J., & Bildfell, A. (2017). Starting from home: Home literacy practices that make a difference. In K. Cain, D. L. Compton, & R. K. Parrila (Eds.), *Theories of reading development* (pp. 169–190). Amsterdam, the Netherlands: John Benjamins Publishing Company.

Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 36, pp. 31–82). Amsterdam, the Netherlands: Elsevier. http://dx.doi.org/10.1016/S0065-2407(08)00002-5

Sirinides, P., Gray, A., & May, H. (2018). The Impacts of reading recovery at scale: Results from the 4-year i3 external evaluation. *Educational Evaluation and Policy Analysis, 40,* 316–335. http://dx.doi.org/10.3102/0162373718764828

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, UK: Sage Publishers.

Snowling, M., Dawes, P., Nash, H., & Hulme, C. (2012). Validity of a protocol for adult self-report of dyslexia and related difficulties. *Dyslexia, 18,* 1–15. http://dx.doi.org/10.1002/dys.1432

Snowling, M. J., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin, 142,* 498–545. http://dx.doi.org/10.1037/bul0000037

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–407.

Suggate, S. P. (2010). Why what we teach depends on when: Grade and reading intervention modality moderate effect size. *Developmental Psychology, 46,* 1556–1579. http://dx.doi.org/10.1037/a0020612

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities, 49,* 77–96. http://dx.doi.org/10.1177/0022219414528540

Swagerman, S. C., van Bergen, E., Dolan, C., de Geus, E. J. C., Koenis, M. M. G., Hulshoff Pol, H. E., & Boomsma, D. I. (2017). Genetic transmission of reading ability. *Brain and Language, 172,* 3–8. http://dx.doi.org/10.1016/j.bandl.2015.07.008

Thijs, J. T., & Koomen, H. M. (2008). Task-related interactions between kindergarten children and their teachers: The role of emotional security. *Infant and Child Development, 17,* 181–197. http://dx.doi.org/10.1002/icd.552

Torgesen, J. (2005). Recent discoveries on remedial interventions for children with dyslexia. In M. J. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 521–537). Malden, MA: Blackwell Publishing. http://dx.doi.org/10.1002/9780470757642.ch27

Torgesen, J. K., Wagner, R., & Rashotte, C. T. (2012). *Test of Word Reading Efficiency (TOWRE). 1999*. London, UK: Psychological Corporation.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91,* 579–593. http://dx.doi.org/10.1037/0022-0663.91.4.579

Torppa, M., Eklund, K., van Bergen, E., & Lyytinen, H. (2011). Parental literacy predicts children's literacy: A longitudinal family-risk study. *Dyslexia, 17,* 339–355. http://dx.doi.org/10.1002/dys.437

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,* 1389–1401. http://dx.doi.org/10.1007/s10802-015-0003-1

van Bergen, E., Bishop, D., van Zuijen, T., & de Jong, P. F. (2015). How does parental reading influence children's reading? A study of cognitive mediation. *Scientific Studies of Reading, 19,* 325–339. http://dx.doi.org/10.1080/10888438.2015.1050103

van Bergen, E., de Jong, P. F., Maassen, B., & van der Leij, A. (2014). The effect of parents' literacy skills and children's preliteracy skills on the risk of dyslexia. *Journal of Abnormal Child Psychology, 42,* 1187–1200. http://dx.doi.org/10.1007/s10802-014-9858-9

van den Bos, K. P., Lutje Spelberg, H. C., & Eleveld, M. A. (2004). Voorspelling van leessnelheid van leerlingen in groep 3 met tests afgenomen in groep 2 en 3 [Prediction of reading speed of children in Grade 1 with tests executed in KG and Grade 3]. *Tijdschrift voor Orthopedagogiek, 43,* 312–324.

van den Bos, K. P., Lutje Spelberg, H. C., Scheepstra, A. J. M., & de Vries, J. R. (1994). *De Klepel, vorm A en B: Een test voor de leesvaardigheid van pseudowoorden*. Lisse, the Netherlands: Swets & Zeitlinger.

van der Leij, A. (2013). Dyslexia and early intervention: What did we learn from the Dutch Dyslexia Programme? *Dyslexia, 19,* 241–255. http://dx.doi.org/10.1002/dys.1466

van der Leij, A., & van Daal, V. H. (1999). Automatization aspects of dyslexia: Speed limitations in word identification, sensitivity to increasing task demands, and orthographic compensation. *Journal of Learning Disabilities, 32,* 417–428. http://dx.doi.org/10.1177/002221949903200507

van Otterloo, S. G., & van der Leij, A. (2009). Dutch home-based pre-reading intervention with children at familial risk of dyslexia. *Annals of Dyslexia, 59,* 169–195. http://dx.doi.org/10.1007/s11881-009-0030-0

van Viersen, S., de Bree, E. H., Zee, M., Maassen, B., van der Leij, A., & de Jong, P. F. (2018). Pathways into literacy: The role of early oral language abilities and family risk for dyslexia. *Psychological Science, 29,* 418–428. http://dx.doi.org/10.1177/0956797617736886

Verhoeven, L. (1993). *Grafementoets. Toets voor auditieve synthese. Handleiding*. Arnhem, the Netherlands: CITO.

Verhoeven, L. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading, 4,* 313–330. http://dx.doi.org/10.1207/S1532799XSSR0404_4

Verhoeven, L., & van Leeuwe, J. (2003). Ontwikkeling van decodeervaardigheid in het basisonderwijs [Development of recoding skill in primary education]. *Pedagogische Studiën, 80,* 257–271.

Verhoeven, L., & Vermeer, A. (1996). *Taaltoets allochtone kinderen*. Tilburg, the Netherlands: Zwijsen.

Visser, J., van Laarhoven, A., ter Beek, A. (1994). *AVI-toets. Leeskaart A en B*. 's Hertogenbosch, the Netherlands: Katholiek Pedagogisch Centrum.

Wadsworth, S. J., Corley, R. P., Hewitt, J. K., Plomin, R., & DeFries, J. C. (2002). Parent-offspring resemblance for reading performance at 7, 12, and 16 years of age in the Colorado Adoption Project. *Journal of Child Psychology and Psychiatry, 43,* 769–774. http://dx.doi.org/10.1111/1469-7610.00085

Wolgemuth, J. R., Abrami, P. C., Helmer, J., Savage, R., Harper, H., & Lea, T. (2014). Examining the impact of ABRACADABRA on early literacy in Northern Australia: An implementation fidelity analysis. *The Journal of Educational Research, 107,* 299–311. http://dx.doi.org/10.1080/00220671.2013.823369

Zijlstra, A. H., Koomen, H., Regtvoort, A., & Van der Leij, A. (2014). Effects of quantitative and qualitative treatment fidelity delivered by non-professional tutors. *Learning and Individual Differences, 33,* 55–62. http://dx.doi.org/10.1016/j.lindif.2014.04.004

Received February 15, 2019
Revision received April 3, 2020
Accepted April 3, 2020 ■