Early home-based intervention for children at familial risk of dyslexia

van Otterloo, S.G.

Publication date
2011

2

Early home-based intervention in the Netherlands for children at familial risk of dyslexia

Sandra G. van Otterloo, Aryan van der Leij, & Lotte F. Henrichs
This chapter is published in: *Dyslexia*, *2009*, *15*(3), 187–217.

Dutch children at higher familial risk of reading disability received a home-based intervention programme before formal reading instruction started to investigate whether this would reduce the risk of dyslexia. The experimental group (n = 23) received a specific training in phoneme awareness and letter knowledge. A control group (n = 25) received a non-specific training in morphology, syntax, and vocabulary. Both interventions were designed to take 10 min a day, 5 days a week for 10 weeks. Most parents were sufficiently able to work with the programme properly. At post-test the experimental group had gained more on phoneme awareness than the control group. The control group gained more on one of the morphology measures. On average, these specific training results did not lead to significant group differences in first-grade reading and spelling measures. However, fewer experimental children scored below 10th percentile on word recognition.

2.1 Introduction

People with severe reading disability that cannot be explained by sensory deficits, attention deficit, general language problems, intelligence, poor home-literacy environment, or insufficient schooling (poor education) suffer from developmental dyslexia (Snowling, 2000). Because literacy is essential in modern society, dyslexia is a serious handicap. Problems dyslexic people experience turn out to be persistent. It is estimated that about 10–15% of the population suffers from dyslexia (Vellutino, Fletcher, Snowling, & Scanlon, 2004).

Dyslexia runs in families. Hallgren (1950) indicated that the probability that a child, with one dyslexic parent, will become dyslexic is approximately 41%. Other researchers also have reported considerable hit rates (e.g. Finucci, Gottfredson, & Childs, 1985; Gilger, Pennington, & De Fries, 1991; Scarborough, 1990). Because of this phenomenon, the risk of developmental dyslexia can be detected early. Moreover, this offers an opportunity to investigate whether early intervention may reduce the risk of dyslexia. The present paper reports a study on the effects of early intervention based on phoneme awareness and letter knowledge for children at higher familial risk of dyslexia. The programme was home-based and conducted in the Netherlands before the onset of formal reading instruction.

Dyslexia is a specific reading disability indicated by weaknesses in word recognition, identification of unfamiliar words (phonological recoding), and production of the written form of the word (spelling) (Lyon, 1995). It is presumed that the core deficit underlying this manifestation of dyslexia is a phonological deficit (Vellutino *et al.*, 2004). Before reading starts and in the initial phase of reading acquisition, segmental phonology (e.g. phoneme awareness) and letter knowledge are strong predictors of reading and spelling achievement and, therefore, of deviations of the normal pattern (Elbro & Scarborough, 2004a; Vellutino *et al.*, 2004). In later stages, when many words are recognized, slow naming speed of symbols correlates with dyslexia (de Jong & van der Leij, 2003; Wimmer, Mayringer, & Landerl, 2000; Wolf & Bowers, 1999). In addition, general language skills such as vocabulary, morphology and syntax, are also predictors of later reading development, although the correlations are less strong (Gallagher, Frith, & Snowling, 2000; Scarborough, 1998).

Correlational studies, such as the studies mentioned above, show that phoneme awareness and letter knowledge are early predictors of reading skill, but do not demonstrate that the relation is causal. Intervention studies, on the other hand, may support a causal relation. Several researchers have studied the effect of phonological awareness training of young children. In their meta-analyses, Bus and van IJzendoorn (1999) and Ehri *et al.* (2001) reported overall effect sizes of intervention on phoneme awareness of d = 1.04 and 0.86, respectively. These effect sizes are large, according to the criteria of Cohen (1988). On reading skill, however, the reported effect sizes were moderate, d = 0.44 and 0.53 respectively. At follow-ups, the effect sizes tended to decline. It seems possible to train phonological awareness successfully, but transfer effects on reading skill are limited.

Comparability across intervention studies may be hampered by differences in content of training, in sample selection, in the type of trainer (experimenters, teachers or parents), and in the country the study is conducted in, leading to differences in languages and teaching methods. These four differences are examined below.

2.1.1 Content of training

According to several researchers (e.g. Ball & Blachman, 1991; Bradley & Bryant, 1985; Bus & van IJzendoorn, 1999; Ehri *et al.*, 2001; Hatcher, Hulme, & Ellis, 1994; Schneider, Roth & Ennemoser, 2000) a preventive training programme for children at risk of dyslexia should include teaching of letter–sound correspondences to be effective. In addition, it should focus on the segmental phonological level (phoneme awareness) and not on the level of supra-segmental linguistic segments (syllable, words, or sentences) (Cary & Verhaeghe, 1994; Elbro & Scarborough, 2004b). Furthermore, Ehri *et al.* (2001) found that instruction in phoneme awareness was most effective when one or two phoneme awareness skills were trained instead of more skills. There seems to be a need of focus.

Lyster (2002) suggested that training of phoneme awareness might not be the only way to improve reading in later stages. In her study, she compared the effects of morphological versus phonological awareness training of Norwegian non-reading kindergartners of about 6 years old. Both interventions included some exposure to print. At the end of first grade, the morphology group and the phonology group outperformed an untrained control group on most measures related to reading. Important to the present study was the finding that the morphology group performed better than the phonology group on 'picture–word matching', a measure of word recognition. Although the phonology group outperformed the morphology group on several post-test measures, this did not result in any group difference in favour of the phonology group at transfer tests at the end of first grade. A study of Snowling, Gallagher, and Frith (2003) also gives a rationale for training these skills. They reported of children at familial risk for dyslexia who never manifest the phenotype because of strong language skills. These children scored weak on phoneme awareness skills, but high on vocabulary and morphology. Moreover, children with familial risk of dyslexia who later on turned out to become dyslexic often had general language delay in early school years.

Overall, it can be concluded that the best results are to be expected from intervention programmes focussing on phoneme awareness and letter knowledge, but training morphological awareness and vocabulary might also give positive effects.

2.1.2 Sample selection

A problem that hampers straight comparisons between the studies in the meta-analyses of Bus and van IJzendoorn (1999) and Ehri *et al.* (2001) is that they include both unselected samples and samples with children at risk. Moreover, children at risk were defined in a variety of ways; for instance 7-year-old children experiencing difficulties in the early stages of learning to read (Hatcher *et al.*, 1994), pre-school children with developmental delays (O'Connor, Jenkins, Leicester, & Slocum, 1993) or kindergartners with low phonological processing scores (Schneider *et al.*, 2000). In only two of the intervention studies that focussed on phoneme awareness and letter knowledge, children at higher familial risk were selected: an Australian study of Hindson *et al.* (2005), and a Danish study of Elbro *et al.* (see Borstrøm & Elbro, 1997; Elbro, Borstrøm, & Petersen, 1998; Elbro & Petersen, 2004; Petersen, 2000). Similar to the present study, children were selected because one (or even two) of the parents had dyslexia. Both studies were not included in the meta-analyses.

These two studies will be described in more detail because they relate directly to the present study. The intervention programmes can be called specific, because they focussed on phoneme awareness and letter knowledge. In addition, in the study of Hindson *et al.* (2005), the programme included 'dialogic reading', a method for structured book reading developed by Whitehurst *et al.* (Arnold, Lonigan, Whitehurst, & Epstein, 1994; Whitehurst *et al.*, 1994). In this Australian study, the teaching sessions were held individually, at home or at pre-school, once a week, by a trained assistant, and took about 30 min. The number of sessions was partly governed by a criterion test and ranged from 11 to 17. At post-test, the trained at-risk group gained more than an untrained at-risk group on measures of phoneme identity, rhyme, and print familiarity. Effect sizes were not given. On follow-up tests for word identification, pseudoword identification, and spelling, a group of trained not-at-risk children outperformed the trained at-risk children. This at-risk group, however, scored similar to untrained unselected children from an earlier study (Byrne & Fielding-Barnsley, 1993). Hindson *et al.* concluded that the intervention might have raised the performance of at-risk children to grade average. It should be noted, however, that their follow-up did not include a no-training at-risk group.

In the Danish study (Elbro & Petersen, 2004), training took place every school day for 17 weeks at the schools of the children and lasted about 30 min. The children worked on the programme with the whole class and with their own kindergarten teacher. The programme was more specific than the Australian programme because it did not contain the structured book reading. At post-test a trained at-risk group had progressed more on phoneme awareness than an untrained at-risk group. Effect sizes (Cohen's *d*) were between 0.47 and 0.67. At follow-up measurements, at the beginning of second and third grade, the trained at-risk children outperformed their controls on most (Petersen, 2000) or all (Elbro & Petersen, 2004) reading measures. Effect sizes (Cohen's *d*) were between 0.40 and 0.69. The number of trained at-risk children with recognizable reading or spelling problems, turned out to be smaller than expected (Petersen, 2000). In second grade 19% of the trained at-risk children were called dyslexic because they scored below a cut-off point on a composite measure of phonological recoding. Petersen chose a cut-off point corresponding with roughly the lowest 10% of children not at risk. For the untrained at-risk children the proportion was 47%. In the not-at-risk groups the prevalence was 11% in a trained and 8% in an untrained condition. Petersen concluded that the training programme tended to reduce the prevalence of dyslexia in at-risk children.

2.1.3 Kind of trainer

In all studies on early intervention described above, the trainers were schoolteachers (e.g. Ball & Blachman, 1991; Elbro & Petersen, 2004; Schneider *et al.*, 2000) or experimenters (e.g. Hindson *et al.*, 2005; O'Connor *et al.*, 1993). However, we decided to use parents as tutors. It is to be expected that experimenters can be controlled best, leading to good internal validity of the results. However, the use of schoolteachers or parents as trainer gives results with higher ecological validity. An advantage of parents, compared with schoolteachers, is that the training is individual. For schools, individual instruction always presents a financial burden. Another argument for the choice of parents is that the motivation of parents who have experience with dyslexia will be high. Moreover, our study on treatment integrity (van Otterloo, van der Leij, & Veldkamp, 2006[1]) indicated that most parents of children at familial risk were able to carry out a similar home-based intervention with sufficient quality and quantity.

Apart from practical reasons, there is evidence to support our choice of trainers. Several researchers showed that parents can have substantial effect on the pre-reading development of their children by directly teaching letters at home (e.g. Sénéchal & LeFevre, 2002). Torppa *et al.* even concluded that parental teaching of letter names to children at familial risk of dyslexia may compensate for the risk (Torppa, Poikkeus, Laakso, Eklund, and Lyytinen, 2006).

Fielding-Barnsley and Purdie (2003) investigated the effect of a home-based programme to train children at higher familial risk of dyslexia in the year prior to formal schooling. The home-based intervention programme used dialogic reading (Arnold *et al.*, 1994; Whitehurst *et al.*, 1994), with additional instruction focussing on rhyme awareness, alliteration, print familiarity, and alphabet knowledge. In the first year of formal schooling, the children were tested twice; at the beginning of the curriculum (time 1) and again a few months later (time 2). At time 1, the children outperformed untrained controls on receptive

---

1 Chapter 3 of this thesis

vocabulary, initial consonant recognition, rhyme, and print familiarity. The effect on alphabet knowledge also reached significance. At time 2 the trained children outperformed the controls on final consonant recognition, spelling, reading, and again on print familiarity. This shows that early home-based intervention programmes for children at familial risk can be effective.

2.1.4 Country

Results of intervention studies cannot be generalized across countries without taking into account differences in languages and educational traditions. Orthographic depth is a main source of differences between languages (Seymour, Aro, & Erskine, 2003). Languages with a shallow orthography (e.g. Finnish) have consistent grapheme–phoneme correspondences. Deep orthographies (e.g. Danish and English) are much more inconsistent and irregular. The Dutch orthography is somewhere in between the two extremes, comparable to German (Borgwaldt, 2003). These differences in orthography may influence the process of learning to read and consequently the results of early intervention. Ehri *et al.* (2001) report in their meta-analysis that the effects of phonological awareness training on reading outcomes were larger in English-speaking countries (deep orthography) than in non-English-speaking countries (different orthographies). The training effects on phoneme awareness and spelling outcomes were also larger in English-speaking countries.

The age at which formal reading instruction starts also differs across countries. Dutch children, as in most countries in Western Europe and in the United States, enter first grade in the year they become seven. Most children in Great Britain and Australia, however, get their first reading instruction a year earlier and children in the Scandinavian countries, including Denmark, a year later. Consequently, children at first grade may differ across countries in the maturity for the development of recoding skills. In the study of Seymour *et al.* (2003), however, age did not seem to be an important factor for the differences in the foundation of literacy acquisition. Therefore, age would probably not be of much importance in the comparability of intervention studies from different countries.

2.1.5 Research question, summary of design, and expectations

It may be concluded that several studies showed that phoneme awareness and letter knowledge are strong predictors of reading skills. Other studies supported the notion of causality, because it is possible to train these skills successfully, leading to better reading skill. However, studies with children at higher familial risk or conducted with parents are rare and they are all conducted in languages with relatively deep orthographies. The aim of the present study was to investigate whether a specific training focussing on phoneme awareness and letter knowledge, for Dutch kindergarten children at higher familial risk, and administered by their parents, lead to immediate effects on the trained skills and later transfer to reading (phonological recoding and word recognition) and spelling. Could the results of the studies of Elbro and Petersen (2004) and Hindson *et al.* (2005) be replicated when the native language had a relative shallow orthography? To control for Hawthorne effects, this specific training programme (experimental condition) was compared to a non-specific training programme focussing on morphology, syntax, and vocabulary (control condition) that may have an effect on word recognition (see Lyster, 2002) but not on phonological recoding and spelling.

We selected 5- or 6-year-old children with at least one dyslexic parent. These children attended kindergarten and did not yet receive formal reading instruction, which starts in first grade, the following school year. The children were assigned to the two training conditions and were measured on four occasions. At time 1, halfway the final kindergarten year, the children were pre-tested on three phoneme awareness measures (first sound identity, phoneme blending, and phoneme segmentation), a receptive letter knowledge measure, two morphology measures (morphology and plural form), and two rapid naming measures (rapid naming pictures and rapid naming colours). Between time 1 and time 2, the training programmes were executed in the two groups. The children were post-tested at the end of kindergarten (time 2) with the same measures. At time 3, in first grade after a few month of formal reading instruction, they were tested on two phoneme awareness measures (phoneme blending and phoneme segmentation), two letter knowledge measures (receptive letter knowledge and productive letter knowledge), and one reading measure (word recognition). At time 4, at the end of first grade, they were tested on one phoneme awareness measure (phoneme segmentation), two reading measures (word recognition and phonological recoding), and a spelling measure.

Because the specific programme focussed on phoneme awareness and letter knowledge, we expected more progress in the experimental group than in the control group on these measures between time 1 and time 2. In addition, we expected to find group differences in favour of the experimental group on these skills in first grade, at time 3 and 4, as well as transfer effects on phonological recoding, word recognition, and spelling. In contrast, the control children could be expected to make more progress on morphology in kindergarten, because the non-specific programme focussed on these skills. Possibly, this would lead to better word recognition in first grade. Neither programme would affect rapid naming, because that skill was not trained.

2.2 Method

2.2.1 Participants

All participants had at least one parent who gave a self-report of dyslexia. The parent was selected for participation if his or her scores on two reading tests (see below) were equal to or below the 20th percentile of the total population, or the score on one of the reading tests was equal to or below the 10th percentile. In addition, we selected the parent if his or her percentile score on the verbal IQ measure was much higher than the percentile score on one of the reading tests, with a disparity of at least 60 (e.g. the score on the verbal competence was above the 90th percentile and the score on one of the reading tests was beneath the 30th percentile). The rationale behind this criterion was that highly educated adults might show reading scores just above the other criteria, probably due to extended experience with reading. Still, when their reading scores were much lower than could be expected from their high verbal competency, it meant a serious reading handicap for these people.

After the screening, 49 out of 70 families met the criteria and participated in the experiment. Almost half of the dyslexic parents (23) scored below or equal to the 10th percentile on both reading measures. Nineteen parents scored

Table 2.1: Means and standard deviations on age and control measures

Measures	Experimental group n^{a} = 23		Control group n^{a} = 25	
	M	SD	*M*	SD
Age[b] (in month)	71.77	4.98	71.08	4.60
Receptive vocabulary	68.09	8.21	71.48	10.62
Nonverbal IQ	24.52	5.25	27.26	3.40

a Kindergarten (time 1 and time 2).
b Measured at pre-test (time 1).

below or equal to the 10th percentile on one of the reading measures and two parents scored below or equal to the 20th percentile on both tests. Five parents were included because of a large disparity between reading skill and verbal competence. If both parents were dyslexic (at least in four families), only the scores of the parent with the lowest scores were included. To indicate that it takes considerable effort to select even a sample of moderate size, it should be noted that, with an estimated incidence of 10–15% in the population, the 49 families were derived from about 350–500 families.

The children came from 30 schools. They were randomly assigned to the conditions within two constraints. (1) Children from the same school participated in the same experimental condition to keep detailed information about the other training programme away from the participating parents. (2) The distribution of girls and boys was kept equal over both conditions.

During the experiment one participant left the experimental group, because the programme was too difficult for her. At post-test (time 2), data sets from 23 at-risk experimental children (12 males and 11 females) and 25 at-risk control children (14 males and 11 females) were available for data analyses. During follow-up measurements halfway first grade (time 3) five children had left the experiment because of personal reasons, leaving for special education, or staying an extra year in kindergarten. After this, we still had 20 participants in the experimental group and 23 participants in the control group. At the second follow-up in June (time 4) we lost another child from the experimental group through migration.

Table 2.1 describes the participants. All children were native Dutch speakers. The groups did not differ in mean scores on the selection measures of the

Table 2.2: Background variables related to home literacy environment

Variables		Experimental group $n = 15$	Control group $n = 19$
		%	%
Education level mother	High	40	47
	Middle	53	42
	Low	7	11
Education level father	High	40	58
	Middle	27	32
	Low	33	11
Only father dyslexic		67	68
Only mother dyslexic		20	26
Both parents dyslexic		13	5
Reading problems direct family		60	53
Early problems with language (Child)		47	35
Number of books in the home	< 100	33	32
	100–200	20	21
	> 200	47	47
Library membership child		87	74
Shared reading	$< 2 \times$ per week	7	0
	$2–4 \times$ per week	40	32
	$\geq 5 \times$ per week	53	68

dyslexic parents. At pre-test the groups did not differ in age or performance on receptive vocabulary. In comparison with age related norms the groups scored on level B (above mean; percentiles 50–75) on this measure. In general the groups also had a nonverbal IQ-score above mean. However, the control group scored significantly better, $t(42) = 2.07, p < 0.05$. We did not include nonverbal IQ as a covariate or between factor in final analyses, because it did not correlate with any dependent variable.

Table 2.2 provides background information given by the parents, using the questionnaire of Eleveld (2005). Chi-square tests indicated that the groups did not differ significantly on any measure. It is clear that a substantial number of selected children had early language problems and that about half of their direct-family members (siblings, uncles, aunts, grandparents) had reading problems. Because we excluded environment as a causal factor, it is important to note that Table 2.2 indicates that environmental conditions as the home-literacy environment (books, library membership, reading to the child) and education level of the parents do not seem to be weak in most cases.

2.2.2 Experimental programme

The experimental programme, aimed to train phoneme awareness and letter knowledge, was called 'Klinkende Klanken en Lollige Letters' ['Soundings sounds and jolly letters'] (unpublished). The programme was home-based with parents as tutors. It was designed to take about 10 min a day, 5 days a week for 10 weeks (a total of 8–8.5 h). Taken from the Danish programme of Borstrøm and Petersen (1996), the programme focussed on single speech sounds right from the start. It progressed with approximately two speech sounds a week. Every speech sound and the corresponding letter were introduced in several ways. First a rhyme or song was read to the child, in which the speech sound played an important role. Then the child was shown a picture of the letter while the parent produced the corresponding speech sound. If possible the child was also given a semantic cue (/m/ is the 'taste-good' sound). After that, the parent and child wrote down the letter several times on a slate. Together they thought of people they knew whose name started with the sound, followed by an articulation exercise with attention to rounding (vowels), place of articulation, and manner of articulation. For this exercise a mirror was used. The newly learned speech sounds and corresponding letters were repeated frequently during the programme in nursery rimes and by means of phoneme awareness exercises. The exercises primarily focussed on phoneme blending and sound identity of both initial and final sound. In one repeated blending exercise, the children were taught how to put a consonant in front of a (pseudo)word. The (pseudo)word was visualized by a box. The parent drew a box on the slate and produced the sound that was represented by the box, e.g. *ice* (/aIs/). After this the parent explained to the child that they would obtain a new word when they put a consonant (e.g. /m/ or /r/) in front of the box (e.g. *mice* or *rice*). The parent illustrated this by drawing the consonant on the slate in front of the box. In several other exercises the parents and child were invited to use a glove puppet. The puppet gave a riddle, which the child had to solve, or made language mistakes, which the child had to correct. The purpose of using the doll was to stimulate correct phonological representations of words in a playful way. During the final week, all the sounds of the programme were practised in several language games.

We only used long and no short vowels in the programme. In Dutch, a homogeneous digraph -as for instance aa or ee- represents long vowels. Twelve (highly frequent) letters and speech sounds were used in the programme: aa, oo, ee, uu, m, s, p, l, t, r, v and b (/a/, /o/, /e/, /y/, /m/, /s/, /p/, /l/, /t/, /r/, /v/, and /b/) The training took place during the months March, April and May 2002 (kindergarten). The training materials consisted of a programme book, a case with pictures and games, a slate with chalk, a glove puppet (a fox), an instruction sheet, and a progress table for the child with funny stickers. The programme book contained detailed descriptions of the daily lessons, including the nursery rime of the day and references to which materials to use. The instruction sheet explained the aims of the programme and gave general instructions. The parents were, among others, explicitly instructed to use speech sounds instead of letter names and were explained why.

2.2.3 Control programme

To control for the non-specific effects of intensive instruction (Hawthorne effects), the present study was designed as a comparative study between two training programmes, both conducted by parents. The control programme, 'Vrolijke Verhalen en Wonderlijke Woorden' [Happy stories and peculiar words] (unpublished), was designed to take the same time as the experimental programme, both per day and in total. The programme aimed to train morphology, syntax, and vocabulary. It can be called non-specific, as the programme did not aim to influence the core phonological deficit of dyslexia.

Every lesson started with a story from the 'Kijk en Luistermethode' [talking-books method] (van der Leij, 1982) read to the child by the parent. Every week the stories were set around a certain theme. In the second part of the lesson the child had to answer a few questions about the story to check for comprehension of the text. The third and fourth part consisted of language exercises and games focussing on morphological skills and syntax. The following topics were included: sentence structure or word order, plural form, past tense, negations, interrogative sentences, (definite and indefinite) articles, diminutives, working memory exercises, and vocabulary exercises such as riddles or contradictions (adversatives). As in the experimental programme, the glove puppet was used in the programme about 2 days a week. The training materials consisted of a programme book with detailed descriptions of the daily lessons, a case with story cards, a glove puppet, an instruction sheet, and a progress table for the child with funny stickers.

2.2.4 Measures

Screening

To investigate whether one (or even both) of the parents was dyslexic, both word recognition and phonological recoding were measured. Because of the important role of reading speed in the Netherlands compared with reading accuracy, we used fluency measures (number of words read correctly in 1 or 2 min). To investigate a possible disparity between reading level and verbal language level, a test of verbal competence was added. The same measures were used in other studies in the Netherlands that selected children at familial risk of dyslexia (e.g. Eleveld, 2005; Koster *et al.*, 2005; Regtvoort & van der Leij, 2007). The screening measures used to select parents are described below.

Fluency of word reading. (EMT [One-minute test]; Brus & Voeten, 1980). Word recognition was measured by a card with four columns of words (total 116) with increasing difficulty. The experimenter measured how many words the subject could read correctly in 1 min. Mean parallel-test reliability (between form A and B) was r_{tt}= 0.90 (range 0.76–0.96) (van den Bos, lutje Spelberg, Scheepstra, & de Vries, 1994).

Fluency of pseudoword reading. (de Klepel [the clapper]; van den Bos *et al.*, 1994). Phonological recoding was measured by a card with four columns of pseudowords with increasing difficulty, analogous to the EMT. The experimenter measured how many pseudowords the subject could read and pronounce correctly in 2 min. Reported mean parallel-test reliability (between form A and B) was r_{tt}= 0.92 (range 0.89–0.95).

Verbal competence. As a control measure, we used the subtest 'Similarities' from the Dutch translation of the Wechsler Adult Intelligence Scale—Revised (WAIS-R; Wechsler, 1981). The experimenter asked the subject to give the similarity between two concepts. Every answer gave 0, 1 or 2 points with a maximum score of 26.

Control measures children

Both non-verbal IQ and verbal IQ (receptive vocabulary) was measured to exclude children with very low IQ scores.

Non-verbal IQ. The Coloured Progressive Matrices (Raven, Court, & Raven, 1984) has three subtests with 12 test items each. Every item consists of a rectangular pattern in which a part is missing. The child has to look for the missing part and must choose between six alternatives. The maximum score is 36. Split-half reliability (corrected for test reduction) was 0.82 for 6-year olds and 0.84 for 7-year olds (van Bon, 1986). The test was administered during first follow-up (time 3).

Receptive vocabulary. This test is part of the Taaltoets Allochtone Kinderen [Language test for children of ethnic minorities] (Verhoeven & Vermeer, 1996). Each item consists of four pictures and a spoken word. The child has to choose the alternative that best matches the given word. The test had 98 items that increased in difficulty. The administration of the test was stopped when the child failed six or more of the last eight items. For this test, the Cronbach's alpha was 0.89. The test was administered during pre-test (time 1).

Manipulation check and usability of the programmes

To gather information about the application and usability of the programme, we asked the parents to fill in logbook forms. In addition, we asked the parents three times (after 1 week of training, after 3 weeks and again after 7 weeks of training) to fill in a short questionnaire about the quality of the instruction, the level of difficulty, experiences with their role of tutor, and the motivation of both children and parents.

Table 2.3: Overview of constructs, measures and occasion of administration

Construct	Measures	Kindergarten		First grade	
	Intervention:	⇓			
	Time:	1	2	3	4
Phoneme awareness	First sound identity	X	X		
	Phoneme blending	X	X		
	Phoneme segmentation	X	X	X	X
Letter knowledge	Receptive letter knowledge	X	X	X	
	Productive letter knowledge			X	
Morphology	Morphology	X	X		
	Plural form	X	X		
Rapid naming speed	Rapid naming pictures	X	X		
	Rapid naming colours	X	X		
Word recognition	Fluency of word reading I			X	
	Fluency of word reading II				X
Phonological recoding	Fluency of pseudoword reading				X
Spelling	Spelling				X

Effect measurement

Table 2.3 presents an overview of constructs, measures used for effect measurement, and their occasion of administration.

At pre- and post-test (time 1 and time 2) we measured phoneme awareness and letter knowledge to investigate whether the experimental programme gave results on the trained skills. Morphology measures were included to investigate results of the non-specific programme and rapid naming measures were included to investigate whether training effects were specific. Because the children did not receive formal reading instruction yet, we did not include reading or spelling measures. In first grade, after several month of reading instruction (time 3 and time 4) we measured reading and spelling to investigate whether the specific training effects resulted in transfer effects on these skills. If possible, we also included measures of phoneme awareness and letter knowledge at time 3 and time 4, but several measures could not be used as ceiling effects were to be expected. Measures used for effect measurement are described below.

First sound identity. This measure (de Jong, van Otterloo, & Regtvoort, 2006) was developed after the example of the 'end-sound identity task' of de Jong, Seveke, and van Veen (2000). The child had to choose from four alternatives the word that had the same first sound as a given word (e.g. *saw* begins with /s/). Each item consisted of a row of five pictures. The first picture represented the given word and was separated from the other pictures by a vertical line. The experimenter told the names of the pictures and also the first sound of the given word. The child was instructed to identify the picture of the word with the same sound at the beginning (e.g. *tree*, *sun*, *ball*, or *comb*). The test consisted of 2 practice items and 10 test items and had a maximum score of 10. Reliability for this test was 0.83.

Phoneme blending. Toets voor Auditieve Synthese [Test for phoneme blending] (Verhoeven, 1993a). The experimenter gave the child a word in separate phonemes and asked the child to blend the phonemes and pronounce the correct word (e.g. /p/-/l/-/a/-/n/-/t/ makes *plant*). The maximum score was 20. Cronbach's alpha was above 0.85 (Verhoeven, 2000).

Phoneme segmentation. Toets voor Auditieve Analyse [Test for phoneme segmentation] (Verhoeven, 1993b). The experimenter gave a whole word and the child was required to segment the word in separate phonemes (e.g. *raam* [window] makes /r/-/a/-/m/). The maximum score was 20. Cronbach's alpha was above 0.85 (Verhoeven, 2000).

Receptive letter knowledge. Receptive knowledge of 32 letters or digraphs was tested. Letters with a very low frequency in Dutch (e.g. c, q, x, y) were not presented. The experimenter showed the child a sheet with rows with six printed lowercase letters each. Then she gave a speech sound and asked the child to indicate the printed letter in that row that matched the sound. Every correct answer gave one point (Verhoeven, 2002). The reliability (Cronbach's alpha) in second kindergarten year was 0.87—0.89 (Eleveld, 2005).

Productive letter knowledge. In the 'Grafementoets'[Grapheme test] (Verhoeven, 1993a) the child has to read out loud 34 separate graphemes. We measured both accuracy and speed. The maximum accuracy-score was 34. Cronbach's alpha was above 0.85 (Verhoeven, 2000).

Morphology. This test was part of the Taaltests voor Kinderen [Language test for children] (van Bon, 1982), Woordvormen-beoordelingstest [Test for the judgement of word forms]. The child listens to an audiotape to hear a man's voice as well as a woman's voice say a particular sentence. The man or the woman makes a mistake and the child has to say who did. The tests covered subjects such as formation of the plural, past participle, past tense, comparative degree, and superlative degree. There were 3 practice items and 48 test items. The maximum score was 48. The 'Taaltest voor Kinderen' was evaluated as reliable (> 0.70) by the Dutch Committee for test affairs (Evers *et al.*, 2002).

Plural form. This test was part of the TAK, Taaltoets Allochtone Kinderen [Language test for children of ethnic minorities] (Verhoeven & Vermeer, 1996). The experimenter shows the child two pictures. On the first picture, the child can see a single object (e.g. a ship), on the second picture the child can see more than one of the same object (e.g. three ships). The experimenter tells the name of the object and stimulates the child to say the plural form. ("Here you see one ship and here you can see three .....") The Plural form tests consisted of 2 practice items (both regular plurals) and 12 test items (mostly regular plurals: three times ending on -en, three times ending on -den, and three times ending on -s, and three times an irregular plural). The maximum score is 12. The 'Taaltoets Allochtone Kinderen' was evaluated as sufficiently reliable (> 0.60) by the Dutch Committee for test affairs (Evers, van Vliet-Mulder, & Groot, 2000).

Rapid naming pictures. The experimenter showed the child a sheet with five rows of 10 pictures. There were five different pictures (a bike, a tree, a fish, a bed, and a chair) represented in a varying order. The experimenter required the child to name the pictures as quickly as possible and measured the time. The time per item (in seconds) was calculated. Split-half reliability for kindergarten was 0.73. For grade one the mean of split-half and test–retest reliability was 0.81 (van den Bos, 2003).

Rapid naming colours. This measure was similar to the Rapid naming-pictures measure, but with coloured blocks instead of pictures. There were five different colours used: black, yellow, red, green, and blue. Split-half reliability for kindergarten was 0.80. For grade one the mean of split-half and test–retest reliability was 0.88 (van den Bos, 2003).

Fluency of word reading I. Word recognition at time 3 was measured by the first test-card (1C) of the Drie-Minuten-Toets [Three-minute test] (Verhoeven, 1995), which contains five columns of CV (consonant-vowel), VC, and CVC words, increasing in difficulty. The children are instructed to read the words correctly and as quick as possible. Cronbach's alpha was 0.88 (Verhoeven & van Leeuwe, 2003).

Fluency of word reading II. Word recognition at time 4 was measured with the use of the cards 1A, 2A, and 3A of the Drie-Minuten-Toets [Three-minute test] (Verhoeven, 1995). Card 1A contained five columns of CV (consonant-vowel), VC, and CVC words, increasing in difficulty. Card 2A also contained five columns with one-syllable words but this time with consonant clusters (CCVC, CVCC, CCVCC, and CCCVC words). Card 3A contained four columns with words with more than one syllable. For every card the instruction was the same. The children were asked to read the words correctly and as quick as possible. The score per card was the number of words read correctly in 1 min. Cronbach's alpha of the first, second, and third card was respectively 0.88, 0.94, and 0.92 (Verhoeven & van Leeuwe, 2003).

Fluency of pseudoword reading. See section about parents.

Spelling. We used the dictation E3A from the 'Schaal Vorderingen in Spellingvaardigheid 1' [Progress in spelling skill scale] (van den Bosch, Gillijns, Krom, & Moelands, 1993). The experimenter dictated a target word, which the child had to write down. The dictation consisted of two parts. The first part contained 20 one-syllable words with consonant clusters (CVCC, CCVCC, VCCC, CVCCC, and CCCVC words). The second part was a bit easier and contained 17 one-syllable words (CVC, CVCC, and CCVC words). Every correctly spelled word gave 1 point, with a maximum of 37. Reliability of the test was 0.87 (Moelands & Kamphuis, 2001).

2.2.5 Procedure

An instruction meeting was organised separately for the participating parents of the experimental programme and for the participants of the control programme. During these meetings the material was presented, the aims of the programmes were explained, and exercise were demonstrated and discussed. We motivated the participants to read the instructions carefully before starting the lesson. The parents also had the opportunity to ask questions. Participants who were not present at the meeting were contacted individually.

All participants were phone called in the first 2 weeks to give them the opportunity to ask questions and to find out whether there was any lack of clarity. After about 4 weeks the researchers organised two other meetings. During these meetings the second part of the programmes were presented and questions were answered.

There were several experimenters involved in the testing of the children. The children were tested individually in a separate room at the school. The test sessions lasted about 20–40 min each. Both pre- and post-test (March and June 2002, kindergarten) were split in two sessions. For both follow-ups one session was sufficient. The tests were presented in fixed order.

2.3 Results

2.3.1 Manipulation check and usability of the programmes

Almost all participants, of both the experimental and the control programme, were in general very positive about the first week of the intervention. After 3 weeks there were more individual differences, mainly in the experimental programme. After 7 weeks many parents working with the experimental programme found it hard to keep up working with the programme every day. The control programme seemed a bit less demanding, with fewer participants reporting decreasing motivation. The children in both conditions showed more individual differences: some stayed quite enthusiastic, and some were hard to motivate. The instruction of both programmes was sufficient, according to the parents, and most parents found the programmes sufficiently diverse. The participants hardly reported problems with their role of tutor. Most parents working with the experimental programme found the first week, a general introduction, quite easy for their children. Later on, the level of difficulty was more suitable and individual differences became larger. The participants of the control programme found the exercises focussing on comprehension of text quite easy. By contrast, they found some morphology and syntax exercises more difficult. Overall, there seemed to be fewer individual differences in the participants of the control programme. The lessons of the experimental programme took about 10 min a day. The range was three till 30 min. This differed between subjects and lessons. The parents in this group were content with the duration of the daily training. When card games were involved, the duration was the longest. The lessons of the control programme took a bit less time, about 8 min a day, with a range of four till 15 min. Most parents were content with the duration of the daily training, but some argued that it was too short. About 60% of the participants in both conditions worked with the programme in the evening after dinner or before bedtime. Most other people worked with the programme in the afternoon after school or before dinner. Only a few people preferred the mornings or irregular times.

We may conclude that parents of both the experimental group and the control group were sufficiently willing and able to put effort in the tutoring, but some reported decrease in motivation over time. Overall, the level of difficulty, the instruction and the diversity of both programmes seemed sufficient. The reactions of parents and children hardly depended on the kind of programme except that the control programme took a bit less time and seemed a bit less demanding.

2.3.2 Effect measurement

Analyses

Group differences at pre-test (time 1), post-test (time 2), and at both follow-ups (time 3 and time 4) were investigated with independent t tests. Because several variables were not normally distributed we controlled whether non-parametric Mann-Whitney U tests gave different results, which was not the case. Therefore t-test results are represented. Immediate training effects at group level (interaction of time $\times$ condition) between pre- and post-test were analysed with repeated measures analyses of variance (MANOVA) with 'time' as within factor and 'condition' as between factor. We also used repeated measures analyses of variance to analyse differences in progress on the kindergarten measures that were repeated at follow-up (time 3 and time 4). To judge the importance of the effects, we also calculated the corresponding effect sizes. The effect size η^2, usually given in MANOVAs, was transformed in the more common Cohen's *d* to be able to compare the results with other studies.

In order to gain insight into the meaning of the effects at the individual level, we also calculated the distribution of the participants (in proportions) over different percentile classes at time 4. These percentile classes were based on a normative sample of educational age controls. *Post hoc* we examined if there were differences in treatment effects on subgroups.

Pre-test (time 1)

The results on all pre- and post-test measures are presented in Table 2.4. The groups did not differ on the pre-test measures, except for a significant difference in morphology, $t(46) = 2.72$, $p < 0.01$. The controls scored significantly better than the children from the experimental group. Because morphology at pre-test correlated significantly with several other pre- and post-test measures, it was included as an extra between factor in the MANOVAs. To do so a dichotomous morphology variable was made based on the median. Because no interaction with morphology was significant, MANOVAs without this factor were interpreted.

Post-test (time 2)

As expected, the experimental group did better on receptive letter knowledge at the post-test, $t(46) = 2.14$, $p < 0.05$, $d = 0.59$, and the control group did better on morphology, $t(46) = 2.96$, $p < 0.01$, $d = 0.83$, and plural form, $t(46) = 3.92$, $p < 0.001$, $d = 1.16$. Table 2.4 shows main effects of time for nearly all measures, except for plural form and rapid naming colours. We found significant interactions of time × condition (two repeated measurements) for two phoneme awareness measures: first sound identity and phoneme blending (see Table 2.4 for F values). As expected, the experimental children made more progress on these measures than the controls did. It should be noted that the first sound identity measure at time 2 was skewed to the left, skewness (SE) = −1.17 (0.34). This tends to a ceiling effect, which means that the test does not sufficiently discriminate between children with high test scores. The training effect on first sound identity might therefore be underestimated. Contrary to the expectations, we did not find any significant interactions for phoneme segmentation or for receptive letter knowledge.

Table 2.4: Means and standard deviations on pre- and post-test measures, MANOVA results and effect sizes

Measures	Experimental group $n = 23$				Control group $n = 25$				Main effect	Interaction	Effect size
	Pre-test (Time 1)		Post-test (Time 2)		Pre-test (Time 1)		Post-test (Time 2)		Time	Time × condition	Interaction Time × condition
	M	SD	*M*	SD	*M*	SD	*M*	SD	*F*(1, 46)	*F*(1, 46)	d^a
Recep. letter knowledge	13.78	6.68	17.91	5.39	11.57	6.21	14.40	5.94	32.32***	1.12	0.27
First sound identity	5.92	2.61	8.22	2.33	6.75	2.66	7.68	2.21	31.27***	5.63*	0.70
Phoneme blending	4.83	4.55	9.35	5.69	7.81	6.05	9.12	5.61	30.18***	9.14**	0.89
Phoneme segmentation	2.83	3.68	5.17	5.34	4.44	4.38	5.88	4.23	10.58**	0.61	0.23
Morphology	26.26	8.08	30.17	6.00	31.76	5.85	35.48	6.38	22.68***	0.02	0.06
Plural form	8.91	1.28	8.70	1.26	9.28	1.17	10.08	1.19	1.44	4.39*	0.62
Rapid naming pictures	1.79	0.45	1.60	0.39	1.80	0.59	1.56	0.47	13.69***	0.14	0.11
Rapid naming colours	1.79	0.52	1.62	0.35	1.80	0.60	1.77	0.55	3.05	1.50	0.36

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$
Note. Recep. = Receptive
a Transformation of η^2 into Cohen's *d* using the formula: $d = \left((4\eta^2)/(1-\eta^2)\right)^{1/2}$ (see Rosenthal, Rosnow, & Rubin, 2000, p.15)

In the final column of Table 2.4, effect sizes of the interaction of time × condition are presented. According to Cohen's criteria (1988), the effect size with regard to first sound identity and phoneme blending can be called moderate (d = 0.70) and large (d = 0.89), respectively. The three phoneme awareness measures together (first sound identity, phoneme blending, and phoneme segmentation) also showed a significant interaction of time × condition, $F(3, 44) = 4.80$, $p < 0.01$. This corresponds with a large effect size (d = 1.15).

Plural form showed a significant interaction of time × condition (Table 2.4). This corresponds with a moderate effect size (d = 0.62). As expected, the control group made more progress than the experimental group on this measure. Actually, the scores of the experimental group slightly decreased over time. With a t test for paired samples, however, we could demonstrate that the difference between pre-test and post-test on this measure was significant for the control group, $t(24) = 2.41$, $p < 0.05$, but not for the experimental group. The groups did not differ in the progress they made on morphology. As expected, the groups did not differ in their scores or progress on the rapid naming measures, either.

January grade one (time 3)

Table 2.5 shows the results of the analyses on the first grade measures tested in January (time 3). Although the experimental groups seemed to score better on all measures, independent t tests showed that the groups did not differ significantly on any measure. In terms of effect sizes, the differences were small or moderate-to-small. It should be noted that all accuracy measures of phoneme awareness and letter knowledge were skewed to the left. This was the case for receptive letter knowledge, skewness (SE) = –2.06 (0.36), phoneme blending, skewness (SE) = –1.66 (0.36), phoneme segmentation, skewness (SE) = –0.99 (0.36), and productive letter knowledge accuracy, skewness (SE) = –1.23 (0.36). However, there was still a significant interaction of time × condition (three repeated measurements) on phoneme blending (Table 2.5). This corresponds with a moderate to large effect size (d = 0.76). As expected, the experimental group made more progress in time. Contrast analysis (simple contrast) showed that time 1 versus time 2 gave a significant result, $F(1, 41) = 9.69$, $p < 0.01$, as we already pointed out. The same was true for time 1 versus time 3, $F(1, 41) = 9.17$, $p < 0.01$. We were also able to show a significant interaction of time × condition (three repeated measurements) on phoneme segmentation

Table 2.5: Means and standard deviations on January measures in grade one, MANOVA- and *t*-test outcomes and corresponding effect sizes

Measures	**Exp. group** $n = 20$		**Control group** $n = 23$		**Independent** ***t*** **test**	**Effect size January, grade one**	**Time**[a]	**Time × condition**[a]	**Effect size Time × condition**
Follow-up I (Time 3)	*M*	SD	*M*	SD	$t(41)$	d [b]	$F(2, 82)$	$F(2, 82)$	d [c]
Rec. letter knowledge	30.25	1.37	29.45	2.52	1.44	0.32	207.89***	0.95	0.31
Phoneme blending	18.30	1.72	17.70	2.53	0.90	0.24	147.65***	5.84**	0.76
Phoneme segmentation	16.25	3.26	14.43	4.98	1.43 [d]	0.37	165.65***	3.77*	0.61
Fluency of word reading	16.81	7.77	16.00	7.01	0.44	0.12	–	–	--
Prod. letter knowl. accur.	30.81	3.79	30.13	3.21	0.49	0.21	–	–	--
Prod. letter knowl. speed	40.45	15.5	44.61	15.9	0.87	0.26	–	–	--

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Note. Rec. = Receptive, Prod. = Productive, knowl. = knowledge, accur. = accuracy, Exp. = Experimental

a Sphericity assumed

b Calculation based on control group SD

c Transformation of η^2 into Cohen's *d* using the formula: $d = \left((4\eta^2)/(1-\eta^2)\right)^{1/2}$ (see Rosenthal *et al.*, 2000, p.15)

d Levene's test indicated that variances were significantly different, corrected *t* test: $t(38.26) = 1.43$

Table 2.6: Means and standard deviations on June measures in grade one, *t*-test outcomes and corresponding effect sizes

Measures	Exp. group $n = 19$		Contr. group $n = 23$		Independent *t* test		Effect size June, grade one
Follow-up II (Time 4)	*M*	SD	*M*	SD	*t*	*df*	d^{a}
Phoneme segmentation	18.89	2.40	17.26	4.83	1.41^{b}	33.73	0.34
Fluency of word reading II	21.02	13.59	17.32	12.02	0.94	40	0.31
Fluency of pseudoword reading	17.00	13.66	13.00	6.32	1.15^{b}	22.68	0.63
Spelling	30.79	4.55	28.27	6.98	1.55	40	0.36

Note. Exp. = Experimental; Contr. = Control
a Calculation based on control group SD
b Levene's tests indicated that variances were significantly different

(Table 2.5). This corresponds with a moderate effect size (d = 0.61). The experimental group made more progress in time. A contrast analysis (simple contrast) showed that time 1 versus time 2 did not show any effect, but that time 1 versus time 3 did, $F(1, 41) = 6.99, p < 0.05$. The children did not benefit right after training, but seemed to do so later on.

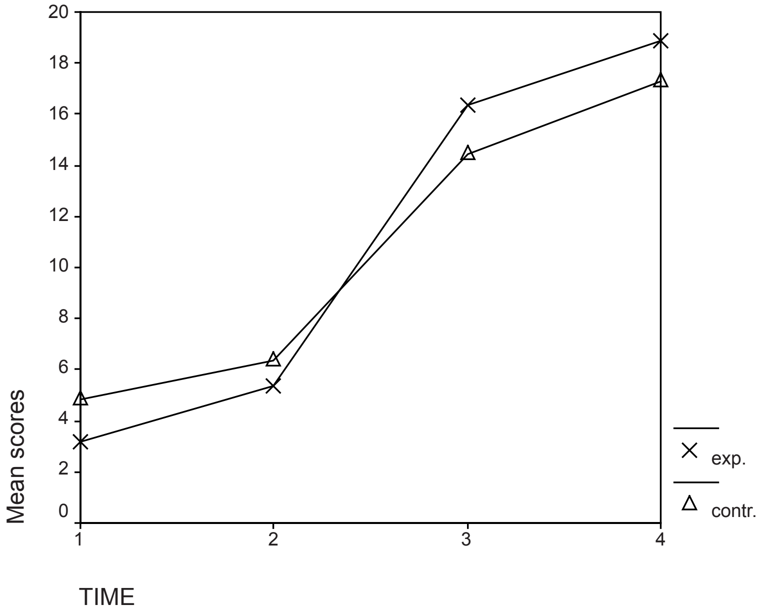
June grade one (time 4)

Table 2.6 shows the results of the second follow-up in first grade (time 4). Again the experimental group seemed to do better on all measures, but independent *t* tests indicated that these differences were not significant. In terms of effect sizes, the differences were small or moderate. The effect size on pseudoword reading (d = 0.63) is relatively large because we used the control group SD to calculate this effect size, which is common in intervention studies. In this case, the control group SD (6.32) was considerably smaller than the experimental group SD (13.66), which explains probably why this moderate effect did not result in a statistical significant group difference.

There was a ceiling effect for phoneme segmentation (skewness = –2.62, SE = 0.37). Many children have the highest possible test score. The test does therefore not discriminate between these children.

Repeated measures analyses of variance still indicated a significant interaction of time × condition (four repeated measurements) on phoneme segmentation, $F(3, 117) = 2.91, p < 0.05, d = 0.55$. The effect size of the interaction

Figure 2.1: Progress in phoneme segmentation



is moderate. Over time the experimental group made more progress than the control group. A contrast analysis (simple contrast) showed that although a significant interaction of time × condition from time 1 to time 4 was detected, $F(1, 39) = 4.47$, $p < 0.05$, this was only caused by the interaction between time 2 and time 3. There is no significant interaction of time × condition between time 1 and 2, as we already pointed out, and between time 3 and 4, which is in line with the ceiling effect we described earlier. Figure 2.1 illustrates the differences in progress. At time 1 and time 2 the variable was skewed to the right, skewness (SE) was 1.18 (0.34) and 0.94 (0.34), respectively, and at time 3 and time 4 the variable was skewed to the left as we already pointed out. This shows that the test is too difficult in kindergarten (time 1 and time 2) and too easy in first grade (time 3 and time 4). Apparently, most Dutch children learn the skill of phoneme segmentation in the first part of first grade.

We did not find a significant interaction of time × condition on fluency of word reading (two repeated measurements of reading card 1), $F(1, 40) = 1.51$, ns, $d = 0.07$. This means there was no significant group difference in word reading progress during the second half of grade one.

To analyse the scores of the two groups on the reading and spelling-measures at time 4 more specifically, we calculated the distribution of the participants in proportions over five percentile classes (Table 2.7). These percentile classes were based on the scores of normative samples of unselected children with the same educational age, as presented by the test constructors (van den Bosch *et al.*, 1993; van den Bos *et al.*, 1994; Verhoeven, 1995). To give a general picture, we also calculated the distribution over percentile classes including all three measures. There were five separate percentile classes: class A includes percentile 76–100, class B includes percentile 51–75, class C includes percentile 26–50, class D 11–25, and class E percentile 10 or below.

Compared with the normative sample, both experimental and control group contained fewer children in class A and B (on average 21% instead of the expected 50% for unselected children). However, there was a difference between the groups in the present study: about 29% of the experimental group scored in the top half of the scores, in contrast to only 14% of the control group. With regard to the lowest 25% of the normative sample, the proportions were in line with the assumption that children from at-risk groups should be overrepresented in these percentile classes. However, the difference between the two groups in the present study was small: 40% of the experimental group and 44% of the control group. Looking at the separate measures, there was one finding clearly in favour of the experimental group. On fluency of word reading, the experimental group contained fewer children (11%) who scored beneath the 10th percentile than the control group (25%).

2.3.3 Differences between subgroups

Because effects might be related to initial differences, especially in an at-risk population, effects of subgroups were investigated. In the introduction we stressed that letter knowledge was one of the most striking predictors of beginning reading skills. Therefore, the data of pre-test and post-test were *post hoc* divided in two based on the letter knowledge of the children at pre-test (time 1). This resulted in two subgroups, a group with high letter knowledge

Table 2.7: Distribution in proportions over different percentile classes on reading and spelling measures at the end of first grade (time 4)

Percentile classes	Percentile	Fluency of word reading II		Fluency of pseudoword reading		Spelling		Overall	
		Exp. $n = 19$ %	Contr. $n = 23$ %	Exp. $n = 18$ %	Contr. $n = 23$ %	Exp. $n = 19$ %	Contr. $n = 22$ %	Exp. $n = 18/19$ %	Contr. $n = 22/23$ %
A	> 75	14	7	17	4	16	9	16	7
B	$> 50 – 75$	12	8	11	4	16	9	13	7
C	$> 25 – 50$	33	40	28	48	32	36	31	41
D	$> 10 – 25$	30	20	22	26	21	27	24	24
E	≤ 10	11	25	22	17	16	18	16	20

Note. Exp. = Experimental group; Contr. = Control group; Overall = mean of three measures

(12 letters or more) and a group with low letter knowledge (less than 12 letters). The group with high letter knowledge consisted of 25 children, 15 from the experimental group, and 10 from the control group. The group with low letter knowledge consisted of 23 children, 8 from the experimental group, and 15 from the control group. We put letter knowledge as an extra between factor in the repeated measures analyses of variance on phoneme blending and first sound identity with time (pre-test and post-test) as within factor and condition as the other between factor. We chose for these measures of phoneme awareness because of the significant interactions of time × condition on these measures (see Table 2.4). A significant three-way interaction of time × condition × letter knowledge was shown on phoneme blending, $F(1, 44) = 6.68$, $p < 0.05$, $d = 0.79$, and a tendency to a significant three-way interaction of time × condition × letter knowledge was shown on first sound identity, $F(1, 42) = 6.88$, $p = 0.07$, $d = 0.55$. Separate analyses on subgroups showed that in the group children with low initial letter knowledge there was no interaction of time × condition on phoneme blending and first sound identity, whereas in the group children with high initial letter knowledge there was. This indicates that only children with sufficient prior letter knowledge profited from the experimental intervention.

Another way to investigate differences between subgroups is to analyse retrospectively if good and poor readers benefited equally from the experimental intervention. To do so, the data were split in two based on the reading level of the children at the end of first grade (time 4). This resulted in two subgroups. One group consisted of 21 relatively good readers, 10 from the experimental group, and 11 from the control group, and one group consisted of 21 relatively poor readers with 9 children from the experimental group and 12 children from the control group. 'Reading level' was put in the repeated measures analyses of variance on phoneme blending and first sound identity as an extra between factor. We did not find significant three-way interactions of time (pre-test versus post-test) × condition × reading level for both phoneme blending, $F(1, 38) = 0.67$, $p = 0.42$, $d = 0.26$ as first found identity, $F(1, 37) = 0.42$, $p = 0.52$, $d = 0.21$. We therefore concluded that both relatively good readers and relatively poor readers in our sample had benefited equally from the experimental intervention.

2.4 Discussion

As expected, the experimental group showed more progress in phoneme awareness immediately after training than the control group. Although the experimental group knew more letters at the post-test in comparison to the control group, the groups did not differ in their progress in letter knowledge, which was contrary to our expectation. The control group, trained with a non-specific programme based on morphology and vocabulary, showed the expected progress on the plural form measure but not on the other morphology measure. Because the experimental programme and the control programme affected different skills, we may conclude that the immediate effects were specifically bound to the intentions of the programmes. This conclusion is also supported by the fact that neither of the programmes had an effect on a reading-related but not-practiced skill (rapid naming). In addition, we may conclude that parents from children at familial risk can teach their children phoneme awareness successfully.

In first grade, at time 3 and time 4, the effect of the experimental intervention on phoneme awareness was still present, but less strong. Contrary to our expectations, we did not find a significant transfer effect of the experimental intervention on phonological recoding and spelling in first grade when we compared average results. We also did not replicate the finding of Lyster (2002) that the morphology trained group (our control group) would do better on word recognition at the end of grade one. When we look at the effectiveness of training in terms of effect size our findings indicate that the experimental group did a bit better, although not significantly better, on spelling, phonological recoding, and even word recognition. We found it justified interpreting these statistical insignificant results, as the effects are in the expected direction and the effect sizes are comparable with those of related studies. This is supported by the repeated appeal (e.g. Kirk, 2001; Wilkinson & the Task Force on Statistical inference, 1999) that the focus of research should not be mainly on statistical significance, but on effects size and practical significance.

The results of the present study are not very different from the results of other studies when we look at the effectiveness. We found a large overall effect of the experimental training on phoneme awareness immediately after training (d = 1.15), somewhat larger than is indicated by meta-analyses (Bus & van IJzendoorn, 1999: d = 1.04; Ehri *et al.*, 2001: d = 0.86). It should be noted that there are differences in the manner of calculation of the effect sizes on phoneme awareness, as in the meta-analyses outcome effect sizes are presented and we corrected for group differences at the pre-test. In spite of these differences we find it justified to compare the effect sizes as they both reflect the impact of the intervention. The largest effect of training was on phoneme blending, which was large immediately after training (d = 0.89), and continued in January of first grade (d = 0.76). This was to be expected because of the emphasis of the experimental programme on phoneme blending skill.

The outcome effect of the experimental programme on fluency of word reading (word recognition) was practically zero in January first grade and only small in June in terms of effect size (d = 0.31), and the effect on pseudoword reading (phonological recoding) was moderate (d = 0.63) (neither effect was significant). These findings are not very different from the results of the meta-analyses that reported moderate to small effect sizes of training on reading measures at post-test (Bus & van IJzendoorn, 1999: d = 0.44; Ehri *et al.*, 2001: d = 0.53) and small effects at follow-up (Bus & van IJzendoorn: d = 0.16, *ns*, after 18.5 months on average; Ehri *et al.*: d = 0.45 at first follow-up, 2–15 months delay, and d = 0.25 at second follow-up, 7–37 months delay). However, we must remark that the long-term effects (follow-up in Bus & van IJzendoorn and second follow-up in Ehri *et al.*) were based on only eight studies, and effects differed considerably across studies.

Investigation of the effects of phoneme awareness training on spelling is less common than on reading. Although we found a significant time × condition interaction with respect to phoneme segmentation in January of first grade (d = 0.61), the outcome effect size at that moment was not large (d = 0.37). In addition, there was no significant transfer effect on spelling (d = 0.36). Ehri *et al.* (2001) reported a moderate effect on spelling immediately after training (d = 0.59). However, there were only small effects at first and second follow-up (d = 0.37 and 0.20), quite comparable to our study. Bus and van IJzendoorn also reported small transfer effects on spelling at follow-up after 18.5 months on average (d = 0.25). Again, these effect sizes were based on a relatively small number of studies, 17, 6, and 8, respectively. It seems that training of phoneme awareness and letter knowledge in the pre-reading phase affects the skill in reading and spelling at follow-up, but only to a small degree. However, sample size in the present study was insufficient to detect small effects with a chance level below 5%.

Another way of looking at the effectiveness is to take the results at the individual level into account (Table 2.7). With regard to fluency of word reading (word recognition) the distribution over percentile classes within the experimental group seems to have moved up: more students performed above the 50th percentile and fewer students performed below the 10th percentile in comparison to the control group. With regard to fluency of pseudoword reading (phonological recoding) the moderate effect size (d = 0.63) of the group difference, although not significant, may be interpreted as support for the hypothesis that the experimental training has an effect on phonological recoding. A closer look at Table 2.7, however, suggests that a larger proportion of experimental children in the higher levels (28 versus 8% in percentile classes A and B) caused this group difference. On the lower levels the proportions hardly differed. This suggests that the weakest children did not benefit from the experimental training for phonological recoding, whereas less weak children did. However, retrospective analyses on subgroups separating relatively good readers from relatively poor readers at the end of first grade indicate that this was not the case.

Although the experimental training focussed on letter knowledge and phoneme awareness, we only find training effects on phoneme awareness, and not on letter knowledge. This is similar to the study of Hindson *et al.* (2005), who also reported an effect on phoneme awareness but not on letter knowledge. A plausible explanation for the fact that we did not find a training effect on letter knowledge could be that the children already knew approximately 13 letters at pre-test, although the dispersion was quite large (M = 12.63, SD = 6.47). The 12 letters we introduced in the experimental programme contained mostly highly frequent, easy to learn letters such as ‘m’ and ‘oo’. More letters or more lower-frequent letters in the programme could possibly have given larger results.

However, in line with the suggestions by Bradley and Bryant (1985), it may be assumed that a training programme does not need to cover all letters to stimulate phoneme awareness. Working with the letters as a visual symbol for each phoneme may have helped to develop phoneme awareness (Adams, Treiman, & Pressley, 1998). On the other hand, when the data of the children in both experimental conditions were *post hoc* split in two based on their initial letter knowledge, children with low initial letter knowledge did not seem to benefit from the experimental training on phoneme blending and first sound identity, whereas children with high initial letter knowledge did so considerably. This indicates that letter knowledge is an important factor in the development of phoneme awareness. Our finding supports the reciprocal relationship between phoneme awareness and letter knowledge as indicated by several studies (e.g. Carroll, Snowling, Hulme, & Stevenson, 2003), and the relationship between poor phonological sensitivity and delayed letter name learning in children at familial risk of dyslexia (Torppa *et al.*, 2006).

The question remains why a training programme with sufficient theoretical validity, carried out properly and at the right time, had no significant impact on reading skill of at-risk children. Comparisons to related studies may shed some light on this. Because the term ‘at-risk’ in the literature covers a variety of definitions, it is best to compare effect sizes of the present study to effect sizes of the study of Elbro and Petersen (2004) who also trained children familial at risk in kindergarten. Moreover, their programme was adapted by us but was essentially comparable in content. At follow-up (beginning of grade two) they reported moderate effects on all reading measures (range d = 0.45–0.69), while in the present study at the end of grade one effect sizes were found that were in the same direction but a bit smaller (range d = 0.31–0.63). However, their findings indicated significant differences at the group level but ours did not. Obviously, sample size (82 in their study, 43 in ours) may have contributed to the differences in significance but probably other differences between the studies did so as well.

It seems unlikely that the difference in results in the present study can be explained by the way the children were selected. Selection criteria were comparable to the Danish study, although no parents were selected because of a large disparity between reading skill and verbal competence (in the present study only about 10% of the parents were selected with this criterion). As a strong indication of a familial factor, Table 2.2 shows that in many cases (roughly 50%) reading problems were present in other family members as well.

Possibly the most striking difference between the two studies relates to the proportion of children with reading scores within the lowest 10% of normative samples of unselected children. Elbro and Petersen (2004) report a large difference on a composite measure between their trained and untrained at-risk groups (respectively 19% en 47%) at the start of grade two while in the present study the proportions at the end of grade one were 16% and 20% (using the overall score as a comparable composite measure, see Table 2.7).

The proportion of at-risk control children with scores on reading and spelling measures in percentile class E in the present study (see the overall-measure in Table 2.7) was much smaller than in the Danish study. Because we may assume that the selection criteria were equal (indicating the same familial risk) and that times of measurement were by and large comparable (at the beginning of grade two and at the end of grade one), the fact that 47% of the Danish control children developed a considerable reading delay in contrast to only 20% of our control children may indicate that the chance of developing dyslexia is larger in Denmark than in the Netherlands, at least in the phase of reading acquisition. Because formal reading instruction in Denmark starts when the children are a year older than in the Netherlands, the larger amount of low-scoring at-risk children in Denmark may not be attributed to cognitive developmental level or socio-emotional 'maturity'. Instead, several environmental factors may explain these findings.

As we already pointed out in the introduction, the Danish orthography is less shallow than the Dutch (Seymour *et al.*, 2003). As a consequence, acquiring reading and spelling may be harder in Danish than in Dutch. Furthermore, at the pre-test, the Danish children knew fewer letters than the Dutch children. Possibly, in the pre-reading phase less attention is paid to emergent literacy skills in Denmark than in the Netherlands. There may be more to gain in Denmark than in the Netherlands by instruction and practice in that phase. Moreover, the quality and quantity of Dutch reading instruction in grade one might be better. In most schools in the Netherlands the phonics approach is used, as in most countries with shallow orthographies (Wimmer, 1993). The whole language approach used in some other countries is not used in the Netherlands. It is to be expected that the effects of phonological awareness training will be larger if the role of phonics in regular education is smaller. This is in line with the suggestion of Lyster (2002) that the small effect of her phonological awareness training in Norway might be due to the relatively shallow orthography and the attention given to phonics in regular class.

An other explanation for the fact that less children, in both the experimental group and the control group, fell in the lowest level of reading and spelling than one might expect based on their at-risk group membership, is the good home literacy environment in our sample. In addition, in a recent Dutch study (van Otterloo & van der Leij, 2009[2]) it is indicated that in Dutch orthography, the reading problems in a sample of children familial at risk of dyslexia tends to manifest itself later on (end second grade and upwards).

A particular aspect of our intervention is that it is carried out at home by parents. It may be argued that the quality of the administration of the programmes might therefore be insufficient, as the parents are no professional treatment agents. Although in the present study our manipulation check was only based on self-report, a study on treatment integrity (van Otterloo, van der Leij, & Veldkamp, 2006) strongly indicated that a similar group of parents was able to carry out a similar home-based intervention with sufficient quality and quantity. In that study, treatment integrity contributed to the prediction of the pre-reading skills at post-test. The quantity of the administration of the programme was a larger predictor than the quality of administration. This indicated that it is important that the programme is continued for a longer time. However, the fact that the quality of the administration was sufficient may be influenced by our sample. Most parents in our sample were motivated, highly educated, and the home literacy environment was relatively adequate. The way we implemented the programme might therefore not be suitable for all parents. In the study on treatment integrity, (van Otterloo, van der Leij, & Veldkamp), it is argued that the implementation could be adapted more to the needs of the parents by, for example, giving less educated parents more support to enhance the quality of the home literacy environment. Because the home

---

[2] Chapter 4 of this thesis

literacy environment was comparable between the experimental conditions (see Table 2.2), we may assume that home literacy environment did not influence the results of the current study.

To conclude, the present study indicates that in the Netherlands early home-based training of phoneme awareness and letter knowledge for kindergartners at familial risk of dyslexia leads to significant gains in phoneme awareness. Evidently, parents of children at familial risk were able to teach their children phoneme awareness. This claim is not hampered by amount of training, treatment integrity or sample size, as the immediate results on phoneme awareness were considerable. Parents were also able to teach their children – at least some – morphology, given the progress that the children from the control group made on the plural form measure.

The experimental training had no effect on letter knowledge. A head start in phoneme awareness alone does not seem to be sufficient for significant effects on reading and spelling measures, as first grade transfer effects on phonological recoding and spelling were small (ns) and the transfer on word recognition was moderate (ns). Moreover, prior letter knowledge influences the development of phoneme awareness in the experimental group. A more intense training, including more letters and low-frequent letters, may possibly give an effect on letter knowledge, larger effects on phoneme awareness, and larger transfer effects on reading and spelling measures in first grade.

2.5 References

Adams, M. J., Treiman, R., & Pressley, M. (1998). Reading, writing and literacy. In I. E. Sigel, & K. A. Renniger (Eds.), *Handbook of child psychology: Child psychology in practice* (Vol. 4, pp. 275–355). New York: Wiley.

Arnold, D. H., Lonigan, C. J., Whitehurst, G. J., & Epstein, J. N. (1994). Accelerating language development through picture book reading: Replication and extension to a videotape training format. *Journal of Educational Psychology, 86*, 235–243.

Ball, E. W., & Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26*, 49–66.

Borgwaldt, S. R. (2003). *From onset to entropy; spelling-pronunciation patterns in six languages.* Doctoral thesis, Amsterdam, University of Amsterdam.

Borstrøm, I., & Elbro, C. (1997). Prevention of dyslexia in kindergarten: Effects of phoneme awareness training with children of dyslexic parents. In C. Hulme & M. Snowling (Eds.), *Dyslexia: Biology, cognition and intervention* (pp. 235–253). London: Whurr.

Borstrøm, I., & Petersen, D. K. (1996). *På vej til den første læsning. Fonologisk opmærksomhed [On the way to initial reading. Phonological awareness].* Copenhagen: Alinea.

Bradley, L., & Bryant, P. (1985). *Rhyme and reason in reading and spelling*: Ann Arbor: University of Michigan Press.

Brus, B. T., & Voeten, M. J. M. (1980). *Eén-minuut-test, vorm a en b. Verantwoording en handleiding. [One-minute-test Justification and manual].* Nijmegen: Berkhout.

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.

Byrne, B., & Fielding Barnsley, R. (1993). Evaluation of a programme to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology, 85*, 104–111.

Carroll, J. M., Snowling, M. J., Hulme, C. & Stevenson, J., (2003). The development of phonological awareness in preschool children. *Developmental Psychology, 39*, 913–923.

Cary, L., & Verhaeghe, A. (1994). Promoting phonemic analysis ability among kindergartners: Effects of different training programmes. *Reading and Writing, 6*, 251–278.

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

de Jong, P. F., Seveke, M. J., & van Veen, M. (2000). Phonological sensitivity and the acquisition of new words in children. *Journal of Experimental Child Psychology, 76*, 275–301.

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

de Jong, P. F., van Otterloo, S. G., & Regtvoort, A. G. F. M. (2006). *Fonologische bewustzijns taken voor kleuters: Rijm en beginfoneem [phonological* *awareness' tasks for kindergarten children: Rhyme and first phoneme].* Unpublished manuscript, University of Amsterdam.

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.

Elbro, C., Borstrøm, I. & Petersen, D. K. (1998). Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly, 33*, 36–61.

Elbro, C., & Petersen, D. K. (2004). Long-term effects of phoneme awareness and letter sound training. An intervention study with children at risk of dyslexia. *Journal of Educational Psychology, 96*, 660–670.

Elbro, C., & Scarborough, H. S. (2004a). Early identification. In P. Bryant & T. Nunes (Eds.), *International handbook of children's reading* (pp. 339–359). Dordrecht: Kluwer.

Elbro, C., & Scarborough, H. S. (2004b). Early intervention. In P. Bryant & T. Nunes (Eds.), *International handbook of children's reading.* (pp. 361–381). Dordrecht: Kluwer.

Eleveld, M. A. (2005). *At risk for dyslexia. The role of phonological abilities, letter knowledge, and speed of serial naming in early intervention and diagnosis.* Doctoral thesis. Antwerp/Apeldoorn: Garant.

Evers, A., van Vliet-Mulder, J. C., & Groot, C. J. (2000). *Documentatie van tests en testresearch in Nederland [Documentation of tests and test research in the Netherlands].* Assen: Van Gorgum.

Evers, A., van Vliet-Mulder, J. C., Resing, W. C. M., Starren, J. C. M. G., van Alphen de Veer, R. J., & van Boxtel, H. (2002). *COTAN; Testboek voor het onderwijs [COTAN; testbook for education].* Lochem: NDC-Boom.

Fielding-Barnsley, R., & Purdie, N. (2003). Early intervention in the home for children at risk of reading failure. *Support for Learning, 18*, 77–82.

Finucci, J. M., Gottfredson, L. S., & Childs, B. (1985). A follow-up study of dyslexic boys. *Annals of Dyslexia, 35*, 117–136.

Gallagher, A., Frith, U., & Snowling, M. J. (2000). Precursors of literacy delay among children at genetic risk of dyslexia. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*, 202–213.

Gilger, J. W., Pennington, B. F., & DeFries, J. C. (1991). Risk for reading disability as a function of parental history in three family studies. *Reading and Writing, 3*, 205–217.

Hallgren, B. (1950). Specific dyslexia ('congenital word-blindness'): A clinical and genetic study. *Acta Psychiatrica et Neurologica KjoBenhavn, 65*, 287.

Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development, 65*, 41–57.

Hindson, B. A., Byrne, B., Fielding-Barnsley, R., Newman, C., Hine, D. W. & Shankweiler, D. P. (2005). Assesment and early instruction of preschool children at risk for reading disability. *Journal of Educational Psychology, 97,* 687–704.

Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement, 61*, 213–218.

Koster, C., Been, P. H., Krikhaar, E. M., Zwarts, F., Diepstra, H. D., & van Leeuwen, T. H. (2005). Differences at 17 months: Productive language patterns in infants at familial risk for dyslexia and typically developing infants. *Journal of Speech, Language, and Hearing Research, 48*, 426–438.

Lyon, G. R. (1995). Toward a definition of dyslexia. *Annals of Dyslexia, 45*, 3–27.

Lyster, S. A. H. (2002). The effects of morphological versus phonological awareness training in kindergarten on reading development. *Reading and Writing, 15*, 261–294.

Moelands, F., & Kamphuis, F. (2001). *Verantwoording van de toetsen uit de pakketten Schaal vorderingen in spellingvaardigheid (SVS) 1, 2 en 3 [Justification of the spellingtests SVS 1, 2, and 3]*. Arnhem: CITO.

O'Connor, R. E., Jenkins, J. R., Leicester, N., & Slocum, T. A. (1993). Teaching phonological awareness to young children with learning disabilities. *Exceptional Children, 59*, 532–546.

Petersen, D. K. (2000). What do you get if you add /mmm/ to ice? Training phoneme awareness in kindergarten: An intervention study with children of dyslexic parents. In N. A. Badian (Ed.), *Prediction and prevention of reading failure* (pp. 247–273). Baltimore: York Press.

Raven, J. C., Court, J. H., & Raven, J. (1984). *Manual for Raven's progressive matrices and vocabulary scales.* London: Lewis.

Regtvoort, A. G. F. M., & 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.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research. A correlational approach.* Cambridge: Cambridge University Press.

Scarborough, H. S. (1990). Very early language deficits in dyslexic children. *Child Development, 61*, 1728–1743.

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities. Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability. A view of the spectrum*. (pp. 75–119). Timonium, MD: York Press.

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 programmes. *Journal of Educational Psychology, 92*, 284–295.

Sénéchal, M., & LeFevre, J. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73(2)*, 445–460.

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

Snowling, M. J. (2000). *Dyslexia* (2nd ed.): Malden MA: Blackwell.

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

Torppa, M., Poikkeus, A.-M., Laakso, M.-L., Eklund, E., & Lyytinen, H. (2006). Predicting delayed letter knowledge and its relation to grade 1 reading achievement among children with and without familial risk for dyslexia. *Developmental Psychology, 42,* 1128–1142.

van Bon, W. H. J. (1982). *Taaltest voor Kinderen [Language test for children].* Lisse: Swets & Zeitlinger.

van Bon, W. H. J. (1986). *Raven's Coloured Progressive Matrices; Nederlandse normen en enige andere uitkomsten van onderzoek [Dutch standards and some other outcomes of research]*. Lisse: Swets & Zeitlinger.

van den Bos, K. P. (2003). *Snel Serieel Benoemen; Experimentele versie. [Rapid naming; experimental version]*. Groningen: University of Groningen.

van den Bos, K. P., lutje Spelberg, H. C., Scheepstra, A. J. M., & de Vries, J. R. (1994). *Handleiding de Klepel [Manual the clapper]*. Nijmegen: Berkhout.

van den Bosch, L., Gillijns, P., Krom, R., & Moelands, F. (1993). *Schaal Vorderingen in Spellingvaardigheid 1. Handleiding. [Scale progress in spelling skill. Manual.]*. Arnhem: CITO.

van der Leij, A. (1982). *Kijk en Luistermethode [Talking books-method]*. Gorinchem: De Ruiter b.v.

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.

van Otterloo, S. G., van der Leij, A., & Veldkamp, E. (2006). Treatment integrity in a home-based pre-reading intervention programme. *Dyslexia, 12,* 155–176.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45*, 2–40.

Verhoeven, L. (1993a). *Grafementoets. Toets voor auditieve synthese. Handleiding. [Grapheme test. Test for phoneme blending. Manual]*. Arnhem: CITO.

Verhoeven, L. (1993b). *Toets voor auditieve analyse. Fonemendictee. Handleiding. [Test for phoneme segmentation. Dictation of phonemes. Manual.]*. Arnhem: CITO.

Verhoeven, L. (1995). *Drie-Minuten-Toets. Handleiding [Three-minute test. Manual]*. Arnhem: CITO.

Verhoeven, L. (2000). Components in early second language reading and spelling. *Sientific Studies of Reading, 4*, 313–330.

Verhoeven, L. (2002). *Passieve Letterkennis [Receptive letter knowledge]*. Arnhem: CITO.

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

Verhoeven, L., & Vermeer, A. (1996). *Taaltoets Allochtone Kinderen. Handleiding [Language test for children of ethnic minorities. Manual]*. Tilburg: Zwijsen.

Wechsler, D. (1981). *WAIS-R: Wechsler Adult Intelligence Scale-Revised*. San Antonio, TX: Psychological Corporation.

Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology, 30*, 679–689.

Wilkinson, L., & the Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604.

Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics, 14*(1), 1–33.

Wimmer, H., Mayringer, H., & Landerl, K. (2000). The double-deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology, 92*, 668–680.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91*, 415–438.