The role of phonological awareness in treatments of dyslexic primary school children

Julia Pape-Neumann[1,2]*, Muna van Ermingen-Marbach[3,6], Marion Grande[1], Klaus Willmes[4], and Stefan Heim[3,5,7]

[1]Division for Clinical Cognition Sciences, Department of Neurology, [2]Department of Neurology, *Email: jpape-neumann@ukaachen.de, [3]Department of Psychiatry, Psychotherapy and Psychosomatics, [4]Section Neuropsychology, Department of Neurology, Uniklinik RWTH Aachen, Aachen, Germany; [5]Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Jülich, Germany; [6]SRH University of Applied Sciences for Health Gera GmBH, Gera, Germany; [7]JARA – Translational Brain Medicine, Aachen and Jülich, Germany

The present study investigated whether phonological awareness training is an effective intervention to significantly improve reading in German dyslexic third and fourth graders with a phonological awareness deficit, and whether these children can equally benefit from a phonology-based reading training or a visually-based reading training. German speaking dyslexic elementary school children (*n*=30; *M*=9.8 years) were matched by forming triplets based on IQ, reading quotient and phonological awareness and then randomly assigned to one out of three interventions (*n*=10): a phonological awareness training, a phonology-based reading training (phonics instruction), and a visually-based reading training (repeated reading of sight words). A total of 20 training sessions (30 minutes each) were distributed over four weeks. Typical readers (*n*=10; *M*=9.5 years) were assigned to the control group. Phonological awareness training directly improves reading comprehension in German dyslexic children with a phonological awareness deficit. However, these children can equally benefit from a visually-based reading training. In contrast, the phonology-based reading training has a direct selective effect on decoding but not on reading comprehension. Despite divergent short-term patterns, long-term improvement of reading comprehension and decoding is similar across all training groups, irrespective of the training method. Phonological awareness may but does not need to be part of reading remediation in dyslexic children with a phonological deficit when learning to read a consistent orthography. Rather, a visually-based reading strategy might compensate for the phonological deficit in dyslexic children after the initial stage of reading acquisition.

Key words: dyslexia, children, intervention, phonology, reading

INTRODUCTION

Reading is an indispensable skill in modern societies, as it provides access to written knowledge. Impaired reading can result in a wide range of problems and can particularly affect children's school achievement and educational career (Arnold et al. 2005). Thus, effective support at an early stage of reading development is of particular importance. Because not all dyslexic children profit equally from the different remediation techniques (Ramus 2003),

remediation strategies targeting the cause of the reading deficit are needed.

One of the dominant theories in the field of dyslexia research is the phonological deficit hypothesis, claiming that phonological deficits in the representation and processing of speech sounds are the direct cause of reading impairment (Snowling 1995, Shaywitz and Shaywitz 2005). Whereas some researchers restrict the role of phonological deficits to mediating between other cognitive skills (e.g., auditory perception, Tallal 1980) and impaired reading, others propose that attentional deficits, which are dissociated from phonological deficits, are the cause of reading impairment (e.g., Vidyasagar and Pammer 2010). Out of the range of phonological processing deficits the impairment of phonological awareness has been

Correspondence should be addressed to J. Pape-Neumann
Email: jpape-neumann@ukaachen.de

regarded as the major cause of developmental dyslexia (Vellutino et al. 2004). Phonological awareness involves the recognition, discrimination, and manipulation of sounds in spoken language, focusing on different sizes of the sound unit like syllables, onsets and rimes, or phonemes (Anthony and Lonigan 2004). Being able to separate and manipulate phonemes is a critical step for insight into the alphabetic principle, as the comprehension of correspondences between letters and sounds hardly emerges spontaneously in young children (Wimmer et al. 1991). This, in turn, requires phonological awareness (Snowling 1995). Thus, dyslexic children with impaired phonological awareness seem to experience difficulties in abstracting letter-sound correspondences and therefore fail to develop phonological recoding of letter patterns into spoken words (Snowling 1981, Manis et al. 1993). Recoding refers to the ability to apply the knowledge of letter-sound correspondences to correctly translate a printed word into sound. Thus, the process of recoding requires access to the mental phonological representation but does not necessarily require access to the meaning of the word (Marx 1998).

Most studies addressing the relationship between phonological awareness and reading were conducted in English-speaking countries, whereas the present study was carried out in Germany. Different from native English readers, German readers are confronted with rather high grapheme-phoneme consistency. In contrast to English-speaking countries reading is not taught before primary school in German-speaking countries (Landerl and Wimmer 2008). German children learning to read usually experience a synthetic PHONICS teaching approach, which first induces graphemes and their corresponding sounds and then immediately builds up the blending of these sounds (Wimmer et al. 2000). The combination of a synthetic PHONICS approach and the consistent German orthography seems to facilitate particularly the awareness of phonemes. Although clearly being on disadvantage in terms of letter knowledge and phonological awareness at the phoneme level assessed in kindergarten, German first and second graders develop equal phoneme awareness skills as well as better letter knowledge and decoding[1], when compared to their American peers

(Mann and Wimmer 2002). In spite of equivalent performance on phoneme awareness, phonemic tasks are more strongly related to decoding in the American than in the German children in the first two years of primary school (Mann and Wimmer 2002). Furthermore, it was shown that impaired phoneme awareness at school entry does not affect the decoding ability of German children about three years later (Wimmer et al. 2000). In contrast, Wimmer and coworkers (2000) reported clear negative effects of phonological awareness deficits at the phoneme and onset-rime level on the accurate reading of foreign words of mainly English origin by the end of third grade.

These findings indicate that, compared to English-based findings, the negative effect of a phonological awareness deficit in German children might emerge only later whereas the early stage of reading acquisition is less affected.

Further studies implemented in countries with rather consistent orthographies reported heterogeneous findings concerning the role of phoneme awareness in reading development. Whereas the effect of phoneme awareness on decoding was found to diminish after the first year of reading instruction in children learning to read the consistent Dutch orthography (de Jong and van der Leij 1999), Müller and Brady (2001) found that there was still an effect of phoneme awareness on decoding in fourth graders learning to read the very consistent Finnish orthography. Finally, in a cross-linguistic comparison of English and Dutch readers aged six to twelve, phoneme awareness was a significant predictor of word and nonword reading in both languages (Patel et al. 2004). Thus, the question whether the importance of phonological awareness is limited to the initial phase of reading development in consistent orthographies is controversially discussed.

Language-related differences concerning the importance of phonological awareness for reading also apply to dyslexic readers. Whereas the impact of impaired phoneme awareness on reading was found to persist up to fifth grade in English-speaking dyslexic children (Torgesen et al. 1997), studies implemented in countries with rather consistent orthographies reported heterogeneous findings. On the one hand, diminishing associations between phoneme awareness and reading by the end of second grade at the latest were found for German (Wimmer 1993, 1996, Landerl and Wimmer 2000). In contrast, Wimmer nad colleagues (2000) still

---

[1] Decoding goes beyond recoding as explained above, as decoding refers to the ability to correctly recode a word and to access the meaning of a word (Marx 1998). Thus, the process of decoding requires not only to access a mental phonological representation but also to establish a relationship between a phonological and a semantic representation, too (Adams 1993).

found a trend towards a negative effect of impaired phoneme awareness on the accurate reading of foreign words in German dyslexic third graders. Foreign words can usually not be decoded. Instead, they can only be recognized from memory by sight after building a vocabulary of these words. Thus, the finding of a negative impact of impaired phoneme awareness on the accurate reading of foreign words indicates that phoneme awareness does not only influence recoding or decoding but word recognition (recognizing words by sight), too.

Furthermore, Heim and others (2008) reported impaired phoneme awareness for one subtype of German children with dyslexia at the end of third grade. Similarly, de Jong and van der Leij (2003) found that the impact of phoneme awareness on Dutch dyslexic readers persists throughout the end of fourth grade. In a cross-linguistic study Caravolas and colleagues (2005) compared the importance of phoneme awareness in the reading development of English and Czech dyslexic children (the latter learning an orthography that is more consistent than the German orthography). Interestingly, significant deficits of phoneme awareness were reported for both groups of English and Czech children at least up to the fifth grade (Caravolas et al. 2005).

Similarly, Landerl and coauthors (1997a) found a comparable phoneme awareness deficit in dyslexic English and German children aged eleven to twelve. Nevertheless, the authors reported different reading difficulties arising from the phonological deficit, when they differentiated between reading accuracy and reading fluency[2]. Reading accuracy refers to the ability to read words without errors and reading fluency is the ability to read accurately, fast, and with proper expression (National Reading Panel 2000). Whereas both, reading fluency and reading accuracy, were impaired in English children, especially reading fluency was impaired in German children (Landerl et al. 1997a). Thus, German dyslexic children were found to read in a slow and laborious way. The finding of a fluency deficit in German children is in line with other studies: The typical problem of German-speaking dyslexic children at older age is impaired reading fluency (Landerl 2001), whereas rather high reading accuracy after first or second grade in German-speaking dys-

lexic children was reported (Wimmer 1993, 1996, Landerl and Wimmer 2008).

Rapid automatized naming (RAN), originally assessed by naming letters, digits, pictured objects or colors (Denckla and Rudel 1976), is of particular interest when trying to explain the persisting reading fluency problem of German dyslexics mentioned above. An early RAN deficit has been reported to be predictive for later reading fluency impairments particularly for the German orthography (Wimmer et al. 1998). Furthermore, it was found that RAN predicts more variance in reading fluency than phonological awareness in poor readers from second to fourth grade learning the German orthography (Moll et al. 2009). Finally, the impact on reading of both, phonological awareness and RAN, was found to be stronger in less consistent orthographies, when dyslexic readers between eight and thirteen years were compared (Landerl et al. 2013). Learning to read requires reading with understanding. Thus, reading development does not only involve the acquisition of accurate and fluent word decoding and word recognition but the acquisition of reading comprehension, too. Reading comprehension goes beyond decoding explained above (Marx 1998). Thus, the process of reading comprehension not only requires access to the meaning of single words, but access to the meaning of written language on the sentence and the text level, too (Hoover and Gough 1990). Reading comprehension is a complex process in itself and it depends upon other reading skills as well as upon oral language skills. Again, most studies examining predictors of reading comprehension have recruited children reading a less consistent orthography than German. However, studies including typical readers in rather consistent orthographies reported predictors similar to those found in studies concerned with English-speaking children. Apart from the predictor phonological awareness, which was found in studies concerned with Hebrew and German typically reading children (Schneider and Näslund 1993, Schiff et al. 2011), further predictors like word decoding, vocabulary knowledge and listening comprehension were reported in Dutch children (de Jong and van der Leij 2002). In addition, reading fluency was identified to be a predictor of reading comprehension in Czech children (Caravolas et al. 2005).

Similar to the predictors identified for typical readers, Landerl (2001, 2003) found that reading comprehension was related to phonological awareness, RAN,

---

[2] As usual in the case of reading studies, reading fluency was assessed as reading time (measured in seconds) in this study, whereas reading accuracy was assessed as an additional measure (number of errors).

reading fluency, basal reading skills and auditory short-term memory in German-speaking dyslexic third graders. Finally, Müller and Brady (2001) found further support for the importance of phonological awareness for reading comprehension when Finnish poor readers in grade four were examined. To summarize, several studies concerned with consistent orthographies suggest, that, apart from non-phonological language skills, phonological awareness plays an important role not only in the development of basal reading skills but in the development of reading comprehension throughout the period of primary school as well.

As reported above, associations between phonological awareness and different reading skills were frequently found. As a consequence, trainings of phonological awareness have been considered a suitable method to support learning to read. Several studies across orthographies differing in consistency including German confirmed reliable effects of phonological awareness training not only on phonological awareness skills but on reading skills, too (Bus and van Ijzendoorn 1999, Schneider et al. 1997). The finding that children with initially low levels of phonological awareness in particular benefit from such programs across different orthographies (Ehri et al. 2001b), was confirmed for German orthography, too (Schneider et al. 2000). According to the meta-analysis of the National Reading Panel (NRP) reported by Ehri and colleagues (2001b), phonological awareness instruction has an impact on recoding, decoding and reading comprehension. Results of the meta-analysis by Bus and van Ijzendoorn (1999) particularly emphasize the long-term effect of phonological awareness interventions on reading comprehension.

Several studies confirmed the phonological linkage hypothesis (Hatcher et al. 1994) across different orthographies including German. Thus, normally developing children as well as children-at-risk benefited most from interventions, which combined phonological awareness training with the training of letter-sound correspondences (Bus and van Ijzendoorn 1999, Schneider et al. 2000, Ehri et al. 2001b, Roth and Schneider 2002). Finally, the effects of phonological awareness interventions for preschoolers on reading were statistically larger than for primary scholars according to findings of the NRP (Ehri et al. 2001b).

In summary, there is plenty of evidence that training phonological awareness reliably enhances phonological and reading skills particularly in children with poor phonological awareness skills. Although effects were larger at the preschool level, phonological awareness instruction was also found to significantly enhance reading in primary school children. Intervention programs exclusively targeting the phonological awareness deficit have usually been administered to preschoolers or first graders in Germany. In contrast, phonological awareness interventions have rarely been administered to older German-speaking dyslexic children. Considering that phonological awareness was repeatedly found to play an important role in reading throughout the stage of primary school across orthographies differing in consistency, the question arises if a phonological awareness training improves reading not only in younger but also in older dyslexic primary school children learning a consistent orthography. In order to explore whether training phonological skills is enough to ameliorate reading in German children with dyslexia it seemed reasonable to compare a training of phonological awareness to a visually-based training of reading. Furthermore, following the phonological linkage hypothesis, the additional comparison with an intervention involving both, phonology and reading, was considered necessary.

Consequently, the major goal of the present study was to investigate, whether phonological awareness training is an effective intervention to significantly improve reading in German dyslexic third and fourth graders with a phonological awareness deficit, and whether these children can equally benefit from a phonology-based reading training or a visually-based reading training.

METHODS

The standardized reading screening Salzburger Lese-Screening für die Klassenstufen 1–4 (Salzburg Screening for Reading from Grade 1–4, SLS 1–4, Mayringer and Wimmer 2003) was initially implemented for recruitment followed by comprehensive testing (t1). During the training period beginning in November 2009 and continuing until February 2011 the interventions occurred in a rolling fashion. Each individual training included 20 sessions (30 minutes each) and was usually performed at five days per week spread over four weeks (M=28 days, SD=4.6 days). All children were reassessed right after the completion of their intervention (t2) and three months after for follow-up (t3) to assess training effects and their persis-

tence. All procedures applied in this study were in accordance with the Declaration of Helsinki and approved by the ethics committee of the Medical Faculty of RWTH Aachen University (Reference number EK 153/08).

Participants

The flow of participants through the stages of the study is illustrated in Figure 1. Initially, 88 potentially dyslexic children were enrolled in the present study after having screened 785 third and fourth graders from 27 primary schools in the Aachen region. To be included for further testing, children had to score below average on a standardized reading screening (reading quotient <90 in the SLS 1–4, i.e. one standard deviation below average) and to be monolingual German. In the course of subsequent testing, the potentially dyslexic children were required to show reading deficits in at least one of the subtests of the well established German reading test Knuspels Leseaufgaben (Knuspel`s Reading Tasks, KNUSPEL-L, Marx 1998, percentile <25), at least average non-verbal intelligence in the German version of the Cattell Culture Fair Test 20 (CFT 20, Weiss 1998, IQ ≥85), phonological awareness deficits according to the commonly used German test Basiskompetenzen für Lese-Rechtschreibleistungen (Basic Skills for Reading and Writing, BAKO 1–4, Stock et al. 2003, percentile <25 in at least one of the subtests) and no severe sensory visual or auditory deficits according to the German perceptual tests Vortests für Wahrnehmungsfunktionen (Pre-Tests for Perception Functions, WAFW, Sturm 2009, ≥80% items correct in each subtest). As illustrated in the flow chart (Fig. 1) 21 children were excluded during the assessment at t1. Furthermore, because the present study was part of a larger project, 35 dyslexic children with predominantly attentional deficits, as assessed by the German attention test battery Kinderversion der Testbatterie zur Aufmerksamkeitsprüfung (Test of Attentional Performance for Children, KiTAP, Zimmermann et al. 2002), were included in another part of the project. Finally, 32 dyslexic children with predominantly phonological deficits attended the trainings of the present study.

In order to parallelize the three training groups, triplets of children were formed such that their non-verbal IQ (CFT 20), reading quotient (SLS 1–4), and phonological awareness score (BAKO 1–4) were as similar as possible, and then randomly assigned to, either, a pure phonological awareness training (PHON), a phonology-based (PHONICS) or a pure visually-based reading training (READ). As illustrated in Figure 1, there was a small initial disparity in the number of group members, although training groups were matched by triplets as far as possible. This was due to the fact that 32 registered dyslexic children were randomly assigned to the final versions of trainings, which resulted in one incomplete triplet. Thus, children from the incomplete triplet were also randomly assigned to the three training groups, resulting in ten children being trained in the PHON group, and eleven children being trained in the PHONICS and in the READ group, respectively.

Although having completed the trainings, data sets of two dyslexics and of five children from the control group had to be dropped *post-hoc* (see Fig. 1). In this context, the issue of IQ requires a detailed explanation. Typically reading children were recruited in order to compare training effects between the three trained groups and an untrained control group. The initial search for control children took place in the context of searching for dyslexic children in primary schools. This initial recruitment of children for the control group proved to be ineffective. Dyslexic children and their families were highly motivated to take part in any of the trainings and therefore were ready to undergo the assessment. In contrast, this did not apply to typical readers and their families during the first recruitment. These children did not take part in any training and thus were not willing to spend that much time on assessment. This led to an additional recruitment via the digital network of the local School of Medicine (RWTH Aachen University, Aachen). Parents working in the Medical School usually are interested in participating in scientific studies and are motivated to learn about their children`s abilities. The second recruitment of control children resulted in 15 control children without reading deficits according to SLS 1–4 and KNUSPEL-L and without severe auditory or visual perception deficits according to WAFW. The second search for controls resulted in a bias, because the group of control children showed a significantly higher IQ than the dyslexic children. In order to avoid group differences in IQ, potentially influencing the improvement of reading dur-

ing the training period, the inclusion criterion concerning IQ was modified *post-hoc*[3]. Thus, three typical readers and one dyslexic child (READ) were excluded, because their IQ (>130) suggested they were intellectually gifted. Since a significant group difference between the PHON group and the control group still remained, one dyslexic pupil in the affected training group showing the lowest IQ across all groups (IQ 85) was excluded as well. The remaining typical readers (CON) received no intervention but took part in all assessments (*n*=10).

The final sample consisted of 40 children (19 girls, 21 boys; 12 third graders, 28 fourth graders). Their average age was 9.7 years (range 8.7–10.7 years; SD 0.5 years). Each of the four groups consisted of ten children.

Nonparametric pairwise group comparisons were conducted to assess comparability of the training groups at t1 with respect to criteria used to select dyslexic children (reading quotient, phoneme awareness, and IQ). An adjusted alpha level of $P<0.017$ (i.e. $P<0.05/3$ for two-sided hypotheses) was employed for each comparison, because the exact Mann-Whitney U tests were carried out on three pairs of groups. Values of $0.017<P<0.034$ were interpreted as trends. Furthermore, the same analyses were conducted to compare the training groups with respect to general inclusion criteria (age and sensory skills).

No significant differences between training groups were found for reading quotient (smallest $P=0.436$), phoneme awareness (smallest $P=0.089$) and IQ (smallest $P=0.529$). Furthermore, no significant differences between groups were found for the general inclusion criteria of age (smallest $P=0.393$) or sensory skills (smallest $P=0.165$). Finally, no significant differences between groups were found according to Fisher`s Exact Probability Test (smallest $P=0.474$, two-tailed probabilities) concerning handedness. The vast majority of recruited children were right handed according to case history. Table I reports descriptive statistics for the selection measures.

---

[3] The authors of the present study are aware that evidence for dyslexia based on operationally defined exclusion criteria has been controversially discussed in recent years (e.g., Gresham and Vellutino 2010). However, given that particularly children in the control group had an IQ above average, there was serious concern that the difference in IQ could enhance the contrast in reading between dyslexic and normal reading children. Thus, these children were considered to be outliers based on their performance rather than typical readers.

Blinding

Neither the participants and their families nor the instructors were blinded regarding group allocation in the present study. Unlike in drug trials, in cognitive treatment trials it is not possible to guarantee double blinding, because participants know they are trained and instructors have to know the training method they are applying. In the case of the present study, ethical standards in the School of Medicine, RWTH Aachen, require to provide general information on all training methods included in a study before written consent of participants can be obtained. Thus, the type of training they were receiving could not be concealed from the children and their families, once the training had begun. However, the speech- and language pathologists who tested the child before treatment and documented children's behavior during the assessments were different from those speech- and language pathologists, who did the trainings afterwards and thus were not aware of the child's group allocation. Moreover, further speech-and language pathologists, who were not involved in testing or training, analyzed the documented behavior in terms of raw and standardized scores according to the respective test manuals` guidelines. These speech-and language pathologists were blinded regarding group allocation, too.

Furthermore, random assignment to the specific treatment was monitored by the first author, thus the instructors of trainings were not involved in the referral process. In addition, although children and their parents as well as the instructors of trainings were not blind to group allocation as soon as trainings had begun, a bias regarding effectivity of different trainings methods is unlikely, as the information on all training methods that was given to children, parents and instructors did not reflect any difference regarding the extent of their potential impact. Finally, the fact that the group allocation was known by the participants should not be relevant because all dyslexic children received treatment and none of the dyslexic groups received a placebo condition.

Assessment battery

All children were administered a large battery of tests measuring non-verbal intelligence, reading skills, phonological awareness, RAN, auditory memory span and perception. Testing was carried out individually in

a fixed order over the course of two or three sessions. At the subsequent times of testing (t2, t3), the battery of tests was repeated, excluding measures of non-verbal intelligence and perception. A detailed description of each test is given in the following section.

Basic reading ability

Reading ability was initially assessed by means of a group test with the standardized screening Salzburger Lese-Screening für die Klassenstufen 1–4 (Salzburg Screening for Reading in Grade 1–4, SLS 1–4, Mayringer and Wimmer 2003) in order to recruit volunteers for the study. The SLS 1–4 measures reading fluency and basic reading ability in simple and short sentences, posing low comprehension demands. Children were asked to silently read short sentences as quickly as possible within a time limit of three minutes and judge their semantic correctness. The SLS 1–4 is scaled like the IQ and provides a Reading Quotient which expresses each child`s individual deviation from the average of the standardization sample. Since the SLS 1–4 was implemented as a recruitment instrument in the first instance, it was not repeated at subsequent times of testing.
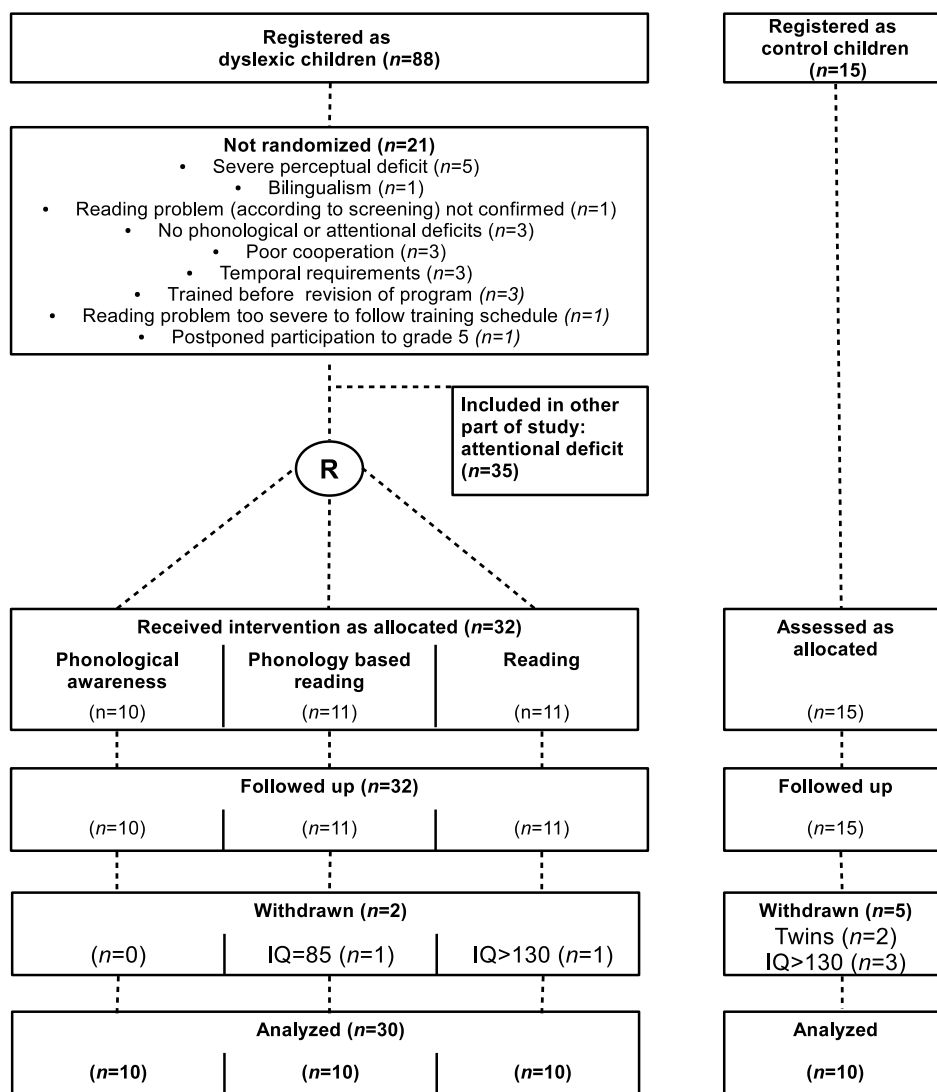
Fig. 1. Flowchart indicating the progress through the stages of the study including withdrawals. (R) randomization. A pair of typically reading twins was excluded from the final analysis, because only after having finished the assessments the parents indicated that their children were monozygotic twins. [Modified according to the flowchart used in the training study of Doesborgh et al. (2004)].

Table I

Descriptive data on measures of selection at t1

Variable	PHON[1]		PHONICS[1]		READ[1]		CON[1]	
Sex (female/male)	5/5		6/4		4/6		4/6	
Handedness[2] (L/R/both)	1/8/1		0/9/1		0/10/0		1/9/0	
	M	SD	M	SD	M	SD	M	SD
Age[3]	9.8	0.4	9.9	0.5	9.7	0.6	9.5	0.4
Non-verbal IQ[4]	107	10	105	11	107	9	117	7
PA[5]	43.8	3.0	41.2	5.5	40.2	5.2	51.7	8.7
Reading quotient[6]	80	7	78	10	76	10	118	13
Perception of brightness[7]	95	4.6	98	1.9	97	2.5	97	2.8
Perception of shape[7]	94	2.2	92	5.8	93	4.7	94	4.2
Perception of loudness[7]	97	2.6	99	1.7	97	3.1	99	1.7

(PHON) Phonological awareness training; (PHONICS) phonology-based reading training; (READ) visually-based reading training; (CON) control group. (1) *n*=10; (2) handedness referring to the number of children that are left/right handed (L/R) or do not prefer any hand (both) as specified in case history; (3) age in years; (4) IQ (Mean 100, SD 15) referring to age norms (5) total scores (*t*-values) of PA on phoneme level of subtests from BAKO 1–4; (6) reading quotient (Mean 100, SD 15) from Reading Screening SLS1–4; (7) percentage of correct responses of subtests from WAFW.

Recoding, decoding and reading comprehension

Recoding and decoding skills at the word level as well as reading comprehension at the sentence level were measured by using the speeded reading test Knuspels Leseaufgaben (Knuspel`s Reading Tasks, KNUSPEL-L, Marx 1998). In the subtest Recoding children had to decide if the pronunciation of orthographically well written, meaningful word pairs differed (e.g., Do these words sound the same: Weizen – Weisen? No). Ten pairs each of visually similar and visually different homophones and visually similar and visually different heteronyms were presented. Children had to apply the knowledge of letter-sound correspondences and to correctly translate a printed word into sound without necessarily assessing the meaning of the word in order to complete the recoding task successfully.

In the subtest Decoding children had to decide if the pronunciation of a pseudo-word sounded like an existing German word [e.g. Does this word sound like a real existing German word: Rogg? Yes (sounds like the word Rock)]. Twenty pseudo-words, which could be pronounced in accord with German orthography rules, sounded like existing German words, whereas the phonological information of another 20 pseudo-words was not associated with meaningful German words. In this task children had to correctly recode the pseudoword first to activate the correct phonological representation. Based on the activated phonological representation they had to decide if the sound of the recoded word matches with a semantic representation. Thus, the task of decoding goes beyond the task of recoding, because access to the meaning of a word is required.

In the subtest Reading Comprehension 14 items each consisting of one question and one request presented in two to three sentences targeted knowledge about the own person, the current test situation or the legendary creature (Knuspel) introduced by the test

(e.g., Which two letters are written at the end of the word Knuspel? Write both letters in the first circle.) In this task children had to precisely understand the meaning of the presented sentences in order to correctly answer the question by executing the appropriate statement.

All subtests had time restrictions. KNUSPEL-L provides norms (*t* values and percentiles) for the middle and the end of first to fourth grade for each subtest. Because children were assessed at different points of the school year, additional norms for the beginning of third and fourth grade were determined for the study. KNUSPEL-L subtests served as an outcome measure and were administered at all three times of testing.

Reading fluency and accuracy

Reading fluency based on text reading was assessed by the subtest Text Reading of the Salzburger Lese- und Rechtschreibtest (Salzburg Reading- and Spelling Test, SLRT Leseteil, Landerl et al. 1997b). In this subtest children were asked to read a text consisting of six or seven simple sentences aloud as fast as possible without making mistakes. The reading of a shorter practice text, for which corrective feedback was given, preceded the assessment. Each child`s reading output was recorded for registration of reading errors and reading time. Unfortunately, the percentiles from first to fourth grade for the time needed to read the text are partly given only as a percentile range and no percentiles for reading errors are available in the test manual. Therefore, the raw data of reading time measured in seconds and of the number of reading errors entered further analyses as the measures of reading fluency and accuracy. Text reading was assessed at all three times of testing.

Non-verbal intelligence

Non-verbal IQ was assessed with the German version of the Cattell Culture Fair Test 20 (CFT 20, Weiss 1998). The four subtests of the short form (Part 1) required children to identify one of several pictures in order to continue a series of pictures and to complete a set of pictures. In addition, children had to classify a picture that had or had not been constructed according to the same idea as the others. Age-related scores were used to rule out initial group differences regarding IQ, which might have influenced the way reading abilities developed in the course of this study.

Phonological awareness

Phonological awareness was measured by the standardized German test Basiskompetenzen für Lese-Rechtschreibleistungen (Basic Skills for Reading and Spelling, BAKO 1–4; Stock et al. 2003). BAKO consists of seven subtests predominantly assessing phonological awareness at the phoneme level. Subtest Segmentation of Pseudo-Words required to segment 8 pseudo-words into their phonemes (e.g., frap: /f/ /r/ /a/ /p/). In the subtest Vowel Substitution the vowel /a/ had to be substituted by the vowel /i/ continuously (e.g., S*a*nd: S*i*nd) in 8 words (two to four syllables) and 4 pseudo-words (two to four syllables). The task of Phoneme Deletion called for pronunciation of 3 words (one beginning with a consonant cluster; one and two syllables) and 4 pseudo-words (one beginning with a consonant cluster; two and three syllables) with the first sound removed (e.g., *F*loß: loß). In the task of Phoneme Commutation children had to pronounce the pseudo-word (two syllables), which resulted from switching the first two phonemes of six words (e.g., *Ar*m: *Ra*m) and five pseudo-words. In the subtest Phoneme Categorization children were asked to identify the one out of four items (one syllable), whose initial or final sound differed from that of the others (e.g., *K*opf – Turm – tief – Trick; three groups of words and five groups of pseudo-words). The task of Vowel Length classification required to identify the one out of four pseudo-words (one syllable, ten items) in which the vowel was of different length (e.g., maar – raas – d*a*ck – laat). In the last subtest Phoneme Reversal children were asked to pronounce 8 words (one or two syllables) and 10 pseudo-words (one or two syllables) in reversed order of sounds (e.g., ruf: fur). The spoken stimuli of each subtest were played from a CD. For each subtest there were two or three practice items, for which corrective feedback was given. Separate norms (*t* values and percentiles) for grades one to four were available. In the present study the average *t* value of all subtests entered further analyses. This measure served as a secondary outcome aspect and was implemented at all three times of testing.

In order to complement the BAKO with tasks of phonological awareness at the levels of syllable and onset-rime, five out of nine subtests of the experimental German version of the Queensland University Inventory of Literacy (Dodd et al. 1996) were administered (Queensland University Inventory of Literacy-Deutsch,

QUIL-D; Hofmann 2000, unpublished master thesis). The 12-item subtest Syllable Identification required indicating whether two syllables in a pair of words (two syllables each) sounded alike. Provided that two syllables sounded the same, the children had to decide about their initial (four items) or final (four items) position (e.g., Fal*ter* – Git*ter*: yes/ at words` ends). In the 12-item subtest Syllable Segmentation participants had to repeat words while segmenting each syllable (three items each of two to five syllables) and to indicate the number of syllables (e.g., Multiplikation: Mul-ti-pli-ka-tion/ five). In the 12-items subtest Rime Identification children were asked to decide whether two words (one syllable, 6 rhyming pairs) rimed (e.g., Boot – Beet: no). In the 20-items subtest Spoonerism children had to pronounce a pair of words (one syllable, 4 pairs containing consonant clusters) after their onsets had been switched (e.g., *f*ein – *D*ach: dein – Fach). The 12-item subtest Sub-Syllabic Synthesis required to blend onsets and rimes (4 items) or phoneme sequences (three or four phonemes, 4 items each; e.g., /p/ - /l/ - /a/ -/n/: Plan). For each subtest there were two to four practice items, to which corrective feedback was given. The spoken stimuli of each of the five subtests were played from a CD, which had been recorded in order to provide standardized stimulus presentation in the context of the present study. As the QUIL-D represents an experimental test-version, no norms were available. Thus, the QUIL-D was not implemented as a recruitment instrument but was administered at all three times of testing to analyze the additional effects of trainings on phonological awareness that were not taken into consideration by BAKO 1–4. The averaged percentage of correct answers across subtests entered further analyses.

Rapid automatized naming

An adapted version of the German matrices for testing Rapid Automatized Naming by Mayer (2008) was constructed for the study (see van Ermingen-Marbach et al. 2014, for details). All stimuli of the adapted version consisted of one syllable. The matrices were presented on four DIN A4 sheets each with 10 lines of five Stimuli. Each test matrix was preceded by two lines of practice items, for which corrective feedback was provided. Children were asked to name the stimuli as fast as possible. The matrix of letters was followed by matrices of digits, colors and objects. The average number of items named per second across matrices entered the

subsequent analyses as measures of rapid automatized naming fluency. Omitted items were subtracted from the total number of items in each matrix (n=50). This measure was used at all three times of testing.

Auditory short-term memory

The Digit-Span task of the German test Psycholinguistischer Entwicklungstest (Test for Psycho-Linguistic Development, PET: Zahlenfolgen-Gedächtnis (ZGF, Angermayer 1974) was conducted in order to measure the storage capacity of short-term memory. Series of digits were auditorily presented at a rate of one digit per second, and children had to immediately repeat them in the given order (28 sequences of digits increasing in length and consisting of two to eight items). To determine the starting point of testing, two series each of the same lengths were provided initially with increasing length, until the child did not repeat correctly in the first trial. Sequences containing one item less were provided subsequently and the test was continued with sequences of increasing length, until two consecutive errors occured (two trials per sequence were accepted). The t-value of this subtest was included in further analyses as one measure of auditory memory span.

To assess phonological short-term memory, the subtest Mottier of the Zürcher Lesetest (Zürich Reading Test, Linder and Grissemann 2000) was conducted. This 30-item task required to repeat meaningless consonant-vowel-combinations increasing in length (two to six syllables). The Mottier was developed in 1951 and norms changed over time with decreasing performance levels. Given that existing norms from diverse studies are hardly comparable, the raw total of items repeated correctly was employed for further analyses. Stimuli from both memory tasks were played from a CD, which had been recorded by the Media Center of the Uniklinik RWTH Aachen to accomplish standardized stimulus presentation in the present study. Identical stress and precise articulation of syllables were cared for at the recording. Both memory tasks were administered at all three times of testing.

Sensory skills

In order to exclude the possibility that basic perceptual impairments may cause poor performance during testing or a poor response to subsequent intervention,

parts of the German screening Vortests für Wahrnehmungsfunktionen (Pre-Tests for Perception Functions, WAFW, Sturm 2009) were administered. Children had to perform three computer-controlled discrimination tasks along the dimensions of brightness, shape, and loudness without time pressure. The first subtest required to press a button when a square or a circle became darker or brighter. In the second subtest children were asked to decide whether two figures presented simultaneously were identical or different and to press corresponding buttons. The last subtest required to indicate changes in loudness. Each subtest was preceded by standardized instructions with practice items. When the child`s behavior indicated that the instructions had not been understood, instruction and practice phase were repeated. The percentage of correct responses for each subtest was used for further analyses. To rule out relevant sensory deficits, children scoring lower than 80% in any of the subtests were excluded from the study. This screening test was assessed at t1 only.

Training procedures

In order to control for confounding effects of other trainings or therapies, the children were not enrolled in other remediation programs during the training interval. The trainings were free of charge. All training sessions took place in silent rooms at children's homes or schools in children's spare time. Two out of seven speech-language pathologists were assigned to one child and alternately administered the individual training. All instructors attended uniform tutorials prior to the training period making them well familiar with the detailed written instructions and standardized training schedules.

In order to ensure treatment fidelity, coaching was provided by the first author during periods of intervention. Additionally, instructors completed standardized daily records for each child. Thus, instructors had to precisely indicate the date and number of each session as well as each training activity undertaken in the course of each session. Moreover, they had to provide detailed information on the success of each session according to standardized criteria (i.e. reading speed or percent of items read correctly) and were asked to add comments on individual strengths or weaknesses that had been noticed in the course of a session. In addition, instructors entered general treatment infor-

mation (i.e., codes for instructors and the trained child, precise time of training) in an online calendar.

Regular controls of the intervention protocols and the online calendar carried out by the first author confirmed that all children received exactly 20 sessions of training. Thus, there was no difference between groups regarding the amount of training. Furthermore, there was only minor variation across all groups concerning the time period over which sessions were spread (M=28 days, SD=4.6 days). Thus, given children's busy schedules and illnesses, it is remarkable that the request to do five sessions for four weeks was met very closely. Adherence to the standardized training schedules was confirmed for every child according to information received during coaching and according to the intervention protocols. Furthermore, it was apparent from the protocols that the progress through training activities was carried out according to the criteria described below (see details for each training separately).

As described in the following section, parts of each training were computerized. All computerized tasks were implemented on lap-tops. The child sat in front of the computer and the instructor sat next to the child. Appearance of items was controlled by the instructor. Thus, the intervals between stimuli were individually adapted according to the child`s needs. As computerized tasks were rather simply designed, reading times and reading errors were not recorded by the computer but monitored by the instructor.

Phonological awareness training

The purely phonological awareness training (PHON) was conducted with the multimedia version of the German Würzburger Trainingsprogramm (Würzburg program for the training of phonological awareness, WT, Küspert et al. 2001). The WT was originally designed to enhance phonological awareness in kindergarten children at risk for dyslexia as a prevention tool, but it was shown to be applicable in primary school remedial reading classes, too. Because the WT contains rather simple tasks, which partly include semantic or visual cues, the WT was complemented by a paper version of the Hörtraining zur Entwicklung der phonologischen Bewusstheit (Hearing training for the development of phonological awareness, HT, Hollbach 1999). Contrary to the WT, the HT focuses explicitly on third- and fourth graders. Thus, tasks from the HT

provided a higher level of stimulus difficulty (i.e. words presented contained more syllables or more consonant clusters) and were presented in a direct training modus without any cues. Within each session tasks from both programs were alternated. Whereas tasks from the WT were done on a computer with legendary creatures orally presenting the tasks, tasks from the HT were orally presented by the instructor according to the program`s worksheets in the order specified by the training schedule. In both programs, tasks that could not be implemented without reference to written language were excluded. The tasks were performed orally. The combination of two programs (WT and HT) was supposed to meet children's phonological training needs as well as to elicit high motivation.

Tasks from both programs were graded for levels of difficulty according to size of sound units and phonological operation required. The training included 11 levels, starting with the identification of rimes. The next levels covered the identification of phonemes, the segmentation of words into syllables, onset-rimes and phonemes, the blending of syllables, onset-rimes and phonemes into words as well as the manipulation of syllables, onset-rimes and phonemes. All stimuli consisted of real words. If 90% of items per level had been completed successfully in two consecutive sessions, tasks from the next higher level were administered.

Visually-based reading training

The visually-based reading training (READ) was administered using the repeated reading program Blitzschnelle Worterkennung (Word recognition at lightning speed, BliWo, Mayer 2009). BliWo has been designed to teach dyslexic children by addressing the automatized process of reading. The program is based on 30 orthographic patterns (e.g., all, ast, ang), which are composed of letter sequences that frequently occur within German words. The program provides ten training units sequenced in a non-hierarchical order. Each unit includes three orthographic patterns and 21 training words containing one of these patterns. The words are similar in linguistic complexity across training units concerning the use of consonant di- and trigraphs, long and short vowels, and use of consonant clusters. Moreover, each of the units predominantly consists of words with one or two syllables, whereas only few words with three syllables are included in the whole program. Nouns and verbs are most commonly

used in each unit, whereas only few adjectives or pronouns are provided. All tasks of the original paper version were administered in a game-like manner. The visually-based reading training was complemented by the computerized version of fast sight word reading included in the program. This was designed as a simple power point presentation with trained items separately appearing at individual positions on the screen (duration of stimuli presentation 0.5 seconds).

Children worked three to four sessions on the orthographic patterns and words of each training unit, respectively. In the first five minutes of the first session, three orthographic patterns were repeatedly presented on cards one by one and pronounced by the instructor before being pronounced by the children. After becoming familiar with orthographic patterns, subsequent games required fast and accurate repeated reading of these patterns in the next 15 to 20 minutes of the first session. Finally, the trained orthographic patterns were repeatedly read in mixed order in the computerized version of fast sight word reading included in the program during the final five to ten minutes. If ≥90% of trained orthographic patterns had been read accurately in the computerized version of fast sight word reading, repeated reading of words was trained in the next session. In case the child failed to read ≥90% of trained patterns accurately, repeated reading of orthographic patterns was trained in the next session again until ≥90% of trained patterns had been read accurately.

In the first ten minutes of the second session the instructor presented and pronounced 21 training words (containing one of the trained patterns each), before the children pronounced these training words. After becoming familiar with the items, in subsequent games of the remaining second session children were explicitly instructed to focus on orthographic similarities between words. Thus, words sharing the same orthographic pattern were repeatedly read in a blocked condition. Alternatively, words were presented in a scrambled condition and children were asked to group words according to the shared orthographic pattern before repeatedly reading them. Furthermore, the targeted orthographic patterns were colored or had to be colored by the children to enhance awareness for orthographic consistencies across words.

In the third session, different games required to read the trained words without highlighting the orthographic patterns in the scrambled condition. Together the

tasks of the second and third session were designed in such a way that each training word had to be read eight to ten times. Finally, the trained words were read in the computerized version of fast sight word reading in the last five to ten minutes of the third session (each training word presented one to two times). If ≥90% of trained words had been read accurately, the next training unit (containing three orthographic patterns and 21 training words) was administered according to the procedure described above. In case the child failed to read ≥90% of trained patterns accurately, repeated reading of words was trained in the next session until ≥90% of trained patterns had been read accurately in the computerized version of fast sight word reading.

After having worked on 2 training units (covering 6 orthographic patterns and 42 words containing these patterns), trained contents were rehearsed in the following session. Again, if ≥90% of trained words per training unit had been read accurately in the computerized version of fast sight word reading at the end of this session, the next training unit was administered. In case the child failed to read ≥90% of trained words accurately, trained contents were rehearsed in the following session once more.

Phonology-based reading training

The phonology-based reading intervention (PHONICS) was administered by using the Kieler Leseaufbau (Kiel program for building up reading skills, KLA, Dummer-Smoch and Hackethal 2007a), which was designed for the instruction of normally reading and dyslexic children. According to descriptions of Ehri and colleagues (2001a), the KLA follows the idea of PHONICS instruction, as it explicitly teaches grapheme-phoneme correspondences sequentially, before teaching to blend phonemes into syllables and syllables into words.

The game-like paper version of the program progresses systematically in 12 steps differing in linguistic complexity. Thus, only consonants that can be prolonged (/m/, /r/, /s/, /n/, /f/, /l/, /w/, /z/) are included from level one to four in order to enhance the technique of blending. Accordingly, plosive sounds are introduced only after level four. For the same reason, short vowels in the first syllable of a word are not provided until the last level of training. Moreover, words with simple word structure (i.e., vcv, cvcv) are provided from level one to ten, whereas words containing

consonant clusters (i.e., ccvcv, ccvcvc, cvccv, cvccvc) are provided at levels 11 and 12 only. Items requiring the application of orthographic rules for correct reading are not included in the program. Forty regular words are trained at each level, respectively (except for level one with 16 words and level 10 with 24 words). Nouns and verbs are most commonly used in this training, whereas only few adjectives and pronouns are provided. Words included in the program predominantly consist of two syllables, whereas only few words consist of one, three or four syllables.

Because in the present study parts of PHON and READ were computerized, the KLA was complemented by computerized tasks requiring visual identification of syllables Der Neue Karolus (The New Carolus, Dummer-Smoch and Hackethal 2007b). These tasks are constructed according to the steps and the items of KLA and designed as a simple memory game (duration of stimulus presentation was triggered by the child).

In the first five to ten minutes of each session children were asked to blend phonemes into syllables by accurately reading a syllable table up to four times. Furthermore, reading fluency was addressed by reading each syllable table as fast as possible. Two to four consecutive tasks during the next 15 to 20 minutes of each session required blending of trained syllables into real words as well as segmenting real words into trained syllables. Each session was completed by a computerized task requiring visual identification of identical syllables or of syllables which add up to a word during the final five to ten minutes.

In order to control for children's progress, accurate and fast reading of syllables included in one training level was tested on the basis of additional syllable tables (Variabolus, Clarkson-Grabs 2006). According to criteria specified in Variabolus, the next level of training was administered if ≥95% of tested syllables ($k=90$) had been read accurately in 70 seconds (level one to ten) or 60 seconds (level 11 and 12).

Statistical analysis

The assumptions of homogeneity of variance and of normal distributions were violated for some variables. Therefore, nonparametric tests were applied consistently. The primary outcome variables in the current study were changes, i.e. difference scores between two assessment occasions, in the KNUSPEL-L subtests

Recoding, Decoding and Reading Comprehension. The secondary outcome variables were changes in phonological awareness performance (BAKO 1–4; QUIL-D). Since we were primarily interested in detailed comparisons of training methods among training groups and the control group, we specifically carried out nonparametric pairwise group comparisons for both performance differences (t2–t1; t3–t1) separately. The exact Mann-Whitney U test was carried out on six pairs of groups. Then, the Bonferroni-Holm correction for multiple comparisons was applied in order to control for a family-wise error rate of 5% for each task (Holm 1979).

Differential improvement between pairs of groups identified by pairwise comparisons cannot tell whether these effects also represent substantial improvements for each single group. Therefore, Wilcoxon signed ranks tests were used to assess the significance of training effects within each single group concerning those variables, for which the pairwise group comparisons had revealed significant differences. An adjusted alpha level of $P<0.0125$ (i.e. $P<0.05/4$ for one-tailed hypotheses) was employed for each comparison, because the Wilcoxon signed ranks test was carried out on each of four groups. Values of $0.0125<P<0.025$ were interpreted as trends.

RESULTS

The present study investigated differential effects of three trainings on reading performance of dyslexic primary school children with weak phonological awareness. Thus, data on improvement of different reading skills are presented first, followed by data on improvement of phonological awareness. Descriptive statistics for each of the reading outcome variables (t1, t2, t3) are reported in Table II, whereas descriptive statistics for phonological awareness are reported in Table III.

Primary outcome measure: Improvement of reading

Direct improvement of reading

The exact version of the Mann-Whitney U test using the Bonferroni-Holm correction for multiple comparisons revealed no significant differences in mean improvement of recoding (t2–t1) for any pair of groups.

Thus, mean improvement of recoding (t2–t1) in PHON (M_{diff}=3.00, SD_{diff}=7.73) did not differ significantly from mean improvement in PHONICS (M_{diff}=0.90, SD_{diff}=4.65), READ (M_{diff}=4.00, SD_{diff}=5.46) and CON (M_{diff}=0.45, SD_{diff}=3.99), (P=0.436, P=0.971 and P=0.315, respectively). Furthermore, mean improvement of recoding (t2–t1) in PHONICS and READ (P=0.353), PHONICS and CON (P=0.631) as well as in READ and CON (P=0.089) did not differ significantly either.

The exact version of the Mann-Whitney U test revealed that mean improvement of decoding (t2–t1) in PHONICS (M_{diff}=10.45, SD_{diff}=6.46) and CON (M_{diff}=0.75, SD_{diff}=4.49) differed significantly (P-adjusted=0.012). Furthermore, mean improvement of decoding (t2-t1) in PHONICS and PHON (M_{diff}=2.15, SD_{diff}=7.19) differed significantly (P-adjusted=0.045). In contrast, mean improvements (t2–t1) of decoding found for PHONICS and READ (M_{diff}=2.85, SD_{diff}=9.37, P=0.052) as well as found for READ and CON (P=0.579) were not significantly different. Finally, mean improvements (t2–t1) of decoding found for PHON and CON (P=0.796) as well as found for PHON and READ (P=1.000) were not significantly different either. These results suggest that the short-term effect (t2–t1) found for PHONICS is more substantial than the effects found for CON and for PHON with respect to decoding.

The exact version of the Mann-Whitney U test revealed that mean improvement of reading comprehension (t2–t1) in READ (M_{diff}=10.80, SD_{diff}=12.51) and PHONICS (M_{diff}=−1.05, SD_{diff}=7.72) differed significantly (P-adjusted=0.042). In addition, mean improvement of reading comprehension (t2–t1) in READ and CON (M_{diff}=0.45, SD_{diff}=3.77) differed significantly (P-adjusted=0.035). Furthermore, mean improvement of reading comprehension (t2–t1) in PHON (M_{diff}=10.95, SD_{diff}=10.50) and PHONICS differed significantly (P-adjusted=0.036). Finally, mean improvement of reading comprehension (t2–t1) in PHON and CON differed significantly (P-adjusted=0.027). In contrast, mean improvement (t2–t1) of reading comprehension found for PHON and READ were not significantly different (P=0.912). Likewise, mean improvement (t2–t1) of reading comprehension found for PHONICS and CON were not significantly different either (P=1.000). These results suggest that the short-term effects (t2–t1) found for PHON and READ each are more substantial than the effects found for PHONICS and for CON with respect to reading comprehension.

Table II

Performance in reading skills over time of the training study (t1, t2, t3)

Variable		PHON M	SD	PHONICS M	SD	READ M	SD	CON M	SD
Reading ability[1]	t1	37.2	11.0	38.1	7.4	36.0	11.0	60.4	6.0
	t2	44.1 a	5.7	42.6 c	11.2	43.5 a	12.7	61.2	6.0
	t3	47.7 a	10.8	42.9 a	8.8	44.2 a	10.1	61.0	5.5
Recoding[2]	t1	42.3	8.9	44.8	7.3	42.1	7.4	60.0	6.2
	t2	45.3	3.0	45.7	8.6	46.1 c	7.9	60.5	5.5
	t3	43.5	9.3	45.4	7.1	45.4	7.8	60.3	6.4
Decoding[2]	t1	42.9	8.7	37.3	5.1	40.8	7.0	58.2	7.4
	t2	45.1	6.2	47.7 a	8.2	43.6	8.6	59.0	8.1
	t3	47.1	11.9	42.0 c b	6.0	44.6	8.1	58.7	8.2
Comprehension[2]	t1	34.3	17.6	39.7	10.0	33.9	17.7	57.3	2.9
	t2	45.3 a	10.5	38.7	14.5	44.7 a	18.5	57.7	3.9
	t3	49.5 a	10.4	45.3	12.8	45.9 a	17.8	57.6	3.5
Fluency[3]	t1	60.2	22.9	60.3	20.1	62.2	27.2	26.1	6.4
	t2	54.5	15.5	54.4	15.0	58.3	33.6	25.8	5.5
	t3	51.3 c	15.7	52.1	19.0	45.3 a	25.0	24.2	3.5
Accuracy[4]	t1	1.9	1.9	2.3	2.0	3.4	2.8	0.3	0.7
	t2	2.0	1.6	1.4	1.2	2.3	1.9	0.5	0.5
	t3	1.6	1.4	1.0	1.2	2.2	1.7	0.0	0.0

(PHON) Phonological awareness training; (PHONICS) phonology-based reading training; (READ) visually-based reading training; (CON) control group. (a, b) Significant differences ($P<0.0125$; one-sided) between test occasions within a single group according to exact Wilcoxon signed ranks tests; (c) trend of difference ($P<0.025$; one-sided) between test occasions within a single group according to exact Wilcoxon signed ranks tests; (a) performance (at t2 resp. t3) differs significantly from performance at t1; (b) performance at t3 differs significantly from performance at t2; (c) results (at t2 resp. t3) show a trend towards better performance compared to t1. (1) Total score (t-values) based on subtests Recoding, Decoding and Reading Comprehension from Knuspel-L; (2) t-values of single subtests from Knuspel-L; (3) reading time (seconds) of subtest Text Reading from SLRT; (4) number of reading errors of subtest Text Reading from SLRT.

Next, paired-sample tests were conducted comparing reading performance at t1 and t2 within each group concerning the variables decoding and reading comprehension, for which the pairwise group comparisons had revealed significant effects. Wilcoxon signed ranks tests comparing decoding performance (t1 *versus* t2) within each group revealed significant improvement for PHONICS ($P=0.002$). In contrast, no significant short-term improvements were revealed for PHON ($P=0.244$), READ ($P=0.221$) and CON ($P=0.313$). These results demonstrate that the short-term effect in decoding found for PHONICS when compared to PHON and CON by the between-group comparisons represents substantial improvement within this group.

Table III

Performance in phonological awareness, auditory short term memory and ran over time of the training study (t1, t2, t3)

Variable		PHON		PHONICS		READ		CON	
		M	SD	M	SD	M	SD	M	SD
PA (BAKO)[1]	t1	43.8	3.0	41.2	5.5	40.2	5.2	51.7	8.7
	t2	47.3	5.6	48.1 [a]	5.8	40.7	5.5	55.4 [c]	4.8
	t3	49.5 [a]	5.9	48.1 [a]	4.8	44.7 [a, b]	5.1	56.5 [c]	3.1
PA (QUIL-D)[2]	t1	0.80	0.05	0.77	0.10	0.69	0.15	0.89	0.06
	t2	0.89 [a]	0.04	0.77	0.11	0.68	0.22	0.90	0.06
	t3	0.88 [a]	0.05	0.81	0.13	0.79 [a]	0.14	0.95 [a, b]	0.03
Memory digits[3]	t1	47.0	2.3	42.4	13.2	40.4	7.9	51.8	7.2
	t2	44.1	8.1	39.8	13.8	40.1	6.1	49.8	11.1
	t3	48.2	4.7	40.6	13.6	41.9	7.6	51.8	6.5
Memory syllables[4]	t1	20.3	3.0	15.9	3.1	16.5	5.6	21.9	4.4
	t2	20.5	4.5	18.2	4.0	17.5	4.2	21.8	4.1
	t3	21.6	3.9	17.5	3.3	19.4	4.0	23.5	2.8
Rapid naming[5]	t1	1.16	0.17	1.20	0.25	1.18	0.17	1.41	0.17
	t2	1.24	0.16	1.18	0.20	1.16	0.21	1.47	0.23
	t3	1.28	0.17	1.22	0.21	1.23 [b]	0.23	1.48	0.27

(PHON) Phonological awareness training; (PHONICS) phonology-based reading training; (READ) visually-based reading training; (CON) control group. (a, b) Significant differences ($P<0.0125$; one-sided) between test occasions within a single group according to exact Wilcoxon signed ranks tests; (c) trend of difference ($P<0.025$; one-sided) between test occasions within a single group according to exact Wilcoxon signed ranks tests. (a) Performance (at t2 resp. t3) differs significantly from performance at t1; (b) performance at t3 differs significantly from performance at t2; (c) results (at t2 resp. t3) show a trend towards better performance compared to t1. (1) Total Score (*t*-value) of phonological awareness (PA) on phoneme level based on all subtests from BAKO 1–4; (2) averaged percentage of correct answers across subtests Syllable Identification, Syllable Segmentation, Rhyme Identification, Spoonerism and Sub-Syllabic Synthesis from QUIL-D; (3) *t*-value of subtest Digit-Span from PET; (4) raw value of correct answers from Mottier; (5) rapid automatized naming (RAN) expressed by the number of items named per second averaged across matrices of letters, digits, colors and objects.

Wilcoxon signed ranks tests comparing reading comprehension performance (t1 *versus* t2) within groups revealed significant improvements for PHON ($P=0.007$) and for READ ($P=0.010$). In contrast, no significant improvements were revealed within PHONICS ($P=0.500$) and CON ($P=0.385$).

To summarize, these results demonstrate that the short-term effect in decoding found for PHONICS when compared to CON and to PHON represents substantial improvement within this group.

Furthermore, the short-term effects in reading comprehension identified for PHON and READ by the between-group comparisons represent substantial improvements within each of the single groups concerned. For a survey of significant improvements (t2–t1, t3–t1) within each group and for each reading variable cf. Table II.

Finally, a series of between-group comparisons on performance differences (t3–t1) and a series of within-group comparisons (t1 *versus* t3) were conducted in

order to explore long-term improvements of reading. Results of these analyses are reported in the following section.

Long-term improvement of reading

The exact version of the Mann-Whitney U test using the Bonferroni-Holm correction for multiple comparisons revealed no significant differences in mean improvement of recoding (t3–t1) for any pair of groups. Thus, mean improvement of recoding (t3–t1) in PHON (M_{diff}=1.20, SD_{diff}=7.78) did not differ significantly from mean improvement in PHONICS (M_{diff}=0.55, SD_{diff}=3.56, P=0.971), READ (M_{diff}=3.30, SD_{diff}=5.75, P=0.280) and CON (M_{diff}=0.30, SD_{diff}=4.35, P=0.912). Furthermore, mean improvement of recoding (t3–t1) in PHONICS and READ (P=0.280), PHONICS and CON (P=0.631) as well as READ and CON (P=0.190) did not differ significantly either.

Similarly, the exact version of the Mann-Whitney U test revealed no significant differences in mean improvement of decoding (t3–t1) for any pair of groups. Thus, mean improvement of decoding (t3–t1) in PHON (M_{diff}=4.20, SD_{diff}=7.93) did not significantly differ from mean improvement in PHONICS (M_{diff}=4.75, SD_{diff}=5.67, P=0.684), READ (M_{diff}=3.85, SD_{diff}=10.16, P=0.796) and CON (M_{diff}=0.45, SD_{diff}=6.37, P=0.353). Furthermore, mean improvement of decoding (t3–t1) in PHONICS and READ (P=0.971), PHONICS and CON (P=0.143) as well as READ and CON (P=0.165) did not differ significantly either.

The exact version of the Mann-Whitney U test revealed that mean improvement (t3– t1) of reading comprehension in PHON (M_{diff}=15.15, SD_{diff}=12.49) differed significantly (P-adjusted <0.001) from mean improvement in CON (M_{diff}=0.30, SD_{diff}=4.51). In contrast, mean improvement of reading comprehension (t3–t1) in PHON did not significantly differ from mean improvement in PHONICS (M_{diff}=5.60, SD_{diff}=10.07, P=0.165) and READ (M_{diff}=12.05, SD_{diff}=10.95, P=0.684). Furthermore, mean improvement of reading comprehension (t3-t1) in PHONICS and READ (P=0.218), PHONICS and CON (P=0.190) as well as READ and CON (P=0.023) did not differ significantly either. These results suggest that the long-term effect (t3–t1) found for PHON is more substantial than the effect found for CON but does not significantly differ from the long-term effects found for the other training groups with respect to reading comprehension.

Next, Wilcoxon singned ranks tests were conducted comparing reading comprehension performance at t1 and t3 within each group. Significant improvement of reading comprehension was revealed for PHON (P=0.002) and READ (P=0.002). In contrast, no significant improvement of reading comprehension was found for PHONICS (P=0.064) and CON (P=0.472). These results indicate that the long-term effect (t3–t1) in reading comprehension found for PHON when compared to CON represents substantial improvement within this group. Interestingly, long-term improvement of reading comprehension represents substantial improvement within READ, although long-term improvement in reading comprehension for READ did not differ significantly from long-term improvement for PHONICS and CON according to the pairwise group comparisons. In contrast, long-term improvements in PHONICS and CON do not represent substantial changes within each of these two groups.

To summarize, the between-group and within-group comparisons indicate that improvements of decoding and reading comprehension differed across groups. The direct effect in decoding found for PHONICS as well as the direct effects in reading comprehension found for PHON and READ by the pairwise group comparisons represent substantial improvements in each of the single groups concerned. Finally, the results indicate that, irrespective of which training the children got, at t3 long-term improvement of reading, as compared to t1, is relatively similar across all groups, with some deviating patterns in decoding and reading comprehension at t2.

Secondary outcome measure: Improvement of phonological awareness

Further, we carried out nonparametric pairwise group comparisons for both performance differences (t2–t1; t3–t1) on phonological awareness separately to compare the effects of different training conditions on phonological awareness. Again, dependent-samples tests were additionally used to assess the significance of training effects within each single group, taking into account those variables for which the pairwise group comparisons had revealed significant effects. Both kinds of analyses were conducted on phoneme awareness scores (BAKO 1–4) first, whereas the subsequent analyses were concerned with phonological awareness scores on the level of syllable and onset-rime (QUILD-D).

Direct improvement of phonological awareness

The exact version of the Mann-Whitney U test using the Bonferroni-Holm correction for multiple comparisons revealed that mean improvement of phoneme awareness according to BAKO 1–4 (t2–t1) in PHONICS (M_{diff}=6.91, SD_{diff}=6.10) and READ (M_{diff}=0.53, SD_{diff}=2.72) differed significantly (P-adjusted=0.024). In contrast, mean improvement of phoneme awareness (t2–t1) found for PHONICS was not significantly different from mean improvement found for PHON (M_{diff}=3.53, SD_{diff}=5.04, P=0.105) and for CON (M_{diff}=3.73, SD_{diff}=4.91, P=0.143). Furthermore, mean improvement (t2-t1) of phoneme awareness found for PHON was not significantly different from mean improvement found for READ (P=0.165) and CON (P=0.971) either. Finally, mean improvement (t2–t1) of phoneme awareness found for READ and CON was not significantly different (P=0.143).

Next, paired-sample tests were conducted comparing phoneme awareness at t1 and t2 within each group, in order to assess whether differential improvement between pairs of groups identified by pairwise comparisons also comprises substantial improvement for each single group. Wilcoxon signed ranks tests revealed significant improvement for PHONICS (P=0.010) and a trend of difference for CON (P=0.013). In contrast, no significant improvement of phoneme awareness was found for PHON (P=0.030) and READ (P=0.322).

The exact version of the Mann-Whitney U test revealed no significant differences in mean improvement of phonological awareness on the levels of syllable and onset-rime according to QUIL-D (t2–t1) for any pair of groups. Thus, mean improvement of phonological awareness (t2–t1) in PHON (M_{diff}=0.09, SD_{diff}=0.05) did not differ significantly from mean improvement in PHONICS (M_{diff}=0.00, SD_{diff}=0.16, P=0.436), READ (M_{diff}=-0.003, SD_{diff}=0.15, P=0.165) and CON (M_{diff}=0.01, SD_{diff}=0.07, P=0.011). Furthermore, mean improvement of phonological awareness (t2–t1) in PHONICS and READ (P=0.971), PHONICS and CON (P=0.739) as well as READ and CON (P=0.912) did not differ significantly either.

Together, these results suggest that the direct effect (t2–t1) found for PHONICS is more substantial than the effect found for READ with respect to phoneme awareness and that this direct effect represents substantial improvement within PHONICS. In contrast

with the between–group difference concerning direct improvement of phoneme awareness, mean improvement of phonological awareness on the levels of syllable and onset-rime (t2–t1) was not significantly different for any pair of groups.

Long-term improvement of phonological awareness

The exact version of the Mann-Whitney U test revealed no significant difference in mean improvement of phoneme awareness (t3–t1) according to BAKO 1–4 for any pair of groups. Thus, mean improvement of phoneme awareness (t3-t1) in PHON (M_{diff}=5.76, SD_{diff}=5.89) did not differ significantly from mean improvement in PHONICS (M_{diff}=6.93, SD_{diff}=4.57, P=0.579), READ (M_{diff}=4.57, SD_{diff}=4.16, P=0.684) and CON (M_{diff}=4.89, SD_{diff}=6.64, P=0.796). Furthermore, mean improvement of phoneme awareness (t3-t1) in PHONICS and READ (P=0.247), PHONICS and CON (P=0.315) as well as READ and CON (P=0.971) did not differ significantly either.

Similarly, the exact version of the Mann-Whitney U test revealed no significant differences in mean improvement of phonological awareness on the levels of syllable and onset-rime (t3–t1) according to QUIL-D for any pair of groups. Thus, mean improvement of phonological awareness (t3–t1) in PHON (M_{diff}= 0.08, SD_{diff}=0.07) did not significantly differ from mean improvement in PHONICS (M_{diff}=0.04, SD_{diff}=0.17, P=0.853), READ (M_{diff}=0.10, SD_{diff}=0.11, P=0.912) and CON (M_{diff}=0.07, SD_{diff}=0.07, P=0.912). Furthermore, mean improvement of phonological awareness (t3–t1) in PHONICS and READ (P=0.684), PHONICS and CON (P=1.000) as well as READ and CON (P=0.853) did not differ significantly either.

Altogether, the results indicate that at t3 long-term improvement of phonological awareness on the levels of phoneme, syllable and onset-rime, as compared to t1, is relatively similar across all groups, with a deviating pattern in phoneme awareness at t2. For a survey of the significant improvements (t2–t1, t3–t1) within each group for phoneme awareness and phonological awareness on the levels of syllable and onset-rime see Table III.

DISCUSSION

The present study compared the effects of three training programs with respect to recoding, decoding

and reading comprehension in German-speaking dyslexic third and fourth graders with weak phonological awareness skills. The major goal of the present study was to investigate whether phonological awareness training is an effective intervention to significantly improve reading skills in German dyslexic third and fourth graders with weak phonological awareness skills, and whether these children can equally benefit from a phonology-based reading training or a visually-based reading training.

First of all, we found that these children can differentially benefit from a phonology-based reading training in terms of directly improved decoding. Furthermore, we were able to show that two effective ways to directly improve reading comprehension involve either a phonological awareness training (PHON) or a visually-based reading training (READ) but not a phonology-based reading training (PHONICS). However, despite divergent patterns at t2, the long-term improvement of decoding and reading comprehension at t3 as compared to t1 was relatively similar across all training groups, irrespective of the training the children got. The direct effect on reading comprehension shared by the two most contrastive interventions deserves special consideration and will be discussed before the direct effect on decoding found for PHONICS in the following section.

Direct effects on reading comprehension

Effects of phonological awareness training on reading comprehension were to be expected on the base of two cross-linguistic meta- analyses (Bus and van Ijzendoorn 1999, Ehri et al. 2001b). The findings of the present study are in line with both analyses, as both found a positive impact of phonological awareness trainings on reading comprehension. Furthermore, an effect of phonological awareness training on reading comprehension seems reasonable, as several studies concerned with consistent orthographies suggest that phonological awareness plays an important role in the development of reading comprehension throughout the period of primary school (Schneider and Näslund 1993, Landerl 2001, 2003, Müller and Brady 2001, Schiff et al. 2011).

Usually, it is suggested that the effect of phonological awareness on reading comprehension is mediated by an effect on decoding. However, in the present study a clear effect on reading comprehension was found although the phonological awareness training did not significantly improve decoding. This is in line with the result of a study which explored the impact of phonological awareness on reading comprehension in a large sample of Norwegian first graders (Engen and Hoien 2002). Thus, Engen and Hoien (2002) found that phonological awareness had a direct impact on reading comprehension above the indirect phonological effect mediated by decoding skills. They proposed that the direct effect of phonological awareness on reading comprehension might be due to the fact that phonological awareness is related to further skills like vocabulary knowledge. Vocabulary knowledge, in turn, has been found to be a critical factor in reading comprehension development in consistent orthographies and the rather inconsistent English orthography (e.g., de Jong and van der Leij 2002, Muter et al. 2004). Thus, future assessment of vocabulary would be an important extension in order to investigate whether vocabulary knowledge contributes to the effects on reading comprehension found in this study.

The direct effect of READ on reading comprehension seems more straightforward than the effect of PHON on this particular skill, since repeated reading of orthographic patterns and sight words was trained. Originally, repeated reading techniques aimed to improve automatized word recognition, which, in turn, provides cognitive capacity for reading comprehension beyond the word level (LaBerge and Samuels 1974, Samuels 1979). Accordingly, meta-analyses for repeated reading techniques reported positive effects on reading comprehension (meta-analysis of the NRP, 2000, Therrien 2004). One typical feature of these analyzed programs was that passages were read aloud several times. Moreover, further studies found that the effect of repeated reading is not limited to repeated reading of passages but applies to word reading, too. Thus, interventions similar to the training used in the current study, focusing on repeated reading of isolated sight words, also transferred to reading comprehension in English speaking dyslexic children (e.g., Levy et al. 1997, McArthur et al. 2013).

Repeated reading has so far not been a common method to enhance reading skills in German. One out of only few published studies on repeated reading programs concerned with German dyslexic primary school children focused on the repeated reading of onset clusters and words (Thaler et al. 2004). The authors reported improved reading fluency of trained

words and small generalization effects to untrained words, but unfortunately did not assess effects on reading comprehension. In the study by Thaler and coauthors (2004) a very limited set of orthographic patterns and training words was clearly presented more often (four different onset clusters occurring in eight training words each were presented up to 150 times) compared to the present study. However, results of the present study suggest that an intervention providing a larger set of orthographic patterns and training words (up to 21 orthographic patterns occurring in seven training words each), which is presented with rather low frequency (eight to eleven presentations per word) may also result in significant effects on reading skills. Therefore, setting a focus on reading a wide range of stimuli sharing a large set of orthographic patterns instead of setting a focus on a very high number of reading repetitions seems suitable to transfer to reading comprehension.

Finally, particularly emphasizing orthographic similarities between words might be a further aspect leading to the transfer on reading comprehension. Children seem to be unable to abstract orthographic representations at the sub-lexical level when trained at the lexical level (e.g., Thaler et al. 2004). Thus, the explicit instruction to focus on orthographic similarities between words used in the present study may not only have set up single orthographic representations but might also have made children generally aware that orthographic consistencies exist within words (Conrad and Levy 2011). This "orthographic awareness" (Conrad and Levy 2011) may have improved reading comprehension in turn.

The absent direct effect on reading comprehension by PHONICS instruction in the current study contrasts with the small but significant effect on reading comprehension revealed by the meta-analysis in older dyslexic readers (Ehri et al. 2001a). The meta-analysis revealed the effectiveness of PHONICS instruction approaches not only concerned with phonemes but with larger sub-syllabic units, too (e.g., onsets and rimes). Unfortunately, in the meta-analysis effects of syllable-based PHONICS instruction were not explored, which was an important part of PHONICS in the current study. Also contrary to the present study, McArthur and coworkers (2013) recently found a direct effect of phonics instruction on reading comprehension in children with dyslexia, too. Thus, it is necessary to discuss, why PHONICS did not lead to a

direct improvement of reading comprehension. Very similar to our study, in the study of McArthur and others (2013) the children were asked to do five 30-min sessions per week, but the children received the double number of sessions (40 sessions in total vs. 20 sessions in the current study). Thus, for future work, it seems reasonable to increase the total duration of PHONICS instruction. This would allow all children to progress through all training levels, because only some of the children in the present study progressed fast enough to practice reading the most difficult grapheme-phoneme associations of the program (e.g., including consonant clusters). Furthermore, findings of the meta-analysis by Suggate (2010) indicate that phonics instruction generally has a larger impact on reading in younger than in older children. Similarly, the meta-analysis of the NRP reported by Ehri and others (2001a) revealed smaller effects on reading when PHONICS instruction was administered beyond first grade and larger effect sizes for fluency instruction than for PHONICS instruction in second through sixth grade. Thus, in view of the age of the dyslexics in the present study it should be considered to follow the proposal of Ehri and colleagues (2001a) to couple PHONICS instruction in older children with other effective methods of reading instruction. This proposal is in line with the study of McArthur and coauthors (2013), who compared phonics training and sight word training in children with dyslexia and explored if different orders of these training have different effects on reading skills. Both kinds of trainings were found to improve reading comprehension. Further, implementing phonics instruction before sight word reading instruction had a small advantage over the reverse order. Thus, considering McArthurs`s findings it seems promising to complement the PHONICS instruction used in the present study with sight word reading. Findings of the present study also suggest that a combination of PHONICS and sight word reading might be a promising approach. The most substantial direct improvement of decoding in the current study was found for PHONICS. Thus, the significant direct improvement of the PHONICS training on decoding as well as the significant effect of sight word training (READ) on reading comprehension in the current study suggest that the combination of both types of training might lead to wide-ranging effects on different reading skills in German children with dyslexia.

Direct effects on decoding

In contrast to PHON and READ, improvement of basic reading skills would rather be expected for the PHONICS condition. In the phonology-based training the routines of grapheme-phoneme blending and of blending larger subunits of words (syllables) were taught explicitly. Indeed, the direct effect on decoding found for this training is in line with findings from a meta-analysis of the NRP reported by Ehri and coworkers (2001a), which revealed that PHONICS instruction had a particular impact on decoding of pseudo-words in dyslexic readers from second up to sixth grade. Furthermore, the finding that a syllable-based PHONICS training improves decoding is in line with a previous German study on training effects for the KLA program used here (Strehlow et al. 1999). In the cited study German dyslexics were trained, but no control group was included. In contrast, a recent study including a KLA training group, a spelling training group, a dyslexic and a normally reading control group found no effect on standardized values of decoding for the KLA (Groth et al. 2013). Differences concerning the inclusion criteria, (reading and spelling problems vs. reading problems in the present study) and the intensity of trainings (twice a week vs. five times a week in the present study) might explain the selective effect of the KLA on decoding in the present study, which was absent in the study of Groth and others (2013).

As mentioned above, the effect on decoding was found for the PHONICS intervention but not for the PHON intervention. In contrast to the present study, earlier work concerned with German dyslexics did report a positive impact of pure phonological awareness training (as provided in the PHON group) on decoding (Schneider et al. 2000). According to findings of the meta-analysis of the NRP, Ehri and coauthors (2001b) reported that phonological awareness instruction focusing on one or two skills has a larger effect on basic reading skills than phonological awareness instruction training multiple phonological awareness skills. Based on these results it seems reasonable to assume that reducing the administered training to only one to two tasks at the phoneme level exclusively instead of multiple tasks operating with either syllables, onset-rhymes, or phonemes would have led to effects on basic reading skills like recoding or decoding. Furthermore, combining phonological awareness and letter training might have facilitated effects for recoding or decoding (e.g., Bus and van Ijzendoorn 1999, Roth and Schneider 2002).

Similar to the PHON intervention, no effect on decoding was found for the READ group. Effects on decoding are not necessarily expected after repeated reading training of sight words. Missing effects of repeated reading on these reading skills in the READ group are thus not surprising, because the primary theoretical motivation of reading orthographic patterns and sight words is not to strengthen single grapheme-phoneme correspondences but to improve automatized visual word recognition (LaBerge and Samuels 1974, Samuels 1979).

Stability of training effects over time

Long-term effects on reading comprehension

In contrast to the effect on reading comprehension found for READ, the effect of PHON on reading comprehension remained stable until follow-up assessment (t3) when compared to CON. Furthermore, long-term effects for PHON and for READ did not exceed the long-term improvements found for PHONICS with respect to reading comprehension any longer. Thus, although PHONICS did not lead to a direct improvement of reading comprehension, this group caught up with the other training groups until t3. However, the long-term improvement found for PHONICS with respect to reading comprehension still was not strong enough to significantly exceed the long-term improvement found for CON. Furthermore, only within PHON and READ but still not in PHONICS the long-term improvements represent a substantial change of reading comprehension, respectively. In summary, despite divergent patterns at t2, the long-term improvement of reading comprehension at t3 as compared to t1 tended to be relatively similar across all training groups, irrespective of the training the children got. The divergent pattern at t2 speaks in favor of different direct mechanisms involved in the improvement of reading, when different approaches are applied. However, the relatively similar pattern at t3 reveals, that, in the long run, reading comprehension can be consistently improved based on these different direct mechanisms.

No significant effects on decoding were found according to between-group comparisons. This aspect will be addressed in the following section.

Long-term effects on decoding

In contrast to the favorable long-term results of PHON on reading comprehension, the long-term effect on decoding found only for PHONICS was less stable. Nevertheless, the fact that there was still a trend for a significant increase in decoding from t1 to t3 (see Table II) puts the slight decrease from t2 to t3 into perspective. In sum, the absent long-term effect on decoding as well as the missing direct impact on reading comprehension found for the PHONICS training, require reconsideration of this intervention. Proposals to introduce improvements have already been discussed above.

Effects on phonological awareness

The between-group comparisons revealed that the direct improvement of phoneme awareness (as measured by the BAKO 1–4 score) found for PHONICS was more substantial than the direct improvement found for READ, and only improvement (t2–t1) of phoneme awareness in PHONICS was significant in the paired-sample test. It is in line with findings concerning the development of phoneme awareness in younger German children that the development of poor phoneme awareness in German dyslexic children in the advanced stage of reading acquisition seems to be facilitated most effectively by the PHONICS training. Thus, the initial development of phonemic awareness in German speaking children is usually facilitated as soon as the phonics instruction begins in primary school (Wimmer et al. 2000). Hence, additional phonics training after the formal phonics instruction in primary school is terminated provides added value with regard to phoneme awareness in older German children with dyslexia and weak phonological awareness.

In contrast to the direct effect on phoneme awareness found for PHONICS, direct improvement of phonological awareness on the level of syllable and onset-rime (as assessed with the QUIL-D score) was not significantly different in any pair of groups. Interestingly, only improvement of phonological awareness at the levels of syllables and onset-rime (according to QUIL-D) within PHON was significant in the paired-sample test, whereas this time, no significant improvement within PHONICS was found. Thus, phonological awareness training in older German primary school children focusing on multiple phonological units without any reference to written language seems to particularly facilitate awareness of phonological units larger than phonemes.

Furthermore, the paired-sample test revealed that phoneme awareness within CON (as measured by BAKO) tended to improve. Whereas the finding that phonological awareness directly improved after both phonologically oriented trainings (PHON and PHONICS) is supposed to reflect specific effects of training, this does not apply to the non-trained control group. Thus, not only in the initial (Wimmer et al. 2000) but also in the advanced stage of reading acquisition typically developing German speaking children can further develop their phoneme awareness in the course of regular reading practice in primary school without any additional phonological training.

In contrast, this was not the case in the dyslexic children in the present study. Thus, children, who received a visually-based reading training, were not able to improve phonological awareness in the short run. As phonological awareness did not significantly improve until t3 in the READ condition, this late phonological increase seems to imply enhanced reading skills, which were directly evident at t2. This leads to two conclusions concerning the reciprocal relationship between reading and phonological awareness: First, a certain level of reading ability seems to be needed, before phonological awareness can further emerge naturally in German dyslexic third and fourth graders. Second, improvements of reading in older German dyslexic primary school children are not necessarily matched by an improvement in phonological awareness, because reading comprehension was directly improved by a visually based reading training, although the phonological deficit remained unchanged at first. Thus, this part of our results indirectly supports Blomert and Willems (2010) who claimed that reading and PA are mainly related reciprocally.

Generalizability and future research

To our knowledge, our study is the first to compare effects of a phonology-based reading training with those of an intervention purely focusing on phonological awareness and those of a visually-based reading training in older dyslexic children learning to read the rather consistent German orthography. First of all, we found that these children can differentially benefit

from a phonology-based reading training (PHONICS) in terms of directly improved decoding. Furthermore, we were able to show that two effective ways to directly improve reading comprehension involve either a phonological awareness training (PHON) or a visually-based reading training (READ) but not a phonology-based reading training (PHONICS).

However, despite divergent patterns at t2, long-term improvement of decoding and reading comprehension at t3 as compared to t1 was relatively similar across all training groups, irrespective of the training the children got. Although reliable direct training effects could be demonstrated, some limitations for generalization should be considered.

Sample characteristics

For ethical reasons, we did not include untreated dyslexic children. Instead, a sample of typically reading children was chosen for control purposes. Thus, effects found when comparing training groups with the control group should be interpreted with caution. However, we found a direct effect on decoding not only when comparing the phonology-based reading training to the control group, but when comparing this group to the phonological awareness training, too. Similarly, we were able to show selective short-term effects on reading comprehension not only when comparing the visually-based reading training or phonological awareness training to the control group, but when comparing these groups to the phonology-based reading training, too. This suggests that the direct effects on decoding and reading comprehension are robust and are not only based on the comparison with typically reading children. Furthermore, in a recent training study concerned with German primary school children it has been doubted that it is realistic to ask parents to abstain from any kind of training for their dyslexic child during the period of a study (Groth et al. 2013). Particularly parents of German third and fourth graders are under considerable pressure to find an immediate training to remediate the reading deficit of their dyslexic child as the assignment to the different forms of secondary school already takes place during the fourth grade. Thus, although we are aware that the inclusion of non-treated dyslexic children might have been a suitable solution in another context, in the present study, the inclusion of non-treated typically reading control children instead of the inclusion of non-

treated dyslexic children was the only possibility to include control children at all. Considering the situation of German parents of dyslexic third and fourth graders described above, including a dyslexic waiting control group would not have excluded the possibility that children in this group might receive an additional intervention during the period of the study.

Furthermore, the study was based on relatively small sample sizes. However, the between-group comparisons, which were corrected for multiple testing per outcome variable, resulted in selective and direct effects. Similarly, results of the paired-sample tests with adjusted alpha levels confirmed the robustness of these selective and direct effects. Still, larger samples are needed to corroborate the selective short-term effects on decoding and reading comprehension found in the present study.

Another limitation concerns the identification of dyslexic children in our study (reading quotient <90 in a reading screening and percentile <25 in at least one of three subtests of the standardized reading test). Trainings implemented with dyslexic children with more severe reading impairments might have led to larger training effects. Nevertheless, third and fourth graders included as dyslexics in our study failed to respond to previous classroom-teaching and were in need of more intensive support (cf. Snowling and Hulme 2011). Further issues concern the generalization across different stages of reading development and across languages. By training children who learn to read the rather consistent German orthography in the advanced stage of reading acquisition, we could yet demonstrate effects of phonological training elements on reading skills. Thus, an important future issue is whether these effects diminish in older dyslexic children learning consistent orthographies vs. equally young dyslexics learning less consistent orthographies.

Outcome measures and analysis

Future assessment of grammatical or lexical processes would be an important extension in order to investigate whether they contribute to the effects on reading comprehension found in this study. Furthermore, particularly the selective short-term effects on reading comprehension found for two of three training methods in contrast to the relatively similar long-term effect on reading comprehension across all training groups

suggest that the process of reading comprehension might be differentially addressed by the training methods used in the present study. Thus, future research should include an investigation of neuro-functional changes induced by each of the single trainings. This might contribute to a more differentiated understanding of the mechanisms by which each of these training types work in dyslexic children.

Implications for the neuro-cognitive basis of reading and dyslexia

Trainings for dyslexic children work – and they have repercussions in their brains. However, in contrast to early suggestions of rather consistent hypo-activation across languages (Paulesu et al. 2001), more recent data revealed some substantial variability in the deviant brain activation patterns in dyslexic readers (Richlan et al. 2011), which also differ depending on task demands (Heim et al. 2010a). Depending on whether a dyslexic child has phonological or visual problems, the underlying neural processes may differ (Heim et al. 2010b).

The current study is relevant for the understanding of these findings in several ways. First, the present data suggest that the task demands must be chosen carefully with respect to recoding, decoding, or reading comprehension. Second, as far as training-induced brain activation changes are concerned (e.g. Temple et al. 2003, Gabrieli 2009), the nature of the training is very likely to interact with these task demands. One example is a recent paper by Heim and coworkers (2014), in which a subset of the dyslexic children who were trained in the present study underwent fMRI scanning before and after training. Due to limited time inside the scanner, the only reading task involved isolated words. The most interesting finding was that all training groups showed a shared pattern of training-induced brain activation increase in the visual word form area of the left inferior occipito-temporal cortex – a finding that was accompanied partly by differential training effects specific for one but not the other training. Thus, the mental force of various training types might, at least in part, differ. The Heim and colleagues (2014) fMRI study was only one first step towards assessing differential training effects for children with unique reading deficit patterns under distinguishable task demands (i.e. reading vs. visuo-spatial attention). The findings of the present study may help to shed further light onto the differential ways by which training of sight words or phonological abilities can trigger neuro-cognitive processes. It would be desirable that they stimulate further research and thus contribute to reducing the apparent – or seeming – variability in studies of dyslexia.

CONCLUSIONS

Phonological awareness training is an effective intervention to significantly improve reading comprehension in German dyslexic third and fourth graders with a phonological awareness deficit. However, these children can equally benefit from a visually-based reading training (repeated reading of orthographic patterns and sight words) in terms of directly improved reading comprehension. Thus, phonological awareness may, but does not need to be part of reading remediation in older primary school children with dyslexia learning a consistent orthography. Rather, an orthographic reading strategy might compensate for the phonological deficit in dyslexic children after the initial stage of reading acquisition. Moreover, German children with dyslexia can benefit from a phonology-based reading training (phonics instruction) in terms of improved decoding but not in terms of directly improved reading comprehension. Thus, a more comprehensive approach, complementing the phonology-based training with the repeated reading of sight words might be an effective method to combine the direct effect on decoding with an additional direct effect on reading comprehension.

REFERENCES

Adams MJ (1993) Beginning to Read: Thinking and Learning about Print. MIT Press, Cambridge, MA, USA.

Angermayer MJ (1974) Test for Psycho-Linguistic Development (in German). Beltz, Weinheim, DE.

Anthony JL, Lonigan CJ (2004) The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. J Educ Psychol 96: 43–55.

Arnold EM, Goldston DB, Walsh AK, Reboussin BA, Daniel SS, Hickman E (2005) Severity of emotional and behavioral problems among poor and typical readers. J Abnorm Child Psychol 33: 205–217.

Blomert L, Willems G (2010) Is there a causal link from a phonological awareness deficit to reading failure in children at familial risk of dyslexia? Dyslexia 16: 300–317.

Bus AG, van Ijzendoorn MH (1999) Phonological awareness and early reading: A meta-analysis of experimental training studies. J Educ Psychol 91: 403–414.

Caravolas M, Volin J, Hulme C (2005) Phoneme awareness is a key component of alphabetic literacy skills in consistent and inconsistent orthographies: Evidence from Czech and English children. J Exp Child Psychol 92: 107–139.

Clarkson-Grabs T (2006) Variabolus (in German). Veris, Kiel, DE.

Conrad NJ, Levy BA (2011) Training letter and orthographic pattern recognition in children with slow naming speed. Read Writ 24: 91–115.

de Jong PF, van der Leij A (1999) Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. J Educ Psychol 91: 450–476.

de Jong PF, van der Leij A (2002) Effects of phonological abilities and linguistic comprehension on the development of reading. Sci Stud Read 6: 51–77.

de Jong PF, van der Leij A (2003) Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. J Educ Psychol 95: 22–40.

Denckla MB, Rudel RG (1976) Rapid „automatized" naming (R.A.N): dyslexia differentiated from other learning disabilities. Neuropsychologia 14: 471–479.

Dodd B, Holm A, Oerlemans M, McCormick M (1996) Queensland University Inventory of Literacy (QUIL). Department of Speech Pathology and Audiology, The University of Queensland, St Lucia, Queensland, AU.

Doesborgh SJ, van de Sandt-Koenderman MW, Dippel DW, van Harskamp F, Koudstaal PJ, Visch-Brink EG (2004) Effects of semantic treatment on verbal communication and linguistic processing in aphasia after stroke: a randomized controlled trial. Stroke 35: 141–146.

Dummer-Smoch L, Hackethal R (2007a) Kiel Program for Building up Reading Skills (in German). Veris, Kiel, DE.

Dummer-Smoch L, Hackethal R (2007b) The New Carolus. Learning Software for the Kiel Program for Building up Reading skills and for the Kiel Program for Building up Spelling (Version 5.0, in German). Veris, Kiel, DE.

Ehri LC, Nunes SR, Stahl SA, Willows DM (2001a) Systematic PHONICS instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. Rev Educ Res 71: 393–447.

Ehri LC, Nunes SR, Willows DM, Schuster BV, Yaghoub-Zadeh Z, Shanahan T (2001b) Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. Read Res Q 36: 250–287.

Engen N, Hoien T (2002) Phonological skills and reading comprehension. Reading and Writing: An Interdisciplinary Journal 15: 613–631.

Gabrieli JDE (2009) Dyslexia: A new synergy between education and cognitive neuroscience. Science 325: 280–283.

Gresham FM, Vellutino FR (2010) What is the role of intelligence in the identification of specific learning disabilities? Issues and Clarifications. Learning Disabilities Research and Practice 25: 194–206.

Groth K, Hasko S, Bruder J, Kunze S, Schulte-Körne G (2013) Measurement of treatment in developmental dyslexia under methodological considerations (in German). Lernen und Lernstörungen 2: 161–175.

Hatcher PJ, Hulme C, Ellis AW (1994) Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. Child Dev 65: 41–57.

Heim S, Tschierse J, Amunts K, Wilms M, Vossel S, Willmes K, Grabowska A, Huber W (2008) Cognitive subtypes of dyslexia. Acta Neurobiol Exp (Wars) 68: 73–82.

Heim S, Grande M, Pape-Neumann J, van Ermingen M, Meffert E, Grabowska A, Huber W, Amunts K (2010a) Interaction of phonological awareness and 'magnocellular' processing during normal and dyslexic reading: behavioural and fMRI investigations. Dyslexia 16: 258–282.

Heim S, Grande M, Meffert E, Eickhoff SB, Schreiber H, Kukolja J, Shah NJ, Huber W, Amunts K (2010b) Cognitive levels of performance account for hemispheric

lateralisation effects in dyslexic and normally reading children. Neuroimage 53:1346–1358.

Heim S, Pape-Neumann J, van Ermingen-Marbach M, Brinkhaus M, Grande M (2014) Shared vs. specific brain activation changes in dyslexia after training of phonology, attention, or reading. Brain Struct Funct: 1–17.

Hollbach H (1999) Hearing Training for the Development of Phonological Awareness (in German). Verlag für Lerntherapeutische Medien, Leer, DE.

Holm S (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6: 65–70.

Hoover WA, Gough PB (1990) The simple view of reading. Reading and Writing: An Interdisciplinary Journal 2: 127–160.

Küspert P, Roth E, Schneider W, Laier R (2001) Würzburg Program for the Training of Phonological Awareness (Multimedia Version, in German). Laier und Becker Psychologie und Multimedia GbR, Dielheim, DE.

LaBerge D, Samuels J (1974) Towards a theory of automatic information processing in reading. Cogn Psychol 6: 293–323.

Landerl K (2001) Word recognition deficits in German: More evidence from a representative sample. Dyslexia 7: 183–196.

Landerl K (2003) Cognitive Deficits in children with poor reading skills (in German). Psychologie in Erziehung und Unterricht 50: 369–380.

Landerl K, Wimmer H (2000) Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. Appl Psycholinguist 21: 243–262.

Landerl K, Wimmer H (2008) Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. J Educ Psychol 100: 150–161.

Landerl K, Wimmer H, Frith U (1997a) The impact of orthographic consistency on dyslexia: A German-English comparison. Cognition 63: 315–334.

Landerl K, Wimmer H, Moser E (1997b) Salzburg Reading- and Spelling Test, SLRT (in German). Hans Huber, Bern, CH.

Landerl K, Ramus F, Moll K, Lyytinen H, Leppänen PH, Lohvansuu K, Schulte-Körne G (2013) Predictors of developmental dyslexia in European orthographies with varying complexity. JCPP 54: 686–694.

Levy BA, Abello B, Lysynchuk L (1997) Transfer from word training to reading in context: Gains in reading fluency and comprehension. Learn Disabil Q 20: 173–188.

Lindner M, Grissemann H (2000) Zürich Reading Test (sixth edition, in German). Hans Huber, Bern, CH.

Manis FR, Custodio R, Szeszulski PA (1993) Development of phonological and orthographic skill: a 2-year longitudinal study of dyslexic children. J Exp Child Psychol 56: 64–86.

Mann V, Wimmer H (2002) Phoneme awareness and pathways into literacy: A comparison of German and American children. Reading and Writing: An Interdisciplinary Journal 15: 653–682.

Marx H (1998) Knuspel`s Reading Tasks, Knuspel-L: Manual (in German). Hogrefe, Göttingen, DE.

Mayer A (2008) Phonological awareness, rapid naming and automatized processes of reading (in German). Shaker, Aachen, DE.

Mayer A (2009) Word Recognition at Lightning Speed. Foundations and Practice (in German). Borgmann Media, Dortmund, DE.

Mayringer H, Wimmer H (2003) Salzburg Screening for Reading in Grade 1-4, SLS 1-4 (in German). Hans Huber, Bern, CH.

McArthur G, Castles A, Kohnen S, Larsen L, Jones K, Anandakumar T, Banales E (2013) Sight word and phonics training in children with dyslexia. J Learn Disabil: 1–17.

Moll K, Fussenegger B, Willburger E, Landerl K (2009) RAN is not a measure of orthographic processing. Evidence from the asymmetric German orthography. Sci Stud Read 13: 1–25.

Muter V, Hulme C, Snowling MJ, Stevenson J (2004) Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. Dev Psychol 40: 663–681.

Müller K, Brady S (2001) Correlates of early reading performance in a transparent orthography. Read Writ 14: 757–799.

NATIONAL READING PANEL (2000) Report of the National Reading Panel:Teaching children to read**:** An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction Reports of the subgroups. National Institute of Child Health and Human Development Clearinghouse, Washington, DC, USA.

Patel TK, Snowling MJ, de Jong PF (2004) A cross-linguistic comparison of children learning to read in English and Dutch. J Educ Psychol 96: 785–797.

Paulesu E, Démonet JF, Fazio F, McCrory E, Chanoine V, Brunswick N, Cappa SF, Cossu G, Habib M, Frith CF, Frith U (2001) Dyslexia: Cultural diversity and biological unity. Science 291: 2165–2167.

Ramus F (2003) Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? Curr Opin Neurobiol 13: 212–218.

Richlan F, Kronbicher M, Wimmer H (2011) Meta-analyzing brain dysfunctions in dyslexic children and adults. Neuroimage 56: 1735–1742.

Roth E, Schneider W (2002) Long-term effects of training phonological awareness and letter-knowledge on literacy acquistion (in German). German Journal of Educational Psychology 16: 99–108.

Samuels SJ (1979) The method of repeated readings. Read Teach 32: 403–408.

Schiff R, Schwartz-Nahshon S, Nagar R (2011) Effect of phonological and morphological awareness on reading comprehension in Hebrew-speaking adolecents with reading disabilities. Ann Dyslexia 61: 44–63.

Schneider W, Küspert P, Roth E, Visé M, Marx H (1997) Short- and longterm effects of training phonolgical awareness in kindergarten: Evidence from two german studies. J Exp Child Psychol 66: 311–340.

Schneider W, Näslund JC (1993) The impact of early meta-linguistic competencies and memory capacity on reading and spelling in elementary school: results of the Munich longitudinal study on the genesis of individual competencies (LOGIC). EJPE 8: 273–287.

Schneider W, Roth E, Ennemoser M (2000) Training phonological skills and letter knowledge in children-at-risk for dyslexia: A comparison of three kindergarten intervention programs. J Educ Psychol 92: 284–295.

Shaywitz SE, Shaywitz BA (2005) Dyslexia (specific reading disability). Biol Psychiatry 57: 1301–1309.

Snowling MJ (1981) Phonemic deficits in developmental dyslexia. Psychol Res 43: 219–234.

Snowling MJ (1995) Phonological processing and developmental dyslexia. J Res Read 18: 132–138.

Snowling MJ, Hulme C (2011) Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. Br J Educ Psychol 81: 1–23.

Stock C, Marx P, Schneider W (2003) Basic Skills for Reading and Spelling, BAKO 1-4 (in German). Hogrefe, Göttingen, DE.

Strehlow U, Haffner J, Busch G, Pfüller U, Rellum T, Zerahn-Hartung C (1999) Concentrating on weaknesses or compensating by strengths? Comparison of 2 strategies for remediation of children with comprehensive reading-spelling disorder (in German). Z Kinder Jugendpsychiatr Psychother 27: 103–113.

Sturm W (2009) Pre-Tests for Perception Functions, WAFW. Schuhfried, Mannheim, DE.

Suggate SP (2010) Why what we teach depends on when: Grade and reading intervention modality moderate effect size. Dev Psychol 46: 1556–1579.

Tallal P (1980) Auditory temporal perception, PHONICS, and reading disabilities in children. Brain Lang 9: 182–198.

Temple E, Deutsch GK, Poldrack RA, Miller SL, Tallal P, Merzenich MM, Gabrieli MDE (2003) Neural deficits in children with dyslexia ameliorated by behavioural remediation: Evidence from functional MRI. Proc Natl Acad Sci U S A 100: 2860–2865.

Thaler V, Ebner EM, Wimmer H, Landerl K (2004) Training reading fluency in dysfluent readers with high reading accuracy: Word specific effects but low transfer to untrained words. Ann Dyslexia 54: 89–113.

Therrien WJ (2004) Fluency and comprehension gains as a result of repeated reading: A meta-analysis. Remedial Spec Educ 25: 252–261.

Torgesen JK, Wagner RK, Rashotte CA, Burgess S, Hecht S (1997) Contributions of phonological awareness and rapid automatized naming ability to growth of word-reading skills in second- to fifth-grade children. Sci Stud Read 1: 161–185.

Vellutino FR, Fletcher JM, Snowling MJ, Scanlon DM (2004) Specific reading disability (dyslexia): What have we learned in the past four decades? JCPP 45: 2–40.

van Ermingen-Marbach M, Verhalen N, Grande M, Heim S, Mayer A, Pape-Neumann J (2014) Standards for rapid automatised naming performances in normal reading children at the age of 9–11 (in German). Sprache Stimme Gehör 38: e28–e32.

Vidyasagar TR, Pammer K (2010) Dyslexia: a deficit in visuo-spatial attention, not in phonological processing. Trends Cogn Sci 14: 57–63.

Weiss RH (1998) German version of the Cattell Culture Fair Test 20, CFT 20. Hogrefe, Göttingen, DE.

Wimmer H (1993) Characteristics of developmental dyslexia in a regular writing system. Appl Psycholinguist 14: 1–33.

Wimmer H (1996) The early manifestation of developmental dyslexia: Evidence from German children. Read Writ 8: 171–188.

Wimmer H, Landerl K, Linortner R, Hummer P (1991) The relationship of phonemic awareness to reading acquisition: More consequence than precondition but still important. Cognition 3: 219–249.

Wimmer H, Mayringer H, Landerl K (1998) Poor reading: A deficit in skill-automatization or a phonological deficit? Sci Stud Read 2: 321–340.

Wimmer H, Mayringer H, Landerl K (2000) The double deficit hypothesis and difficulties in learning to read a regular orthography. J Educ Psychol 92: 668–680.

Zimmermann P, Gondan M, Fimm B (2002) Test of Attentional Performance for Children, KiTAP. Psytest, Herzogenrath, DE.