THE ROLE OF SEMANTICS

IN EARLY READING

Martine Gijsel

The research reported in this thesis was financially supported by Aspasia grant 015.000.028
from the Netherlands Organization for Scientific Research (NWO), provided to dr. Anna M. T. Bosman.

Cover design: Daniël Jansen
Print: PrintPartners Ipskamp, Enschede

ISBN-10: 90-9021409-7
ISBN-13: 978-90-9021409-2

The role of semantics in early reading

Een wetenschappelijke proeve op het gebied van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de Rector Magnificus prof. dr. C.W.P.M. Blom,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op dinsdag 9 januari 2007
om 15.30 uur precies door

Martine Alice Roger Gijsel
geboren op 11 maart 1976
te Terneuzen

Promotor: Prof. dr. L.T.W. Verhoeven

Copromotor: Dr. A.M.T. Bosman

Manuscriptcommissie:
Prof. dr. A.H. Neijt-Kappen
Dr. W.H.J. van Bon
Prof. dr. W.J.M. van den Broeck

In alle wetenschappen is het grootste aantal wetenschappers bezig met het ontwikkelen en het nalopen van ideeën van een klein aantal anderen.

Claude Bernard (1813-1878)

Frans fysioloog

Contents

Chapter 1: General introduction

Research in reading processes has a long history. Already in the 1880's, the first studies on word recognition were carried out by James McKeen Cattell, who was the first PhD-student of Wilhelm Wundt at the University of Leipzig (see for an overview of history of reading research, Venezky, 1984). By 1900, the research on reading processes shifted to American universities and since then, various topics of interest have been examined and different theoretical frameworks have dominated the literature. In the last decennia, the main interest of most studies predominantly concerned the role of phonology in reading (for overviews see Berent & Perfetti, 1995; Bosman & van Hell, 2002; Frost, 1995; Jared & Seidenberg, 1991; Van Orden, Pennington, & Stone, 1990) and the contribution of phonological skills to reading development (e.g., Ehri et al., 2001; Elbro, Borstrøm, & Petersen, 1998; O'Connor & Jenkins, 1999). The role of semantics on reading, on the other hand, is recognized, but has received less attention. This thesis investigates the role of semantics in word-decoding skills of Dutch students in primary grades.

The introduction starts with a discussion of selected theories on visual word-identification which have been most influential in the literature and at the same time assigned different roles to semantics, and concludes with the reading model that I preferred as the theoretical background of the present study. Next, aspects of successful reading intervention will be considered and several intervention programs will be discussed. Then, because the nature of the reading process partly depends on a language's orthography, a concise overview of the Dutch orthography is presented and consequences for beginning reading performance will be discussed. Finally, the outline of this thesis is presented and will be broadly embedded in the literature. A more exhaustive literature overview has been included in each of the chapters. Because all chapters have been written as independent papers for publication, it was inevitable that some relevant literature was discussed more than once. However, I tried to minimize these repetitions as much as possible.

Theories of word identification

In this section, three different types of reading models will be considered. These comprise dual-route models of reading (see for example, Coltheart, 1978), connectionist models (see for example, Seidenberg & McClelland, 1989), and the Phonological Coherence Model, which is a fully recurrent connectionist model (see for example, Van Orden, Pennington

& Stone, 1990). The section is not aimed at providing an extensive overview of all reading models that are postulated the last century.

<u>Dual-route Model</u>

The Dual-Route Model has known several slightly different versions, but the key assumption is the distinction of an internal lexicon and two different routes that operate independently and lead to the identification of a letter string: A lexical route, sometimes called the direct route or visual route, and a non-lexical route, sometimes referred to as the indirect route or phonological route (e.g., Coltheart, 1978; Coltheart et al., 2001; Jackson & Coltheart, 2001). This model is shown in Figure 1.

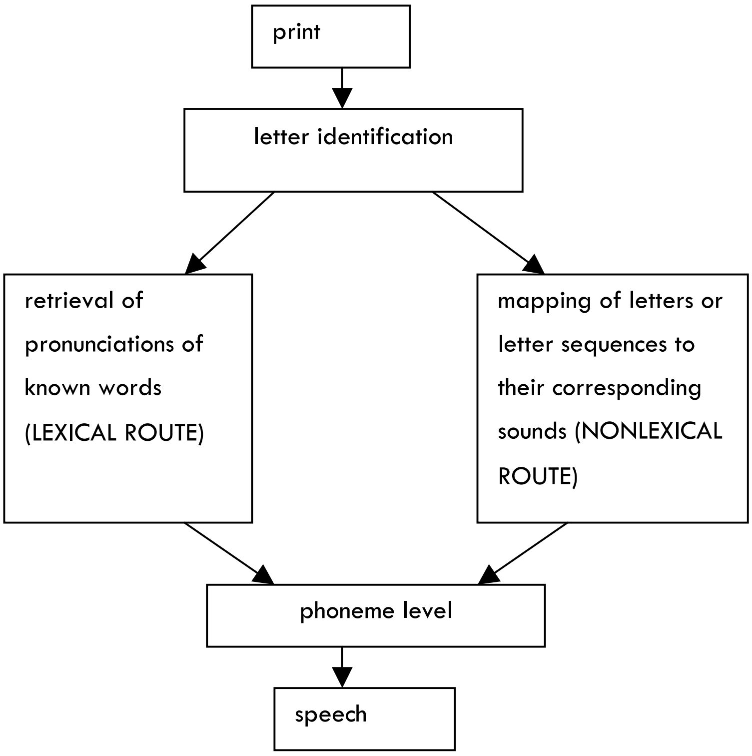

Figure 1. Dual-route model of reading aloud adapted from Jackson & Coltheart (2001, pp. 41)

It is assumed that the indirect route is used in reading regular words, that is, words that obey the predominant grapheme-to-phoneme correspondences of a language’s orthography, and pronounceable nonwords. These kind of letter strings are analyzed into graphemes (referred to as graphemic parsing) and subsequently, the corresponding phonemes are assigned to the graphemes, via a set of grapheme-phoneme correspondence (GPC) rules. As a result, the internal lexicon which contains semantic, phonological, and orthographic information can be accessed, depending on the task requirements. This procedure is not possible for reading ideographs and exception words, that is, words that disobey these rules. Therefore, a second route, also known as the direct route was proposed. In the direct route, the identification of a letter string occurs via visual access, thus without making use of GPC-rules. A letter string as a whole will activate the phonological form of the word in a phonological lexicon. This whole-word representation will in turn activate the meaning of the word in the semantic lexicon. For skilled reading, accuracy in both routes has to be acquired. In the original version of this theory (Coltheart, 1978), semantics played an inferior role and there was no feedback procedure included for semantics to influence the identification process. The Dual-Route Model was the dominant theory of reading in the eighties and nineties but in time, several assumptions of the model have been criticized. For example, experimental results have suggested that similar kinds of knowledge are used for the pronunciation of words and pseudowords (e.g., Glushko, 1979; Rosson, 1983) and it has been suggested that phonology is always activated in reading (e.g., Van Orden et al., 1990). Based on the criticism and experimental data that provided results that could not be explained within the framework, the model has been updated and has been implemented in a computational model, referred to as the Dual-Route Cascaded Model, from now on DRC-Model (Coltheart et al., 2001). The model is a generalization of the Interactive Activation Model of McClelland and Rumelhart (1981). The basic architecture of this model is shown in Figure 2.

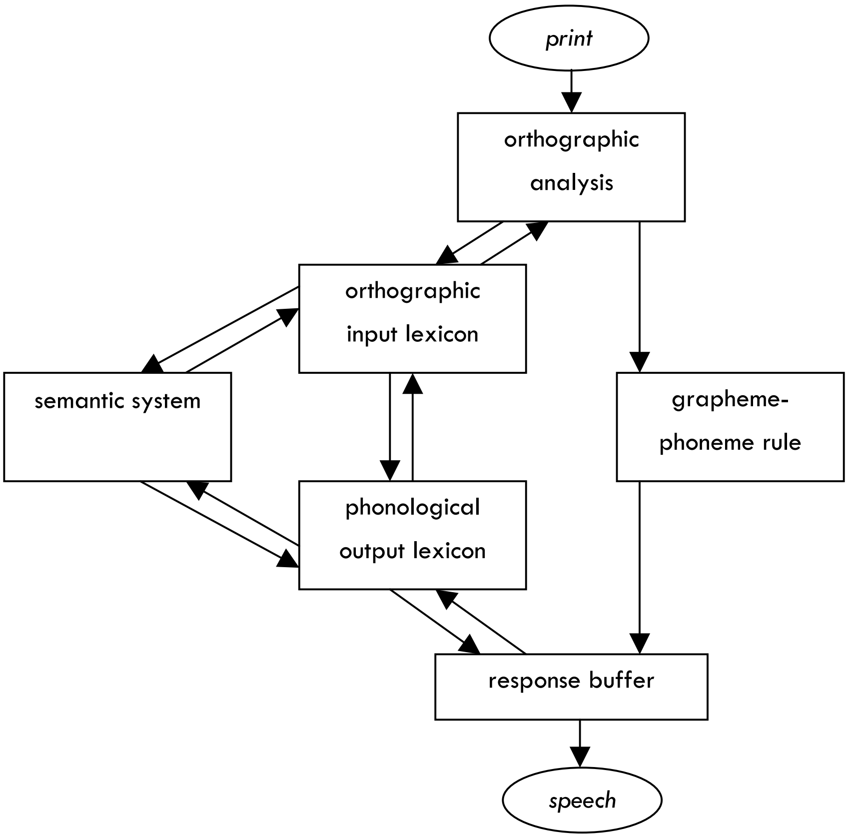

Figure 2. Basic architecture of the DRC-model of visual word recognition and reading aloud adapted from Coltheart et al. (2001, pp. 213)

The DRC-Model is a localist network, which means that concepts are represented by single network nodes. The term ‘cascaded’ refers to the property of the system that as soon as one level of processing is activated, subsequent levels will receive activation in a continuous way. The model consists of three routes: A GPC-route, lexical-semantic route, and lexical-non-semantic route. The lexical route distinguishes between a semantic route, in which activation flows from the orthographic lexicon via a semantic system to the phonological lexicon, and a non-semantic route, which does not include access to a semantic system. As shown in Figure 2, in the lexical route, activation flows bi-directionally. At the time the model was developed, however, feedback in the non-lexical route was not yet implemented. The model has successfully simulated a number of basic phenomena of reading and experimental results. For purposes of the present study, it is interesting to know how semantics has been implemented in the model. The authors did include a semantic system, which sends activation both to the phonological lexicon and orthographical lexicon, and receives information from these lexicons. Based on the results of Strain, Patterson, and Seidenberg (1995), who demonstrated shorter latencies and fewer regularization errors for high-imageability low-frequency words than for low-imageability low-frequency word, it is suggested that the lexical semantic system only contributes to correct word identification in skilled readers when the other two routes operate too slowly. This is the case for reading low-frequency irregular words, which cause competition between the lexical route and the non-lexical route. Semantic contribution is made possible because activation flows in both directions. Because the lexical-semantic route has not been implemented in the model, it is impossible to evaluate the assumed semantic contribution in the word-identification process.

Connectionist model

In the 1990's, a new class of models emerged that diverged from the symbolic view of reading: Connectionist models. These connectionist models used the brain as a metaphor for processing instead of the computer metaphor of the dual-route theories. There is, however, a certain degree of abstraction from the structure of the brain and the similarity to neurobiological networks is often at a coarse grain size. The Interactive-Activation (IA) models and Parallel-Distributed Processing (PDP) models are two types of connectionist models that abstract strongly form biological details.

The Interactive-Activation Model (McClelland & Rumelhart, 1981) consists of local symbol representations and the nodes in the model correspond to different linguistic units, that is, visual features, letters, and words. When a letter string is presented to the model, the visual-feature nodes excite position-coded letter nodes that contain these features and both excitatory and inhibitory activation is sent through the network. The model is interactive because activation from higher levels (e.g., words) influence the activation of lower levels (i.e., letters). The model does not include phonology, which is a serious limitation, because numerous studies have demonstrated phonological effects in reading (e.g., Van Orden et al., 1990) .

Seidenberg and McClelland (1989) took a different approach and developed a PDP-network model. In these types of models, the main focus is on the learning process itself: The reader learns by experience to establish relationships between different types of information (e.g., orthography and phonology). They assumed that words are distributed patterns of activation, rather than local units stored in memory. Another important fundamental difference with the DRC-Model is a rejection of the distinction between lexical and sublexical processes. The authors suggest that one single process can account for reading words and nonwords. The general architecture of the model is shown in Figure 3.

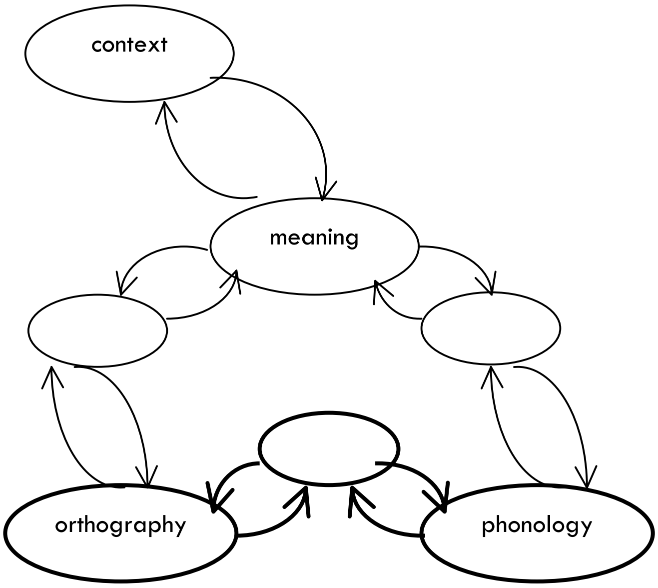

Figure 3. General framework of Seidenberg and McClelland's reading model, with the implemented model in boldface type adapted from Seidenberg & McClelland (1989, pp. 526)

In the model, three kinds of nodes are computed simultaneously, that is, semantic nodes, phonological nodes, and orthographic nodes. Connections between nodes have different strengths, reflected in their weights and are realized through hidden units. It is assumed that word identification takes place either via the phonological pathway (from orthography to phonology), or via the semantic pathway (from orthography to semantics to phonology), however, both pathways operate in a similar way (in contrast to the Dual Route Model in which two different mechanisms operate). The model makes use of supervised or error-correcting learning: If the output is not in the desired direction, the network weights are adjusted to produce the correct output, whereas if the output corresponds to the input, the weights are reinforced. In the computational model (Seidenberg & McClelland, 1989), only the phonological and orthographic nodes have been implemented. Based upon several limitations of the model and due to criticism concerning the poor performance on nonwords (e.g., Besner, Twilley, McCann, & Seergobin, 1990), the model has been updated by Plaut and his colleagues (Plaut, McClelland, Seidenberg, & Patterson,1996). This model accounts for a variety of experimental data and deals with semantics in a preliminary way, to simulate performance in acquired surface dyslexia. The main focus, however, is on the mapping of orthography to phonology. Harm and Seidenberg (2004) did include semantics in their computational reading model and focused on reading for meaning, rather than on the translation from orthography to phonology, as in the model of Seidenberg and colleagues (1989, 1996). They stressed the interdependence of all elements in the system and focused on how meanings are computed in a system with orthographic, phonologic, and semantic components. The authors successfully simulated a number of experimental results in human readers.

Phonological Coherence Model

Another example of connectionist models is the Phonological Coherence Model. This model was first presented in 1990 by Van Orden and his colleagues (see also, Farrar & Van Orden, 2001; Van Orden et al., 1990; Van Orden, Bosman, Goldinger, & Farrar, 1997) and is shown in Figure 4.

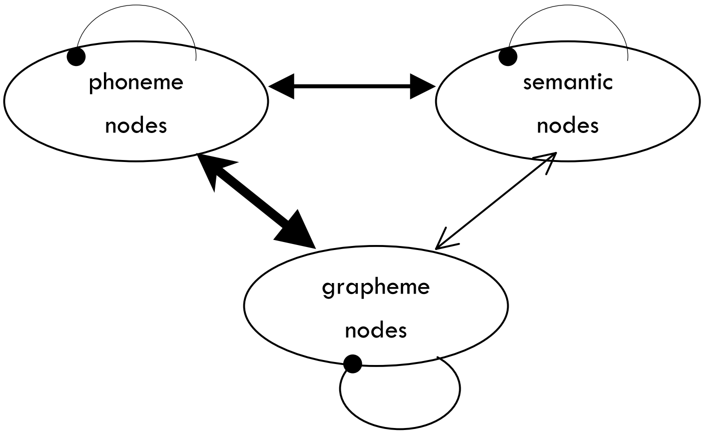

Figure 4. Macrolevel of the Phonological Coherence Model adapted from Bosman & Van Orden (1997)

They assume that reading a word is the result of parallel activation of semantic nodes, orthographic nodes, and phonological nodes in a fully recurrent network. These nodes are functional units or subsymbols, which are only ‘in function’ when they are active. Like the connectionist model of Seidenberg and colleagues, words are considered as distributed patterns of activation across the nodes. Nodes are bi-directionally linked to each other, with different strengths. The weights of these connections are the result of a covariant learning process. This means that patterns that co-occur frequently (i.e., orthographic subsymbols and phonological subsymbols), will have greater strengths than combinations of letters and sounds that are scarce. Thus, patterns of correspondences are detected by the system and are used to adjust the weights. Whereas the network model of Seidenberg and McClelland included a supervised learning rule, the Phonological Coherence Model makes use of an unsupervised learning paradigm. In unsupervised learning, there is no external instruction that is responsible for the correct mappings between nodes. If the output corresponds with the input, weights are strengthened and if the pattern is uncorrelated, the strengths will decrease. The authors stress the fundamental role of phonology because phonology is always activated when reading a word. Numerous studies have provided evidence for this assumption (e.g., Bosman, van Leerdam, & de Gelder, 2000; Van Orden, 1987; Van Orden et al., 1990). Importantly, according to the model, semantics is also activated in the word-identification process and can support the correspondence between grapheme nodes and phoneme nodes. This may be useful when the translations from graphemes to phonemes are slow or error-prone, which is the case in poor readers and in reading inconsistent irregular words. This model was advocated as a theoretical model for word identification in the present thesis, because of the recognition of the role of semantics in word identification.

Intervention studies

Although many intervention studies have been carried out in the past, there is yet no general agreement about the best way to instruct and remediate reading skills. Because intervention studies differ among several dimensions, structural comparisons between studies are complicated. Next, a few decisions related to the design and construction of an intervention study will be considered. There are of course a number of other factors that have to be taken into account (e.g., small-group instruction vs. individual instruction and the intensity of the training). However, I choose only to report on those variables that are directly related to the choices that have been made in the intervention program, as described in Chapter 3.

First, the sample of participants that is included in an intervention study influences the outcome of the study. Ehri et al. (2001) carried out a meta-analysis on the effects of phonemic-awareness instruction for learning to read. They found that although phonemic-awareness instruction yielded positive effects for at-risk readers, disabled readers and normally-progressing readers, the effects were smaller for disabled readers. Swanson (1999) also carried out a meta-analysis of intervention outcomes and contrasted student characteristics on several variables (e.g., reading cutoff-score criteria, and IQ-discrepancy). It was demonstrated that studies meeting cutoff-score criteria (IQ > 85 and reading < 25th percentile) yielded higher effect sizes for word-recognition than studies that did not meet these cutoff-score criteria. These are only a few examples to illustrate that student characteristics influence treatment outcomes. There are of course several other sample characteristics (e.g., the age of the participants, levels of socio-economic status, SLI-diagnosis, ethnic minority) that presumably affect intervention outcomes.

A second dimension on which intervention studies differ is whether the intervention is carried out by a (remedial) teacher or performed on a computer. As computers become more and more available in classrooms, a lot of computer-assisted intervention programs have been developed and tested. Meta-analyses that have been carried out on the effects of technical applications are useful in the evaluation of computer-assisted instruction (CAI). Blok, Oostdam, Otter, and Overmaat (2002) reviewed 42 publications which resulted in 75 experimental studies from 1990 onwards. They concluded that CAI is effective in beginning reading instruction, although the overall effect size was quite small. Importantly, the effect size was higher when the language of instruction was English than in studies with non-English speaking countries. MacArthur, Ferretti, Okolo, and Cavalier (2001) also reviewed several studies concerning CAI and concluded that computer applications are effective tools to increase decoding skills. Finally, Hall, Hughes, and Filbert (2000) reviewed 17 studies on CAI and concluded that computer applications might be successful tools for children with learning disabilities who need additional instruction in reading. Although these studies have demonstrated the effectiveness of CAI, the researchers underline that a lot of studies lack appropriate methodological designs. More empirical research is needed for a systematic investigation of factors that increase success of a computer program. Bishop and Santoro (2006) offer several criteria that may be useful in the selection of software for the early grades. In short, computer programs are effective tools for the improvement of decoding skills, if they are used in addition to instruction by a human teacher and provide systematic instruction with effective correction procedures.

Third, intervention studies differ in instructional contents. What skills are taught in an intervention program aimed at increasing decoding skills? Since phonological awareness has been demonstrated to play a major (causal) role in reading difficulties, many programs included training of phonological or phonemic awareness (see e.g., Gonzalez, Espinel, & Rosquete, 2002, McGuinness, McGuinness, & McGuinness, 1996; see Bus & van IJzendoorn, 1999, and Ehri et al., 2001, for a meta-analysis). Phonological/phonemic awareness training aims at ‘promoting children’s awareness of the phonological structure of spoken language, especially of phonemes’ (Scarborough & Brady, 2002). Examples of activities that promote phonological awareness are rhyming, blending, segmentation, and categorization. A number of studies have demonstrated that training phonological-awareness skills improves reading skills of young children. In particular, a phonological-awareness training combined with the training of letter-sound correspondences, increases reading skills considerably (Bus & van IJzendoorn, 1999; Ehri et al., 2001). Positive effects have been demonstrated for normally developing children, at-risk children, as well as children with dyslexia. The focus on phonological skills is also discernable in psycholinguistic-training programs for children with dyslexia. In a psycholinguistic program, the sound structure of words is explicitly taught and children are instructed to use algorithms concerning grapheme-phoneme correspondences. Results have demonstrated positive effects on reading skills (e.g., Gerretsen, Vaessen, & Ekkebus, 2003; Tijms, Hoeks, Paulussen-Hoogeboom, & Smolenaars, 2003; Tijms & Hoeks, 2005).

Other intervention studies have focused on the orthographic structure of words rather than on phonology. To reach a satisfying level of word-specific orthographic knowledge, repeated-reading programs have been constructed (e.g., Meyer & Felton, 1999; Rashotte & Torgesen, 1985; Therrien, 2004; see for an overview, Van den Bosch & Van Bon, 1994). In general, these studies have demonstrated that repeated exposures to words or short meaningful passages increases reading speed for those words or passages. Positive effects of repeated reading can be explained both in the Dual-Route model of reading and in a connectionist framework of reading. According to the Dual-Route Model, words can be identified either via grapheme-phoneme correspondences or direct, via the visual route, which requires whole-word recognition. Repeated reading is assumed to increase direct recognition of words, thus reading via the direct route. In a connectionist model, the identification of a word entails the parallel activation of graphemic nodes, phonemic nodes, and semantic nodes. If specific words are presented frequently, the weights of the connections between the nodes will increase.

Few intervention studies have focused on the semantic attributes of words (e.g., Berends & Reitsma, 2006; Norbury & Chiat, 2000). If reading assumes the activation of orthographic, phonological, as well as semantic information, semantic activation might contribute to correct word identification. Semantic interventions aim at improvement of explicit connections between orthography, phonology, and meaning. Activities in a semantically-based training require focusing on semantic attributes of words. This can be realized by manipulating the type of response that is requested in a training program. This issue will be considered more thoroughly in Chapter 3, which focuses on the effects of a semantically-oriented reading program.

The Dutch Spelling System

Written Dutch uses all 26 letters of the Roman alphabet (i.e., A, B, C, D, E, F,G, H, I, J, K,

L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z) with the additional digraph IJ, which has a special status. The Roman alphabet is a fully phonetic system, that is, both consonants and vowels are represented by letters or letter clusters. Regarding the description of Dutch orthography, its phonemes serve as the starting point. A phoneme is the smallest phonetic unit in a language capable of conveying a distinction in meaning. Alphabetic languages attempt to define a consistent relationship between the phonemes of the language and a limited set of graphic signs, that is, graphemes. A grapheme is the orthographic counterpart of the phoneme. The graphemes constitute all the letters of the alphabet plus all letter combinations that represent a phoneme. The Dutch language largely consists of two different orthographies, one for native words (85%) and one for non-native words (15%) (Bosman & Mekking, 2006). Because this thesis primarily focuses on word-identification skills of beginning readers, only the spelling of native words will be discussed. Most native-Dutch words contain highly consistent grapheme-to-phoneme relations, which makes reading native words fairly straightforward, provided one has been familiarized with the set of grapheme-phoneme relations and one does not suffer from dyslexia. In the Netherlands, most words in the reading-instruction books used in first grade are native and highly consistent.

The set of native-Dutch words has 35 native phonemes and 39 native graphemes. The native phonemes are further divided in 19 consonants and 16 vowels. The consonants comprise 5 plosives [p b t d k], 7 fricatives [f v s z χ ɣ h], 3 nasals [n m η], 2 liquids [l r], and 2 glides [w j]. The vowels comprise 5 tense or short vowels [I ɛ ɑ ɔ Y], 7 lax or long vowels [i y e a o ø u], 3 diphthongs [ɛi œy ɑu], and the schwa [ə]. The native graphemes comprise 23 monographs (a, e, i, o, u, ij, b, d, f, g, h, j, k, l, m, n, p, r, s, t, v, w, z), 14 digraphs (aa, ee, oo, uu, ie, oe, ei, ui, au, ou, eu, ng, uw, and ch), and 2 trigraphs (ouw, auw).

In Dutch, phoneme-grapheme relations are less consistent than grapheme-phoneme relations. Fewer possibilities exist for the pronunciation of a Dutch letter than for the spelling of a Dutch phoneme. This indicates that spelling in Dutch, at least of native words, is more difficult than reading in Dutch, a characteristic shared by most other languages using the roman alphabet (Bosman & Van Orden, 1997). For example, when presented with the spoken word [ɣɛit] (goat), one has to decide which of the two spelling alternatives GEIT or GIJT is the correct one, because the phoneme [ɛi] can be written in two ways: EI as in WEI (meadow) or IJ as in WIJ (we). Reading the words GEIT or WIJ, however, unequivocally leads to the correct pronunciation, provided that one is familiar with the grapheme-phoneme conversion rules. To conclude, since Dutch grapheme-phoneme relations are fairly consistent, reading accuracy is not the major problem in most of the children in the Netherlands.

However, a few Dutch spelling rules are responsible for slight deviations of the consistent grapheme-phoneme mappings and may complicate word decoding, especially in beginning readers. One example is the vowel-reduction rule. Vowel reduction (i.e., degemination) occurs in words with lax vowels. Generally, the spelling rule for words with lax vowels prescribes a double letter (i.e., a geminate), for example, the word RAAM [ram], (window) has a lax vowel and is thus spelled with two A's. The plural version of RAAM is RAMEN [ramən], (windows), still contains a lax vowel, but its vowel is reduced to a single A. This is the result of a spelling rule that states that vowels in open syllables reduce to one. It has been demonstrated (e.g., Rutjens, 2000) that a substantial number of reading errors in young children are made because of erroneous application of rules, for example,the misreading of the single vowel in an open syllable. Children often pronounced the single A of RAMEN as [ɑ], whereas it should read as [a]. Thus, misapplication of spelling rules appears to contribute mostly to the variety of reading errors, apart from addition and deletion of single letters. If, however, the absolute number of errors is taken into consideration, it has to be concluded that reading in Dutch is actually not a real problem for the majority of children.

Outline of the present thesis

The present study was designed to obtain insight in semantic effects in word identification of Dutch children. In Chapter 2, the issue of predicting reading difficulties is addressed. A longitudinal prediction study was carried out, in which risk factors, cognitive skills, and teachers' perceptions in kindergarten were related to reading performance at the beginning of Grade 1. More specifically, the discriminatory power of these predictors was studied to judge whether these factors could identify reading disabilities at the individual level. In other words, I wanted to know whether these variables could predict a child's likelihood of developing reading difficulties in Grade 1. Although specific variables (e.g., phonological awareness and letter knowledge) are strongly related to subsequent reading performance at group level, a number of studies have demonstrated too many false positives (e.g., Catts, Fey, Zhang, & Tomblin, 2001; Flynn & Rahbar, 1998; O'Connor & Jenkins, 1999; Taylor et al., 2000), and false negatives (e.g., Coleman & Dover, 1993; Hammill et al., 2002; Mantzicopoulus & Morrison, 1994). False positives refer to the number of students who are predicted to have reading difficulties, but who turn out to be normal readers. False negatives refer to the number of students who are predicted to be normal readers but who turn out to have reading difficulties. In addition to this theoretical perspective, the study also served an empirical goal. For intervention purposes in Grade 1, it was necessary to know at an early stage which children were at risk for developing reading difficulties.

Chapter 3 describes the results of two intervention studies, in which effects of a semantically-oriented training program and a phonologically-oriented training program are discussed. Poor beginning readers (mean age 6.5 years) from regular primary schools were assigned to either an experimental group or a control group. Children in the experimental group received either a semantically-oriented reading training or a phonologically-oriented training. In the semantically-oriented program, semantic activation was elicited by the inclusion of pictures or was stimulated by the type of response requested. By focusing on the activated semantic attributes of a word, the intervention aimed at improving decoding skills in poor readers. That is, poor readers are slow or inaccurate in connecting graphemes to phonemes. However, because of the assumed interaction between semantics, orthography, and phonology in visual word perception, it was hypothesized that semantic activation would support the inefficient mappings of the graphemes to the correct phonemes in children with reading difficulties. Importantly, the program aimed at automatic compensation of semantics, rather than intentional use of contextual cues.

Chapter 4 addresses the relationship between semantic skills and word decoding. Few researchers have examined the relationship between semantic-categorization skills and reading performance in children and demonstrated a positive relationship between categorization skills and word identification in primary grades (e.g., Ben-Dror et al., 1995; Howell & Manis, 1986; Vellutino et al., 1995). Two experiments were carried out to examine whether poor readers and good readers differ in their semantic skills. In the first experiment, 99 first graders performed both a semantic-categorization task and a word-association task and results were compared with word-decoding skills. In Experiment 2, 141 children from Grades 1 to 6 participated in two types of categorization tests and the relationship between semantic skills and word decoding was studied from a developmental perspective.

Chapter 5 reports the effect of a semantic variable, that is, imageability in visual word perception in primary grades. The question was whether word-identification skills of Dutch students would be affected by imageability ratings of words. Word-reading skills were assessed by lexical decision (Experiment 1) and naming (Experiment 2). In both experiments, a speeded test as well as a non-speeded test were administered. The majority of researchers investigated imageability effects in printed-word perception in skilled adult readers and demonstrated an imageability advantage in low-frequency (exception) words (e.g., Cortese, Simpson, & Woolsey, 1997; De Groot, 1989; Strain et al., 1995). The question was whether semantic characteristics of words, more specifically, imageability ratings, would influence word-identification skills of children.

Finally, Chapter 6 presents general conclusions and a discussion of the results. Theoretical and practical implications will be considered and recommendations for subsequent research of semantics in word identification will be presented.

REFERENCES

Ben-Dror, I., Bentin, S., and Frost, R. (1995). Semantic, phonologic, and morphologic skills in reading disabled and normal children: Evidence from perception and production of spoken Hebrew. *Reading Research Quarterly, 30,* 876-893.

Berends, I. E., and Reitsma, P. (2006). Addressing semantics promotes the development of reading fluency. *Applied Psycholinguistics, 27,* 247-265.

Berent, I., & Perfetti, C. A. (1995). A Rose is a REEZ: The two-cycles model of phonology assembly in reading English. *Psychological Review, 102,* 146-184.

Besner, D., Twilley, L., McCann, R. S. Seergobin, K. (1990). On the association between connectionism and data: Are a few words necessary? *Psychological Review, 97,* 432-446.

Bishop, M. J., & Santoro, L. E. (2006). Evaluating beginning reading software for at-risk learners. *Psychology in the Schools, 43, 57-70.*

Blok, H., Oostdam, R. Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research, 72,* 101-130.

Bosman, A. M. T., & van Hell, J. G (2002). Orthography, phonology, and semantics: Concerted action in word perception. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.). *Precursors of functional literacy* (pp. 165-187). Amsterdam: John Benjamins.

Bosman, A. M. T., van Leerdam, M., & de Gelder, B. (2000). The /O/ in OVER is different from the /O/ in OTTER: Phonological effects in children with and without dyslexia. *Developmental Psychology, 36,* 817-825.

Bosman, A. M. T., & Mekking, A. J. (2006). *Statistical analysis of the bidirectional inconsistency of spelling and sound in Dutch.* Manuscript in preparation.

Bosman, A. M. T., & Van Orden, G. C. (1997). Why spelling is more difficult than reading. In C. A. Perfetti, L. Rieben, & M. Fayol, (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 173-194). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bosman, A. M. T., & Van Orden, G. C. (2003). Het fonologisch coherentiemodel voor lezen en spellen [The phonologic coherence model of reading and spelling]. *Pedagogische Studiën, 80*, 391-406.

Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91,* 403-414.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten students: a research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32,* 38-50.

Coleman, J. M., & Dover, G. M. (1993). The RISK Screening Test: Using kindergarten teachers' ratings to predict future placement in resource classrooms. *Exceptional Children, 59, 468-477.*

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological-Review,* 108, 204-256.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.). *Strategies of information processing* (pp.151-216). London: Academic Press.

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review,* 4, 226-231.

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

Elbro, C., Borstrøm, I. & Petersen, D. K. (1998). Predicting dyslexia from kindergarten. The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly, 33*, 36-60.

Farrar, W. T., & Van Orden, G. C. (2001). Errors as multistable response options. *Nonlinear Dynamics, Psychology, and Life Sciences, 5*, 223-265.

Flynn, J. M., & Rahbar, M. H. (1998). Improving teacher prediction of students at risk for reading failure. *Psychology in the Schools, 35,* 163-172.

Frost, R. (1995). Phonological computation and missing vowels: Mapping lexical involvement in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 398-408.

Gerretsen, P., Vaessen, A., & Ekkebus,M. (2003). Het effect van een psycholinguïstische behandeling bij kinderen en volwassenen met dyslexie. [The effect of a psycholinguistic treatment in children and adults with dyslexia]. *Tijdschrift voor Remedial Teaching, 2,* 4-11.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 674-691.

Gonzalez, M. R. O., Espinel, A. I. G., & Rosquete, R. G. (2002). Remedial interventions for children with reading disabilities: Speech perception – An effective component in phonological training? *Journal of Learning Disabilities, 35,* 334-342.

de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 824-845.

Hall, T. E., Hughes, C. A., and Filbert, M. (2000). Computer assisted instruction in reading for students with learning disabilities: A research synthesis. *Education and Treatment of Children, 23,* 173-193.

Hammill, D. D., Mather, N., Allen, E. A., & Roberts, R. (2002). Using semantics, grammar, phonology, and rapid naming tasks to predict word identification. *Journal of Learning Disabilities, 35,* 121-136.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111,* 662-720.

Howell, M.J., and Manis, F.R. (1986). Developmental and reader ability differences in semantic processing efficiency. *Journal of Educational Psychology, 78,* 124-129.

Jackson, N. E., & Coltheart, M. (2001). *Routes to reading success and failure: Toward an Integrated Cognitive Psychology of Atypical Reading* (pp. 39-70). NY: Taylor & Francis.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120,* 358-394. MacArthur, C. A., Ferretti, R. P., Okolo, C. M., & Cavalier, A. R. (2001). Technology applications for students with literacy problems: A critical review. *The Elementary School Journal, 101,* 273-301.

Mantzicopoulos, P. Y., & Morrison, D. (1994). Early prediction of reading achievement. Exploring the relationship of cognitive and noncognitive measures to inaccurate classifications of at-risk status. *Remedial and Special Education, 15,* 244-251.

McClelland, J. L., Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88,* 375-407.

McGuinness, C., McGuinness, D., & McGuinness, G. (1996). Phonographix™: A new method for remediating reading difficulties. *Annals of Dyslexia, 46*, 73-96.

Meyer, M. S., Felton, R. H. (1999). Repeated reading to enhance fluency: Old approaches and new directions. *Annals of Dyslexia, 49,* 283-306.

Norbury, C. F., & Chiat, S. (2000). Semantic intervention to support word recognition: a single-case study. *Child Language Teaching and Therapy, 16,* 141-163.

O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading, 3,* 159-197.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S. & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review, 103,* 56-115.

Rashotte, C. A., & Torgesen, J. K. (1985). Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly, 20,* 180-188.

Rosson, M. B. (1983). From SOFA to LOUCH: Lexical contributions to pseudoword pronunciation. *Memory & Cognition, 11,* 152-160.

Rutjens, E. (2000). *Onderzoek naar het pseudowoord-deficiet en het type leesfouten bij dyslectische kinderen.* [A study of the pseudoword deficit and the type of reading errors in children with developmental dyslexia. Unpublished masters' thesis, University of Nijmegen, the Netherlands.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523-568.

Scarborough, H. S., & Brady, S. A. (2002). Toward a common terminology for talking about speech and reading: a glossary of the "Phon"words and some related terms. *Journal of Literacy Research,* 329-336.

Strain, E, Patterson, K, & Seidenberg, M. S. (1995). Semantic effects in single-word naming. Journal of experimental psychology: Learning, Memory, and Cognition, 21(5), 1140-1154.

Swanson, H. Lee. (1999). Reading research for students with LD: A meta-analysis of intervention outcomes. *Journal of Learning Disabilities, 32,* 504-532.

Taylor, H. G., Anselmo, M., Foreman, A. L., Schatschneider, C., & Angelopoulos, J. (2000). Utility of kindergarten teacher judgments in identifying early learning problems. *Journal of Learning Disabilities, 33,* 200-210.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading. A Meta-Analysis. *Remedial and special Education, 25,* 252-261.

Tijms, J., & Hoeks, J. (2005). A computerized treatment of dyslexia: Benefits from treating lexico-phonological processing problems. *Dyslexia, 11,* 22-40.

Tijms, J., Hoeks, J. J. W. M., Paulussen-Hoogeboom, M. C., & Smolenaars, A. J. (2003). Long-term effects of a psycholinguistic treatment for dyslexia. *Journal of Research in Reading, 26,* 121-140.

Van den Bosch, K., & van Bon, W. H. J. (1994). Preventie en remediatie van leesproblemen: Effecten van training op woordniveau en lager [Prevention and remediation of reading difficulties: Training effects on word level and lower levels] . *Stem-, Spraak- en Taalpathologie, 3,* 137-160.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound and reading. *Memory & Cognition, 15,* 181-190.

Van Orden, G. C., Bosman, A. M. T., Goldinger, S. D., & Farrar, W. T. (1997). A recurrent network account of reading, spelling, and dyslexia. In J. W. Donahoe & V. Packard Dorsel (Eds.), *Neural network models of cognition: A biobehavioral foundation* (pp. 522-538). Amsterdam, the Netherlands: Elsevier Science Publishers.

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1269-1291.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97*, 488-522.

Vellutino, F. R., Scanlon, D.M., and Spearing, D. (1995). Semantic and phonological coding in poor and normal readers. *Journal of Experimental Child Psychology, 59, 7*6-123.

Venezky, R. L. (1984). The history of reading research. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal, (Eds). *Handbook of Reading Research* (pp. 3-38). New York : Longman.

Chapter 2: Kindergarten Risk Factors, Cognitive Factors and Teacher Judgments as Predictors of Early Reading in Dutch[1]

Abstract

This study focused on the predictive value of risk factors, cognitive factors, and teachers' judgments in a sample of 462 kindergartners for their early reading skills and reading failure at the beginning of Grade 1. With respect to risk factors, enrollment in speech–language therapy, history of dyslexia or speech–language problems in the family, and the role of gender were considered. None of these risk factors were significantly related to reading performance. Cognitive factors in this study included letter knowledge, rapid naming ability, and nonword repetition skills. Of these skills, letter knowledge seemed to have the highest correlation with reading. Kindergarten teachers' judgments, including a task assignment scale and teachers' predictions, demonstrated a significant relationship with reading. Finally, to judge whether these predictors could identify reading disabilities, the discriminatory power of all predictors was assessed and appeared to be insufficient. Implications for screening purposes are discussed.

---

[1] Accepted for publication as: Gijsel, M. A. R., Bosman, A. M. T., & Verhoeven, L. (in press). Kindergarten Risk Factors, Cognitive Factors and Teacher Judgments as Predictors of Early Reading in Dutch. *Journal of Learning Disabilities*. We are greatly indebted to all children and teachers of the schools that participated in our study and we want to express our appreciation to all students who assisted us in collecting the data. We are also very grateful to Jan van Leeuwe for his statistical support. Finally, I want to thank three anonymous reviewers for their helpful comments on an earlier version of the manuscript.

Introduction

Factors that have been demonstrated to correlate with reading performance in previous research include risk factors, cognitive factors, and teachers' judgments. If a significant relationship between a particular factor and reading performance has been established, the factor is often identified as a predictor of reading ability. Predictors are referred to as *concurrent predictors* if the relationship between the factor and reading is demonstrated at the same point in time. In contrast, in *longitudinal predictors*, the relationship is demonstrated over a certain time interval. In this case, assessments are usually made in kindergarten in order to predict reading skills in higher grades. In the current study, we expand on the existing research and focus on longitudinal predictors of reading performance in Dutch first graders.

For screening purposes, the prediction of reading disabilities (RD) is more important than the prediction of reading ability in general. A teacher wants to know which child is at risk for reading problems in order to prevent difficulties and also to take appropriate preventive measures. Therefore, it is important to know the critical predictive factors for reading failure. These factors could be useful in a screening battery for subsequent reading skills. In this study, we will investigate the importance of discriminating between the prediction of reading abilities in general and the prediction of reading difficulties in particular. In short, in addition to the relationship between reading ability and risk factors, cognitive factors, and teachers' judgments, we will explore the discriminatory power of these predictors.

Predictive Factors for Reading Performance

With respect to the prediction of reading skills, we explored the relationship between reading performance and risk factors (i.e., dyslexia, speech and language disorders in relatives, Specific Language Impairment (SLI), and gender), cognitive factors (i.e., letter knowledge, rapid naming ability, and nonword repetition), and teachers' judgments. With respect to cognitive factors, we included a letter knowledge test, because letter knowledge has been found to be the best predictor of reading performance (Blaiklock, 2004; Braams & Bosman, 2000; Catts, Fey, Zhang, & Tomblin, 2001; Gallagher, Frith, & Snowling, 2000; Hammill, 2004; Pennington & Lefly, 2001; Scarborough, 1998). In addition to letter knowledge, phonological processing skills have also been shown to correlate with subsequent reading skills. Wagner and Torgesen (1987) distinguished three types of phonological processing abilities: phonological awareness, phonological recoding in lexical access (long-term memory), and phonetic recoding to maintain information in working memory (short–term memory). Although most research has aimed at investigating phonological awareness, Scarborough (1998) pointed out that phonological awareness tests appear to be more successful in predicting future superior reading than future reading problems. Moreover, Blaiklock (2004) suggested that letter knowledge and phonological awareness show substantial overlap in the explained amount of variance in reading (for relations between letter knowledge and phonological awareness in preschoolers, see Johnston, Anderson, & Holligan, 1996). Therefore, instead of measuring phonological awareness, it was decided to assess phonological memory by means of a nonword repetition test and to test phonological naming by means of a rapid serial naming test. We decided on these tests because little is known about the relative impact of these abilities on predictions of Dutch students' reading abilities.

Risk Factors

A family history of dyslexia or speech and language disorders in relatives is often considered to negatively affect students' performance on language and reading tests. Most studies of genetic factors in reading performance have been concerned with the prevalence of dyslexia in relatives. Snowling, Gallagher, and Frith (2003) showed that 66.1% of the students with at least one family member with dyslexia experienced reading problems at the age of 8. Lewis, Freebairn, and Taylor (2000) also showed a trend for a family history of RD to predict reading failure. Moreover, a family history of reading problems significantly predicted spelling impairment. A family history of speech–language disorders as a prospective risk factor for reading has been studied less frequently. Lewis et al. (2000) showed a moderate association between a family history of RD and reading impairment at school age if this variable was assessed in a dichotomous fashion (i.e., positive vs. negative). When it is was coded as an ordinal variable (i.e., the number of nuclear family members affected) no significant effect emerged.

A second variable that possibly constitutes a risk factor for RD is speech and language characteristics. In most studies, poor language or speech characteristics have been diagnosed by means of test assessment (see, e.g., de Jong, & van der Leij, 1999; Menyuk et al., 1991; Scarborough, 1990). Usually, the results of speech–language tests are related to reading performance. An alternative way to assess information about speech and language characteristics is to investigate the history of speech–language therapy. Weiner (1985) found a high incidence of later reading problems in preschoolers who were enrolled in language therapy.

Third, differences in gender may play a causal role in lowering test performance. Petersen (2002) demonstrated that although boys showed a significantly higher score on vocabulary and nonverbal IQ at the beginning of kindergarten, they showed lower scores on reading performance in Grade 2. Badian (1999) indicated that girls were significantly better at reading comprehension than boys. No such differences were found for listening comprehension. Blonk and Bosman (2003) found that teenage girls in the first year of secondary school performed better than teenage boys on reading and language skills and, more important, that teenage girls with dyslexia scored better than teenage boys with dyslexia. Flynn and Rahbar (1994) tested 708 students in Grades 1 and 3 on reading achievement and categorized the students as having severe RD (< 10th percentile), RD (11th to 30th percentile), or typical reading abilities. In the severe RD category, the ratio of boys and girls was 1.4 to 1 in first grade and 1.3 to 1 in third grade. In the RD category, the number of boys and girls was approximately the same in each grade.

Cognitive Factors

Letter Knowledge

In the domain of cognitive abilities, letter-name knowledge is a factor that has been studied frequently in relation to reading skills. The combined results of three meta-analyses demonstrated that together with reading itself and knowledge about writing convention, letter knowledge ($r = .52$) was the best predictor of reading (Hammill, 2004). Other studies of both English- and non–English-speaking populations have yielded similar results. De Jong and van der Leij (1999) demonstrated the high predictive value of letter knowledge in the Dutch language. In their test, they presented five letters used relatively frequently in Dutch books. On the productive letter test, both letter names and letter sounds were considered correct. The correlation between receptive and productive letter knowledge in kindergarten and word decoding at the beginning of Grade 1 was moderate ($r = .39$ and $r = .51$, respectively). At the end of Grade 2, however, the correlations between letter knowledge in kindergarten and word decoding in Grade 2 had decreased dramatically. Braams and Bosman (2000) also used a letter naming task in kindergarten to predict the reading and spelling ability of Dutch students in Grade 1. At the beginning and at the end of kindergarten, students had to name 20 letters. Letter knowledge tested at the beginning of kindergarten revealed a moderate correlation with reading performance at the middle of the curriculum in Grade 1 ($r = .44$), but the correlation between letter knowledge in kindergarten and reading declined with increasing reading experience to .36 at the end of Grade 1.

Rapid Naming.

Another factor often related to the prediction of reading performance is rapid naming (e.g., Allor, 2002; Blachman, 1984; Cornwall, 1992; Hammill, Mather, Allen, & Roberts, 2002; Kirby, Parrila, & Pfeiffer, 2003). In rapid naming tests, participants are asked to name a set of items (usually pictures, colors, letters, or digits) as quickly and accurately as possible. The strength of the relationship between rapid naming and reading performance is dependent on the age of the participants. Kirby et al. (2003) demonstrated that naming speed in higher grades had much stronger effects on reading than in kindergarten and Grade 1, with the latter grades showing significant, albeit weak effects. De Jong and van der Leij (1999) also reported small effects of rapid naming ability in kindergarten on reading skills in Grade 1. Van den Bos, Lutje Spelberg, and Eleveld (2004) tested kindergartners on rapid naming (colors and pictures) and visual matching, and they studied the predictive power of these tests for reading ability in Grade 1. They demonstrated a moderate but significant multiple correlation ($r = .52$). Note that this correlation is even higher than the one between letter knowledge in kindergarten and reading ability in Grade 1 as reported by Braams and Bosman (2000). Hammill et al. (2002) also investigated the importance of rapid naming ability in predicting word identification. In their study, the correlation between rapid naming (letters and words) and word identification was .52. In a study by Cornwall (1992), the correlation between rapid naming and the identification of individual words was .49 for letters and .19 for colors. It thus seems that the strength of the relationship between rapid naming and reading also depends on the kind of stimuli used in the test.

Nonword Repetition

Another factor that is often associated with reading ability is nonword repetition. In nonword repetition tests, participants are asked to repeat a set of nonwords. The test items generally obey the phonological structure of the language. A number of studies have demonstrated that students with RD perform relatively badly on a test for repeating nonwords (Gallagher et al., 2000; Lewis et al., 2000; Muter & Snowling, 1998; Snowling, 1981; Snowling, Goulandris, Bowlby, & Howell, 1986; Snowling et al., 2003). These findings are often explained in terms of deficits in phonological working memory, integrity of phonological representations, or phonological decoding. In the study of Lewis et al. (2000), all participants were enrolled in speech-language therapy. Reading impairment in this group of students was associated with their scores on nonword repetition in kindergarten. Bishop (2001) also concluded that in children with SLI, deficits in nonword repetition and poor literacy skills were correlated, suggesting that the same genes were responsible for both deficits. Muter and Snowling (1998) demonstrated that the strength of the correlation between nonword repetition as a longitudinal predictor and reading performance at the end of Grade 5 was significant in a typical population, that is, in children without SLI (*r*s ranging from .31 at the age of 4 to .53 at the age of 6).

Teachers' Perceptions and Predictions

Finally, we were interested in the predictive value of the perceptions and predictions of kindergarten teachers. In a study by Kenny and Chekaluk (1993), teachers' questionnaire scores and teachers' category ratings (advanced, average, or poor reading) showed good predictive value for auditory conceptualization scores and reading achievement. Flynn and Rahbar (1998) demonstrated that teachers correctly predicted 64% of poor readers and missed 36% of those students who failed in reading. In their study, the correspondence between teacher ratings and students' scores was low, except for scores in the area of letter-sound knowledge, and it appeared that a combination of test performance scores and teacher information was most effective with respect to the prediction of reading failure.

Discriminatory Power of Predictive Factors

From a scientific point of view, it is not only interesting to acquire knowledge about predictors of reading performance in general, but also to obtain an accurate prediction of the likelihood of developing RD. Therefore, we were interested in the discriminatory ability of risk factors (e.g., dyslexia or speech-language disorders in relatives, enrollment in speech and language therapy, gender), cognitive factors, and teachers' predictions and perceptions. In other words, we wanted to know whether these variables could predict a child's likelihood of developing RD in Grade 1. A moderate or high correlation demonstrating the predictive validity of a factor does not imply that this factor will discriminate well between students who will develop RD and those who will not.

In the current study, children who performed below the cutoff score—that is, below the 25th percentile—in reading after 2 months of formal reading instruction were considered children with RD. Both practical and theoretical evidence demonstrate the validity of early assessment and stability of reading scores for Dutch students throughout all grades. Verhoeven and van Leeuwe (2003) categorized 2,873 students into five groups at different reading levels, based on their reading performance after 3 months of formal reading instruction in Grade 1. These students showed quasi-stable mean levels of reading skills through Grade 6 for one-syllable CVC words (C = consonant; V = vowel). Bast and Reitsma (1998) reported a similar outcome in 280 Dutch students—that is, children who were diagnosed as poor readers after 3 months of reading instruction remained poor readers during the first three grades.

Moreover, the reading-acquisition rate in consistent alphabetic orthographies such as Finnish (see, e.g., Aro, 2004; Lyytinen et al., 2004) is different from the rates reported for English—a language with a highly inconsistent orthography. In Finnish, one of the most consistently spelled alphabetic languages, the majority of children have acquired accurate decoding skills after the first semester in Grade 1. Thus, RD can be identified much earlier than in English. The same point can be made for Dutch; the Dutch orthography is also relatively consistent in its spelling-sound relationships. After a very short time (within 4 months, before Christmas), children in first grade have learned all the prototypical grapheme-phoneme correspondences. Thus, after a relatively short time, children should be fully aware of the alphabet principle. Moreover, Dutch reading curricula strongly stimulate the awareness of the alphabet principle, because they *all* adhere to the phonics principle.

Moreover, Wentink and Verhoeven (2001) implemented a protocol for the early assessment and intervention of RD, which is now used widely in Dutch primary schools. They recommended assessing reading skills after 6 weeks of formal reading instruction.

Risk Factors

With respect to risk factors, the number of children with speech-language disorders or affected relatives (i.e., with dyslexia or speech-language disorders) or the number of boys in an RD sample may indicate the discriminatory power of these variables to some extent. If a significant proportion of these children at risk is represented in the lower tail of the reading distribution and a very small proportion is represented in the highest tail, the particular risk factor might discriminate quite well between typical children and children with RD. As mentioned earlier, Snowling et al. (2003) showed that 66.1% of the students with at least one family member with dyslexia experienced reading problems at the age of 8. Scarborough (1989, 1991) also demonstrated that family incidence of RD was an accurate predictor of reading ability; outcomes were correctly predicted for 72.6% to 80.6% of the participants, depending on the way in which family incidence was identified (Scarborough, 1989).

Cognitive Factors

Letter knowledge seemed to be a useful factor in predicting reading ability. But the question is, will letter knowledge discriminate between poor readers and typical readers? To answer this question, the number of valid positives and valid negatives versus false positives and false negatives can be calculated. *Valid positive rate* (hits) refers to the number of students who were predicted to have RD who did turn out to be poor readers. *False positive rate* (false alarm) refers to the number of students who were predicted to have RD but who turned out to be typical readers. *Valid negative rate* (correct rejection) refers to the group of students in which RD were not predicted and not observed. Finally, *false negative rate* (misses) refers to the number of students who were predicted to be typical readers but who turned out to have RD. In this way, the correctness of classification (poor vs. typical readers) based on performance on the predictor tests can be evaluated. Muter and Snowling (1998) considered this issue and selected a group of poor readers (reading accuracy scores below 25th percentile) and typical readers (reading scores above the 75th percentile) in their sample.

They used rhyme detection, phoneme deletion, nonword repetition, and letter-name knowledge as predictors. Scores were obtained at the ages of 4, 5, and 6 years. These four variables did not classify the students well at all test moments. At the age of 4, no significant predictor set was found. When test scores were obtained at the age of 5 or 6, 80% of the students were classified correctly when rhyme detection was eliminated. Analyses revealed that phoneme deletion and nonword repetition were the best predictors. Letter knowledge was a significant predictor as well, but its discriminatory power was relatively low. These results should be interpreted carefully—first, because their sample size was quite small (only 20 children), and second, because high correlations do not imply high discriminatory power, and the latter is desirable for a screening battery.

Hammill et al. (2002) also assessed the discriminatory power of tests. For this purpose, participants were divided into two groups (scoring above or below the 25th percentile) based on performance variables. Levels of agreement (i.e., the score on a test or subgroup of tests correctly identifies the level of word identification) were calculated. Results indicated 26 out of 56 false positives (46%). The percentage of false negatives was 15%; none of the test variables in the study (semantics, grammar, phonology, rapid naming, and rapid marking) were effective predictors of poor reading. With respect to nonword repetition, Muter and Snowling (1998) showed that scores on nonword repetition and phoneme deletion obtained at the age of 5 or 6 significantly discriminated between good and poor readers.

Teachers' Perceptions and Predictions

Satz and Fletcher (1988) reviewed a set of major studies regarding teachers' predictions. A comparison between kindergarten teachers' predictions and test results in Grades 1 and 2 revealed that overall hit rates were almost identical for test results and teacher predictions. Kenny and Chekaluk (1993) studied the utility of teachers' perceptions and test outcomes from kindergarten through Grade 2. Teachers completed a questionnaire and had to categorize the students into advanced, average, and poor readers. In teachers' classifications, false positive rates (ranging from .30 in kindergarten to .13 in Grade 2) were higher than false negative rates (.23, and .06, respectively). This means that the number of students who were predicted to have RD but who turned out to be typical readers was higher than the number of students who were predicted to be typical readers but who turned out to have RD.

Flynn and Rahbar (1998) presented teachers with a rating scale with behavioral descriptions reflecting 10 kindergarten skills and assessed their predictive value for reading skills in third grade. Teachers correctly classified 64% of the poor readers, in contrast with a screening test, which yielded 80% valid positives. A combination of teachers' ratings and the screening test resulted in 88% valid positives. However, the number of false positives increased from .23 (teachers' ratings alone) to .39, which indicated an overidentification of students at risk.

In addition to level of agreement, the sensitivity, specificity, and positive predictive value can be calculated to explore the practical value of a factor. The *sensitivity* index reflects the ability of a test to correctly identify individuals who have the disorder. The *specificity* index reflects the ability of a test to correctly identify individuals who do *not* have the disorder. The *positive predictive value* reflects the proportion of valid positives among all individuals whom the screening measure identifies as at risk. Petersen (2002) demonstrated a better sensitivity of predictors for advanced reading than for poor word reading.

To sum up, in the present study, we seek answers to the following two questions:

1. What is the relationship between risk factors, cognitive factors, and teachers' predictions and perceptions on the one hand and reading performance on the other hand?
2. How well do these factors identify students who will develop RD?

Method

The data collection in this study comprised a kindergarten test battery, two questionnaires designed for the parents and teachers of the kindergartners, and a Grade 1 test to assess reading skills.

Participants

Of the 462 students participating in this study, 241 were boys (52.2%) and 221 were girls (47.8%) from 20 general education primary schools in the Netherlands. All students were first tested when attending kindergarten, and their mean age was 70.8 months (*SD* = 4.4). In Grade 1, when they were tested a second time, the mean age was 79.1 months (*SD* = 4.4).

Materials

Questionnaires

Parents' Questionnaire.

This questionnaire consisted of four questions concerning the native language of the child, the child's enrollment in speech or language therapy, and the presence or absence of dyslexia or speech and language difficulties in relatives. These questions had to be answered by marking the option *yes* or *no*. If the answer was yes, further information was required (e.g., which of the first-degree relatives had dyslexia or speech-language problems, or what the child's native language was if not Dutch). The parents of all kindergartners received a questionnaire.

Teacher Questionnaire.

This questionnaire consisted of two parts. In Part 1, teachers were asked to write down the names of students whom they believed would develop reading or spelling difficulties in Grade 1. With respect to the analyses, this variable was dichotomous (*yes* vs. *no*). Part 2 was a subscale of the *Task Assignment Scale* (*Werkhoudingslijst*) of van Doorn (1996) and consisted of 10 items concerning concentration, motivation, and attitude of the child. The teacher was asked to make judgments for each of the students in their classroom about the frequency of the described behavior on a 5-point scale, ranging from *never* to *always*. Examples are "the child tries to solve problems by himself or herself," "the child needs encouragement during work," and " the child starts working immediately after instruction." With respect to the analyses, scores on task assignment were assigned to the answers and categorized either in the lowest quartile (scores below the 25th percentile) or in the highest quartile (scores above the 75th percentile).

Kindergarten Test Battery

Nonword Repetition Test.

The *Nonword Repetition Test* (*Nonwoord Repetitietaak*; Irausquin, 1999) required the repetition of 22 pronounceable nonwords. To control for possible articulation errors or hearing problems of the child, the test was preceded by 15 real words. Thus, errors in nonword repetition (e.g., substitutions due to certain articulation errors) that already occurred in real word repetition were counted as correct. In case of hearing disorders, test assessment was stopped. Performance on real words was not included in the test score; only repetition of the nonwords determined performance. The length of the items increased, ranging from one to four syllables. Nonword examples are *kloda* and *bledistot*. Three practice items preceded the nonwords. The experimenter pronounced each word and nonword with a piece of paper in front of the mouth in order to prevent the use of visual information. The child was asked to repeat the words and nonwords after the experimenter. With respect to the analyses, the test score was the number of nonwords correctly repeated. The maximum score on the test was 22.

Rapid Naming Picture Test.

The *Rapid Naming Picture Test* (*Benoemtaak Plaatjes*; van den Bos, 2004) consisted of five different pictures (tree, duck, chair, scissors, and bicycle) depicted in five columns of 10 pictures each, yielding a total number of 50 pictures to be named. Each picture was presented 10 times at random positions. The child was asked to name all the pictures as fast as possible and as correctly as possible in a vertical direction (i.e., column by column). The test was preceded by a short training, in which the child had to name the pictures in the final column. The time to complete the task and the number of errors were registered. With respect to the analyses, the mean time to name one picture (correctly or incorrectly) was computed. The number of errors was not included as a variable because few errors were made.

Rapid Naming Colors Test.

The *Rapid Naming Colors Test* (*Benoemtaak Kleuren*; van den Bos, 2004) consisted of five columns of 10 colored squares (black, yellow, red, green, and blue). The test procedure and the scoring were exactly the same as with the *Rapid Naming Picture Test*. Note that we did not use alphanumeric stimuli because the students had not received formal reading and writing instruction yet.

Letter Knowledge Test.

The *Letter Knowledge Test* considered productive letter knowledge as well as receptive letter knowledge. First, productive letter knowledge was tested. This test consisted of all 26 letters of the alphabet, presented on a card in six rows of 4 letters and one row of 2 letters. The letters were typed in Helvetica regular font, size 24. If the capital letter had a different shape than the lowercase letter (which is the case for most letters), the capital letter was presented just below the lowercase letter. Thus, all of the letters had two presentation forms, except the letter 'a', which had three different representations (a, A, and *a)*. Three different orders of letters were created. The child was presented one row of letters at a time; a sheet of paper covered the other letters. The child had to name the letters in a horizontal direction. Responses were considered correct if the child named either the lowercase letter or the capital letter correctly; both letter names and letter sounds were considered correct. In case of doubt about the production of a letter, the child was asked to name a word with that letter in the initial position. Uncertainty only emerged in the distinction between voiced sounds and their voiceless counterparts like *s/z* and *f/v*. When the child did not know a certain letter, he or she was allowed to guess and continue with the next letter.

Subsequently, receptive letter knowledge was tested. The child was presented with the same card. The experimenter named a letter sound (in random order), and the child was asked to point to that letter on the card. Guessing was permitted. With respect to the analyses, two variables were computed, namely, the score on the productive and the score on the receptive letter knowledge test. The test score was the total number of correct responses. The maximum score on each test was 26.

Grade 1 Reading Test

The *Word Reading Test* (*Toets Woorden Lezen;* Wentink & Verhoeven, 2001) consisted of three parts, each of them including10 (C)VC words. The first part consisted of well-known words that had been taught in the classroom. The second part consisted of new words that differed by one letter from words that had been taught in the classroom. In this way, the items resembled the words that had already been taught. The third part consisted of new words that differed by two letters from the words taught in the classroom. In this condition, there was hardly any resemblance to the words that had already been taught. The child was asked to read aloud the words as fast and as accurately as possible. For each part, the test was stopped when the child failed on three consecutive items. With respect to the analyses, the number of items read correctly and the time needed to perform each part of the test were converted into the number of items read correctly in one minute. The first part of the test (decoding well-known words) was not considered in the analyses. Therefore, the results of the second and third part are henceforth referred to as Word Reading 1 and 2, representing the ability to decode new words.

Procedure

Students were tested in kindergarten and at the beginning of Grade 1. In kindergarten, no structural (reading) program was used and, in most schools, there was no structural remediation in kindergarten concerning reading or reading-related skills. In Grade 1, the majority of the children were instructed with *Veilig Leren Lezen* (*Learning to Read Safely*), the most widely used reading program in Dutch schools (Mommers, Verhoeven, & van der Linden, 1979, 1994). The emphasis in this method is on the structure of the orthographic system and the relationship between letters and sounds (i.e., phonics). Initially, only consistently spelled words are used. After 4 months of instruction, the children are familiar with the main grapheme-phoneme correspondence rules. It is a fairly rigid, preprogrammed curriculum, which imposes a strict day-by-day and week-by-week progression.

In kindergarten, the students were tested halfway through the curriculum (January-March, 2002; henceforth Time 1; see Note 1). This took about 20 min for each child. The order of presentation of the tests varied among students, such that each test was administered equally often first, second, third, and so forth. At the end of kindergarten (May, 2002), the parents' questionnaires were provided to the teachers with the request to present them to the parents. Parents were asked to complete the list and return it in 2 weeks. Furthermore, the teachers' questionnaires were provided. In Grade 1, all students were tested 8 weeks after formal reading instruction had started (Fall 2002; henceforth Time 2). All students completed the *Word Reading Test*. In Grade 1, additional instruction or intervention was implemented at the earliest after the first test assessment—that is, after 2 months of formal reading instruction. This coincided (not accidentally) with the time that RD were diagnosed.

Analysis

To evaluate the predictive value of cognitive factors for reading performance, a structural analysis was performed. Furthermore, a multiple regression was carried out to determine the best combination of variables to predict reading. To investigate the relationship between risk factors and teachers' judgments, *t* tests were performed on scores on the *Word Reading Test* in Grade 1. To investigate the discriminatory power of all variables, percentages of valid and false positive and negative outcomes were calculated. Moreover, the sensitivity and specificity indexes were computed. For these analyses, the outcome had to be labeled dichotomously (i.e., developing RD or not). We defined students with RD as students with reading scores below the 25th percentile, and good readers as students demonstrating reading scores above the 75th percentile. The reading score was the mean number of items read correctly in one minute on *Word Reading Test* 1 and 2. We chose the 25th percentile to represent the group of students that performed below standard, because this criterion is used in standardized Dutch. Finally, we performed a discriminant function analysis in order to establish which combination of variables discriminated best between poor and typical readers.

Results

Sample Characteristics

The majority of students (95%) had Dutch as their native language; 22 students (5%) were originally from other countries (both in and outside Europe) and had Dutch as their second language or were bilingually educated. The number of reported speech and language problems was 71 (15.4%), whereas 6.3% of the parents of all participants had coped with speech or language problems. Finally, 6.9% of the parents reported a history of dyslexia in relatives—that is, a first- or second-degree family member with dyslexia. The lack of response on the questionnaires was very low: Of the parents' questionnaires, 400 copies (86.6%) were returned, and of the teachers' questionnaires, 439 copies (95%) were returned.

Statistics

Because 22 students in our sample were native speakers of a language other than Dutch, we first established whether the performance of these students differed from that of their peers whose mother tongue was Dutch. A one-way ANOVA on the mean scores of all tests in kindergarten and on the mean scores on the *Word Reading Test* in Grade 1 indicated that students whose native tongue was not Dutch did not differ significantly from students with Dutch as native language (all *p*s > .05). Therefore, it was decided to include in subsequent analyses all students who attended kindergarten and participated in the *Word Reading Test* at the beginning of Grade 1. Students who had to repeat kindergarten or who repeated Grade 1 were excluded. Despite this exclusion criterion, a few missing values were inevitable. To resolve this issue, we made use of the “nearest neighborhood” method to estimate the missing scores.

Predictive Factors for Reading Performance

Table 1 shows the mean scores and standard deviations on all tests in kindergarten and Grade 1. Scores were computed for all students together and for poor readers and typical readers separately. Table 2 shows the characteristics for the poor reading group and the typical reading group. To explore the predictive value of the reported tests, a structural analysis was performed. In this analysis, all variables were included. Figure 1 depicts all variables and standardized regression weights. The fit of the model was good, $\chi^2(9, 462) = 16.2$, $p = .06$, *GFI* = .99, *AGFI* = .97, *NFI* = .99, *RMSEA* = .04.

Table 1. Mean Scores, Mean Naming Times per Word and Standard Deviations in Kindergarten and Grade 1

	All students (N = 462)		*Lowest reading quartile (n = 118)*		*Highest reading quartile (n = 123)*		
	M	*SD*	*M*	*SD*	*M*	*SD*	*d*
Kindergarten							
Rapid Naming							
colors	1.6	0.5	1.9	.6	1.4	.4	.97
pictures	1.6	0.4	1.8	.4	1.4	.4	.96
Letter Knowledge							
productive	10.9	7.8	5.4	5.3	17.5	6.8	1.99
receptive	10.5	7.4	5.4	4.8	16.7	6.8	1.92
Nonword Repetition	16.8	2.8	16.2	2.9	17.2	3.0	.35
Grade 1							
Word Reading 1							
score	7.5	2.4	4.5	2.3	9.3	1.0	
time (sec)	54.7	30.7	92.2	25.6	22.4	8.6	
wpm	13.8	14.1	3.2	2.1	30.4	17.7	2.16
Word Reading 2							
score	7.3	2.9	3.4	2.5	9.1	1.4	
time (sec)	59.9	31.5	94.9	29.6	27.9	11.7	
wpm	11.9	13.0	2.1	1.7	26.0	17.6	1.91
Word Reading 1/2							
score	7.4	2.5	3.9	2.1	9.2	1.1	
time (sec)	57.3	29.4	93.6	22.6	25.1	9.4	
wpm	12.8	13.1	2.7	1.6	28.2	16.7	2.15

Note. The scores on Rapid Naming are the number of seconds to name an object (colors/pictures). The scores on Letter knowledge (maximum score = 26) and Nonword repetition (maximum score is 22) are the number of items responded correctly. Word Reading score is the number of items read correctly (maximum score = 10). Wpm is the number of words read per minute.

Table 2. Frequency Distribution of Poor-Reading (Lowest Quartile) and Superior-Reading (Highest Quartile) Groups with respect to Risk Factors and Teachers' Predictions

	All students (*N* = 462)	Lowest reading quartile (*n* = 118)	Highest reading quartile (*n* = 123)
Risk factors			
Speech/language therapy	71	30	11
Dyslexia	32	13	7
SLI	29	11	4
Boy	241	67	62
Teachers' judgments			
negative prediction	91	50	5
low task assignment [a]	127	59	15

Note. [a] Lowest quartile task-assignment group

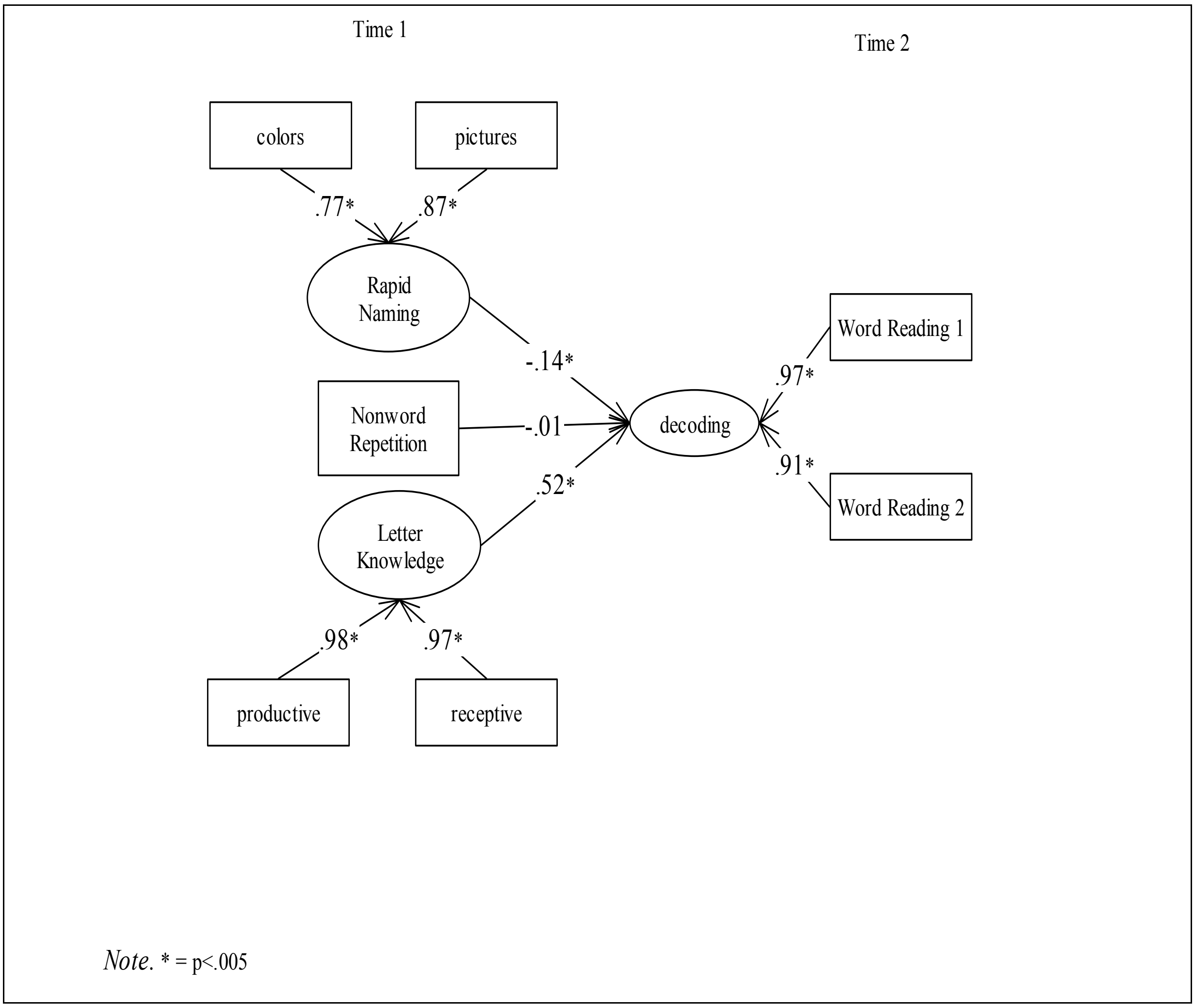

Figure 1. Structural Model of the Relationship between Kindergarten-test Scores and Decoding Skills in Grade 1 (N = 462).

Next, a stepwise multiple regression was carried out in order to find out whether a combination of factors increased the strength of the correlation. The dependent variable was the mean number of words read correctly on *Word Reading Test* 1 and 2. All risk factors, cognitive factors, and teachers' predictions and perceptions were included as independent variables. Receptive letter knowledge was excluded from the analysis because of its strong correlation ($r = .96$) with productive letter knowledge. The results showed that productive letter knowledge was the best predictor of reading performance, $r = .53$, $p < .001$. A second factor that contributed significantly to the variance was the rapid naming of colors. If this variable was added to letter knowledge, 29% of the variance was explained ($r = .54$, $p < .001$). Despite these two factors, no other predictor turned out to be significant. These results support the outcome of the structural analysis. Furthermore, to investigate the relationship between risk factors and teachers' predictions and perceptions and reading ability, *t* tests were performed. Table 3 shows that only the teachers' predictions and perceptions significantly discriminated students' scores on word reading.

Table 3. Results of t-tests for All Students concerning Risk Factors and Teachers' Predictions

		Word-Reading Test 1/2		
		M	*SD*	t-test
Risk factors				
Speech/language therapy				
	yes	10.3	14.0	$t(394) = 1.87, p = .06$
	no	13.5	12.5	
Dyslexia				
	yes	9.17	7.07	$t(366) = 1.81, p = .07$
	no	13.50	13.36	
SLI				
	yes	8.90	7.27	$t(390) = 1.73, p = .08$
	no	13.19	13.17	
Gender				
	boy	12.49	13.78	$t(460) = -.59, p = .56$
	girl	13.20	12.38	
Teachers' judgments				
Prediction				
	negative	6.3	6.7	$t(306) = 8.21, p < .001$
	positive	14.7	13.5	
Task assignment				
	below 25th percentile	8.3	10.6	$t(201) = -6.0, p < .001$
	above 75th percentile	18.9	16.7	

To further investigate the role of the teacher, scores on all tests were compared between students with positive and negative predictions (see Table 4) and between students with scores within the lowest quartile on task assignment and those with scores within the highest quartile (see Table 5). For screening purposes, it would be useful to know whether the students with negative ratings represented the lower tail of the distribution of reading scores. Analyses revealed that this was indeed the case. About half of the students who were assigned to the lowest task assignment group and who were predicted to develop RD were represented in the

25th reading percentile. The scores of the students without predicted RD were more widely distributed.

Table 4. Mean Scores and Standard Deviations of Students with Predicted Reading Difficulties (Negative Prediction) and Students without Predicted Difficulties (Positive Prediction) by their Kindergarten Teacher

		Prediction				
		Negative (*n* = 91)		Positive (*n* = 323)		
Kindergarten		*M*	*SD*	*M*	*SD*	*t-test*
Rapid Naming						
	pictures	1.7	0.5	1.6	.4	*t*(412) = -3.30, *p* = .001
	colors	1.8	0.6	1.6	.5	*t*(412) = -3.45, *p* = .001
Letter Knowledge						
	productive	5.1	5.3	12.4	7.6	*t*(206) = 10.48, *p* < .001
	receptive	5.2	4.9	11.9	7.2	*t*(213) = 10.18, *p* < .001
Nonword Repetition		15.4	3.0	17.1	2.7	*t*(412) = 5.38, *p* < .001
Grade 1						
Word-Reading 2		6.8	6.8	15.8	14.6	*t*(324) = 8.27, *p* < .001
Word Reading 3		5.7	6.8	13.7	13.4	*t*(295) = 7.68, *p* < .001
Word Reading 2/3		6.3	6.7	14.7	13.5	*t*(306) = 8.21, *p* < .001

Table 5. Mean Scores and Standard Deviations on all Tests for both Task-Assignment Groups

		Task assignment				
		Lowest quartile (*n* = 127)		Highest quartile (*n* = 121)		
		M	*SD*	*M*	*SD*	*t-test*
Kindergarten						
Rapid Naming						
	pictures	1.74	.51	1.46	.37	*t*(231) = -5.02, *p* < .001
	colors	1.85	.65	1.42	.36	*t*(198) = 6.50, *p* < .001
Letter Knowledge						
	productive	7.1	6.9	14.5	7.5	*t*(246) = -8.13, *p* < .001
	receptive	7.1	6.2	14.0	7.3	*t*(236) = -8.05, *p* < .001
Nonword Repetition		16.0	2.9	17.1	2.5	*t*(246) = -3.23, *p* = .001
Grade 1						
Word-Reading 2		8.9	11.5	19.9	17.6	*t*(205) = -5.81, *p* < .001
Word Reading 3		7.7	10.7	17.9	16.8	*t*(202) = -5.71, *p* < .001
Word Reading 2/3		8.3	10.6	18.9	16.7	*t*(201) = -6.00, *p* < .001

Discriminatory Power of Predictive Factors

To answer our second question concerning the discriminatory power of the variables, we calculated the number of false positive and valid positive outcomes and false negative and valid negative outcomes. Recall that the false positive rate reflects the number of students who were predicted to have RD but who turned out to be typical readers. The valid positive rate refers to the number of students who were predicted to have RD who indeed turned out to be poor readers. The valid negative rate refers to the group of students for whom RD were not predicted and not observed. Finally, the false negative rate refers to the number of students who were predicted to be typical readers but who turned out to have RD. These percentages were computed for all variables. In this way, we were able to evaluate different predictors of RD. Concerning the cognitive factors, difficulties were predicted if the score on the test was below the 25th percentile. Recall that RD were defined by scores below the 25th percentile on decoding new words (*Word Reading Test* 1 and 2). Next, for all predictors, we calculated the sensitivity and specificity. The sensitivity index is the number of valid positives/(valid positives + false negatives), and the specificity index is the number of valid negatives/(valid negatives + false positives). These data are presented in Table 6. The results demonstrated that teachers' predictions have more predictive power than some cognitive factors (i.e., rapid naming and nonword repetition).

Table 6. Percentages of Valid and False Positives and Negatives, Sensitivity and Specificity of Risk Factors, Cognitive Factors, and Kindergarten Predictors

	Valid positives*	False positives	Valid negatives	False negatives	Sensitivity	Specificity
Risk factors						
Speech/language therapy	42.2%	57.8%	78.2%	21.8%	.30	.86
dyslexia	40.6%	59.4%	77.7%	22.3%	.15	.93
SLI	37.9%	62.1%	75.5%	24.5%	.11	.94
Gender	27.8%	72.2%	76.9%	23.1%	.57	.49
Cognitive factors						
Rapid Naming						
pictures	36.5%	63.5%	78.1%	21.9%	.36	.79
colors	38.5%	61.5%	79.1%	20.9%	.40	.78
Letter Knowledge						
productive	49.2%	50.8%	84.1%	15.9%	.56	.80
receptive	50.8%	49.2%	83.9%	16.1%	.54	.82
Nonword Repetition	33.3%	66.7%	77.8%	22.2%	.39	.73
Teachers' judgments						
prediction	55.0%	45.0%	83.0%	17.0%	.48	.87
task assignment	46.5%	53.5%	82.0%	18.0%	.51	.79

Note. Sensitivity index is the number of valid positives/(valid positives + false negatives), and specificity index is the number of valid negatives/(valid negatives + false positives)

* the percentage of valid positives is equal to positive predictive value

Finally, we submitted all variables to a stepwise discriminant function analysis in order to establish which combination of variables discriminated best between poor readers (reading scores below the 25th percentile) and typical readers. All risk factors, cognitive factors, and teachers' perceptions were included in the analysis. Productive letter knowledge, teachers' predictions, and rapid naming of colors contributed significantly to the classification ($p < .001$). The canonical correlation between these variables and group membership (above or below the 25th percentile) was .49. Standardized discriminant function coefficients reflecting the contribution of each variable showed that productive letter knowledge contributed most to correct classification; rapid naming of colors contributed .36, teachers' predictions .47, and productive letter knowledge −.64. The discriminant analysis had an overall accuracy rate of 70.8%. The number of valid positives was 46.3%, whereas 12% of the children were misclassified (i.e., false negatives). The number of valid negatives was 88%, and 53.7% of the children were predicted to develop RD but turned out to be typical readers (false positives).

Discussion

Our first question concerned the relationship between risk factors, cognitive factors, and teachers' predictions and perceptions on the one hand and reading ability on the other hand. Risk factors turned out to play a minor role in the prediction of the reading performance of Dutch children in Grade 1. Students without speech-language therapy only seemingly outperformed those students enrolled in speech-language therapy in Grade 1; the difference, however, did not reach significance. In kindergarten, students with a history of speech-language therapy performed significantly worse on all tests. The lower performance on letter knowledge and the lower (though not significant) reading performance in students with speech or language disorders correspond with the findings of earlier studies (e.g., Carroll & Snowling, 2004; Scarborough, 1990). Note that most studies include formal speech or language tests, and few have used enrollment in speech-language therapy as an indicative variable for reading problems. Further research is required to establish the value of this variable in identifying students at risk. We suggest gathering detailed information about the treatment (kinds of problems, duration of the treatment) to increase the accuracy of the identification of students at risk.

Furthermore, and surprising enough, the current study provided no evidence for a hereditary factor in reading or speech-language problems: The test performance of students with and students without a family history of dyslexia or speech-language difficulties was statistically equal for all measures. This result is in contrast with a number of studies performed with English-speaking children (e.g., Pennington & Lefly, 2001; Scarborough, 1989, 1991; Snowling et al., 2003). Scarborough (1998) reviewed several studies and concluded that the family incidence of dyslexia certainly increases the risk for RD in a child. However, the degree of risk varies among studies and may be caused by the way in which information was gathered. In the current study, parents were interviewed by means of a questionnaire. If we had tested the reading ability of the parents and their speech and language performance, our results might have been different. Scarborough (1989) showed that although self-reported RD in parents significantly predicted children's reading performance, predictions were more accurate when test results were used. With respect to gender, the difference between boys and girls did not reach significance for reading performance at the beginning of Grade 1.

Scarborough (1998) also concluded in another review that boys are only slightly more at risk for RD than girls.

With respect to cognitive factors, the relative importance of kindergarten test scores for reading performance was analyzed with the use of structural equation modeling. Letter knowledge turned out to be the strongest predictor, followed (remotely) by rapid naming. The important role of letter knowledge in predicting reading performance has been found in previous studies (e.g., Braams & Bosman, 2000; Catts et al., 2001; de Jong & van der Leij, 1999; Gallagher et al., 2000; Hammill, 2004; Pennington & Lefly, 2001). However, the adequacy of this predictor might be restricted to the beginning stage of reading. Walsh, Price, and Gillingham (1988) demonstrated that letter-naming speed correlated much more strongly with reading achievement among kindergartners than among students in Grade 2. Other studies yielded similar results, demonstrating declining correlations between letter knowledge and reading ability in higher grades (Braams & Bosman, 2000; de Jong & van der Leij, 1999; Wesseling & Reitsma, 2000).

The relationship between rapid naming and reading performance was much weaker, although it also reached significance. The predictive value of rapid naming has been demonstrated in several studies. However, the present study only provides weak evidence for the predictive value of rapid naming in kindergarten. The strength of the relationship in the current study would probably have been larger if we had used letters in the test instead of pictures or colors. However, because of the limited letter knowledge of children in kindergarten, this was not possible at the time of testing. Some researchers have suggested that rapid naming is a better predictor of poor reading than it is of typical reading (Hammill et al., 2002; Meyer, Wood, Hart, & Felton, 1998; Petersen, 2002).

The ability to correctly repeat a set of nonwords was not significantly related to reading skills and thus played a minor role in the prediction of decoding skills. The role of nonword repetition in identifying students at risk for RD and the interpretation of the test outcome are not yet quite clear. In English, Snowling and her colleagues (Gallagher et al., 2000; Muter & Snowling, 1998; Snowling, 1981; Snowling et al., 1986; Snowling et al., 2003) demonstrated that poor readers scored significantly worse on the repetition of nonwords than controls. However, effects were influenced by phonological complexity (group differences were most apparent when four-syllable nonwords had to be repeated). It might be that our selection of items was not appropriate, because we included relatively many nonwords with a low level of phonological complexity.

Next, we were interested in whether kindergarten teachers' predictions and perceptions (ratings of task assignment) would be predictive for reading outcome. Several studies (e.g., Coleman & Dover, 1993; Flynn & Rahbar, 1998; Kenny & Chekaluk, 1993; Taylor, Anselmo, Foreman, Schatschneider, & Angelopoulos, 2000; Teisl, Mazzocco, & Myers, 2001) have demonstrated significant relations between teachers' ratings and future school performance. The results of the current study supported these findings and demonstrated that test performance supported the predictions of teachers quite well. Both in kindergarten and in Grade 1, the performance of students who were expected to develop RD was lower than the performance of students with positive predictions of reading ability. This pattern was clear in all tests. The role of kindergarten teachers' predictions may be underestimated because students who repeated kindergarten were excluded from the analyses. Those students were probably expected to develop RD in Grade 1 with great certainty. The power of teachers' judgments was further demonstrated by the results of the task assignment scale. Kindergartners who were assigned to high levels of task assignment performed significantly better (both in kindergarten and Grade 1) than students with moderate or low task assignment. This result corresponds with that of Kenny and Chekaluk (1993), who reported a significant relation between teachers' perceptions and reading achievement. Flynn and Rahbar (1998) reported high correspondence between teacher ratings and letter-sound knowledge, but low correspondences between teacher ratings and all other measures (vocabulary, syntax, visual discrimination, form copy). In sum, the present findings demonstrated the usefulness of teachers' predictions and ratings for classifying children at risk for RD.

The second question concerned the discriminatory ability of predictive factors. Rather than looking at the ability to predict actual scores, we examined the ability to predict whether a child will fail in reading or not. This should be an import property of a screening battery. To evaluate this capacity, calculations of the sensitivity index, specificity index, and positive predictive value were performed. For practical use in a screening battery, these indexes need to reach a minimum value of .75 (Hammill et al., 2002). In our study, only the specificity index matched this criterion. This means that most of our predictors were only able to correctly identify those students who did *not* develop RD. For all predictors, the sensitivity was much lower than the specificity, which is a general trend in prediction studies (e.g., Hammill et al., 2002; Pennington & Lefly, 2001; Petersen, 2002; Schneider & Näslund, 1993; Teisl et al., 2001; see Scarborough, 1998, for an overview). Thus, in general, the number of false positives and false negatives was too high. In particular, the large number of false positives—that is, the percentage of students who were predicted to have RD but who turned out to be typical readers (ranging from 45% to 72.2%)—was troublesome. Several other studies also demonstrated the occurrence of too many false positives (e.g., Catts et al., 2001; Flynn & Rahbar, 1998; O'Connor & Jenkins, 1999; Taylor et al., 2000) or false negatives (e.g., Coleman & Dover, 1993; Hammill et al., 2002; Mantzicopoulos & Morrison, 1994). In contrast to the findings of other studies, RD were identified very early in Grade 1—that is, after 2 months of formal reading instruction. However, as noted in the introduction, results from other Dutch studies demonstrated strong stability of reading scores during primary school (Bast & Reitsma, 1998; Verhoeven & van Leeuwe, 2003). Thus, the negative results in our study are unlikely to be due to early assessment. In sum, risk factors, cognitive factors, and teachers' perceptions were not sufficiently adequate on their own to correctly identify those students who exhibited RD.

The results of a discriminant function analysis, however, demonstrated that a combination of productive letter knowledge, rapid naming of colors, and teachers' predictions increased the accuracy of prediction to an overall accuracy rate of 70.8%. These results are consistent with the results of Pennington and Lefly (2001), who performed a discriminant function analysis and demonstrated that letter-name knowledge and rapid serial naming of colors and objects were most important in predicting RD.

In conclusion, the results of this study suggest that group membership (RD or not) in Dutch students at the beginning of Grade 1 can be moderately predicted in kindergarten. Although letter knowledge seems to be a strong correlate of word reading, it cannot on its own correctly identify students who will develop RD. To improve the accuracy of classification, a combination of variables is needed (see also Scarborough, 1998). Surprising enough, teachers' predictions and perceptions seem at least as effective as cognitive factors and might, therefore, contribute to accurate prediction. A combination of productive letter knowledge, teachers' predictions, and rapid naming of colors accurately classified 71% of the children. Unfortunately, however, too many students who were predicted to have RD turned out to be typical readers, and too many poor readers were not identified in kindergarten. These results suggest that kindergarten measures moderately predict subsequent reading skills in Dutch children. However, the results might have been different if, in addition to letter knowledge, the battery had included other measures, such as one or more tests of phonemic awareness and a test of verbal memory. In the Netherlands, a great deal of attention is already focused on the early screening and prediction of RD, especially since the implementation of the *Protocol Leesproblemen en Dyslexie* (*Protocol Reading Problems and Dyslexia*; Wentink & Verhoeven, 2001), which includes a checklist for kindergartners. Letter knowledge, phonological awareness, and phonemic awareness are part of this checklist. In addition to the assessment of letter knowledge, we recommend the use of kindergarten teachers' predictions. These two measures are low cost and accurate and require only a limited amount of time. Finally, reading skills in Grade 1 have to be assessed as early as possible-that is, after 2 months of formal reading instruction, as recommended by Wentink and Verhoeven (2001). After all, the best predictor of future reading is reading itself (see, e.g., Hammill, 2004).

REFERENCES

Allor, J. H. (2002). The relationships of phonemic awareness and rapid naming to reading development. *Learning Disability Quarterly, 25, 47–57.*

Aro, M. (2004). Learning to read; The effect of orthography. Jyväskylä, Finland: University of Jyväskylä.

Badian, N. A. (1999). Reading disability defined as a discrepancy between listening and reading comprehension: A longitudinal study of stability, gender differences, and prevalence. *Journal of Learning Disabilities, 32,* 138–148.

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology, 34,* 1373–1399.

Bishop, D. V. M. (2001). Genetic influences on language impairment and literacy problems in children: Same or different? *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 189–198.

Blachman, B. A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology, 76,* 610–622.

Blaiklock, K. E. (2004). The importance of letter knowledge in the relationship between phonological awareness and reading. *Journal of Research in Reading, 27,* 36–57.

Blonk, A. M. T., & Bosman, A. (2003, November). Girls are better, even girls with dyslexia?! Gender differences in reading performance of teenagers with and without dyslexia. Paper presented at the International Conference on Dyslexia, University of Amsterdam, the Netherlands.

Braams, T., & Bosman, A. M. T. (2000). Fonologische vaardigheden, geletterdheid en lees- en spellinginstructie. [Phonological skills, literacy, and reading and spelling instruction]. *Tijdschrift voor Orthopedagogiek, 39,* 199–211.

Carroll, J. M., & Snowling, M. J. (2004). Language and phonological skills in students at high risk of reading difficulties. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45,* 631–640.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten students: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32,* 38–50.

Coleman, J. M., & Dover, G. M. (1993). The RISK Screening Test: Using kindergarten teachers' ratings to predict future placement in resource classrooms. *Exceptional Children, 59,* 468–477.

Cornwall, A. (1992). The relationship of phonological awareness, rapid naming, and verbal memory to severe reading and spelling disability. *Journal of Learning Disabilities, 25,* 532–538.

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

Flynn, J. M., & Rahbar, M. H. (1994). Prevalence of reading failure in boys compared with girls. *Psychology in the Schools, 31,* 66–71.

Flynn, J. M., & Rahbar, M. H. (1998). Improving teacher prediction of students at risk for reading failure. *Psychology in the Schools, 35,* 163–172.

Gallagher, A., Frith, U., & Snowling, M. J. (2000). Precursors of literacy delay among children at genetic risk of dyslexia. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41,* 203–213.

Hammill, D. D. (2004). What we know about correlates of reading. *Exceptional Students, 70,* 453–469.

Hammill, D. D., Mather, N., Allen, E. A., & Roberts, R. (2002). Using semantics, grammar, phonology, and rapid naming tasks to predict word identification. *Journal of Learning Disabilities, 35,* 121–136.

Irausquin, R. (1999). Nonword repetition test. Unpublished test materials, University of Nijmegen, the Netherlands.

Johnston, R. S., Anderson, M., & Holligan, C. (1996). Knowledge of the alphabet and explicit awareness of phonemes in pre-readers: The nature of the relationship. *Reading and Writing, 8,* 217–234.

Kenny, D. T., & Chekaluk, E. (1993). Early reading performance: A comparison of teacher-based and test-based assessments. *Journal of Learning Disabilities, 26,* 227–236.

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology, 95,* 453–464.

Lewis, B. A., Freebairn, L. A., & Taylor, H. G. (2000). Academic outcomes in students with histories of speech sound disorders. *Journal of Communication Disorders, 33,* 11–30.

Lyytinen, H., Aro, M., Eklund, K., Erskine, J., Guttorm, T., Laakso, M.-L. et al. (2004). The development of children at familial risk for dyslexia: Birth to early school age. Annals of Dyslexia, 54, 184–220.

Mantzicopoulos, P. Y., & Morrison, D. (1994). Early prediction of reading achievement: Exploring the relationship of cognitive and noncognitive measures to inaccurate classifications of at-risk status. *Remedial and Special Education, 15,* 244–251.

Menyuk, P., Chesnick, M., Liebergott, J. W., Korngold, B., D'Agostino, R., & Belanger, A. (1991). Predicting reading problems in at-risk students. *Journal of Speech and Hearing Research, 34,* 893–903.

Meyer, M. S., Wood, F. B., Hart, L. A., & Felton, R. H. (1998). Selective predictive value of rapid automatized naming in poor readers. *Journal of Learning Disabilities, 31,* 106–117.

Mommers, M. J. C., Verhoeven, L., & van der Linden, S. (1979). Veilig leren lezen [Learning to read safely]. Tilburg, the Netherlands: Zwijsen.

Mommers, M. J. C., Verhoeven, L., & van der Linden, S. (1994). Veilig leren lezen [Learning to read safely] (2nd ed.). Tilburg, the Netherlands: Zwijsen.

Muter, V., & Snowling, M. (1998). Concurrent and longitudinal predictors of reading: The role of metalinguistic and short-term memory skills. *Reading Research Quarterly, 33,* 320–337.

O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading, 3,* 159–197.

Pennington, B. F., & Lefly, D. L. (2001). Early reading development in children at family risk for dyslexia. *Child Development, 72,* 816–833.

Petersen, D. K. (2002). Prediction of poor and superior word reading. Working Papers, 50, 17–32.

Satz, P., & Fletcher, J. M. (1988). Early identification of learning disabled children: An old problem revisited. *Journal of Consulting and Clinical Psychology, 56,* 824–829.

Scarborough, H. S. (1989). Prediction of reading disability from familial and individual differences. *Journal of Educational Psychology, 81,* 101–108.

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

Scarborough, H. S. (1991). Antecedents to reading disability: Preschool language development and literacy experiences of children from dyslexic families. *Reading and Writing, 3,* 219–233.

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

Schneider, W., & Näslund, J. C. (1993). The impact of early metalinguistic competencies and memory capacity on reading and spelling in elementary school: Results of the Munich Longitudinal Study on the genesis of individual competencies. *European Journal of Psychology of Education, 8,* 273–287.

Snowling, M. J. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research, 43,* 219–234.

Snowling, M., Goulandris, N., Bowlby, M., & Howell, P. (1986). Segmentation and speech perception in relation to reading skill: A developmental analysis. *Journal of Experimental Child Psychology, 41,* 489–507.

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

Taylor, H. G., Anselmo, M., Foreman, A. L., Schatschneider, C., & Angelopoulos, J. (2000). Utility of kindergarten teacher judgments in identifying early learning problems. *Journal of Learning Disabilities, 33,* 200–210.

Teisl, J. T., Mazzocco, M. M., & Myers, G. F. (2001). The utility of kindergarten teacher ratings for predicting low academic achievement in first grade. *Journal of Learning Disabilities, 34,* 286–293.

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

van den Bos, K. P. (2004). Test serieel benoemen en woorden lezen [Test of serial naming and word reading]. Unpublished test materials. Groningen: RUG.

van Doorn, F. (1996). Werkhoudingslijst [Task-assignment scale]. Amsterdam: Advies- en Begeleidingscentrum.

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

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192–212.

Walsh, D. J., Price, G. G., & Gillingham, M. G. (1988). The critical but transitory importance of letter naming. *Reading Research Quarterly, 23,* 108–122.

Weiner, P. (1985). The value of follow-up studies. *Topics in Language Disorders, 5,* 78–92.

Wesseling, R., & Reitsma, P. (2000). The transient role of explicit phonological recoding for reading acquisition. *Reading and Writing, 13,* 313–336.

Wentink, H., & Verhoeven, L. (2001). Protocol voor leesproblemen en dyslexie [Protocol for reading difficulties and dyslexia]. Nijmegen, the Netherlands: Expertisecentrum Nederlands.

Chapter 3: The effects of a semantically-oriented reading intervention for poor beginning readers[1]

Abstract

This study was designed to investigate the effects of a semantic-reading program for children at risk for reading disabilities. In Experiment 1, 121 poor beginning readers (mean age 6.5 years) from 22 regular primary schools were assigned to a semantically-oriented training, a phonologically-oriented training or a control group. Results showed an advantage for the semantic training over the phonological training after four months of training. At post-test, however, children in both experimental training programs showed similar gains in word-identification skills. In Experiment 2, the experimental-training programs were modified and extended. About 83 poor beginning readers participated in one of both experimental groups or were assigned to a control group. Letter knowledge, word-identification skills, text-reading skill and receptive vocabulary were assessed during the training and reading skills were assessed at follow-up, mid-Grade 2. Result showed that all groups performed statistically equally across all measures.

---

[1] I am are very grateful to Ton Mekking from ATI-software who designed and developed the computer programs. I also thank Marianne van Lith, who was of crucial importance in collecting the data, supporting teachers, and motivating children, and all students who assisted in collecting the test data. Finally, I am grateful to all children who participated in the study and to the teachers who invested a considerable amount of time in the training experiment.

Introduction

Reading involves the transformation of a string of graphemes into the correct phonemic units and semantic units. The important role of phonology in the word-recognition process has been extensively demonstrated by several researchers (for overviews see Berent & Perfetti, 1995; Bosman & van Hell, 2002; Bosman & Van Orden, 1997; Frost, 1995; Jared & Seidenberg, 1991; Van Orden, Pennington, & Stone, 1990). Since phonological deficits are assumed to be a major cause in reading difficulties, many researchers have constructed intervention studies that are phonologically oriented (see e.g., Gonzalez, Espinel, & Rosquete, 2002; McGuinness, McGuinness, & McGuinness, 1996; for an overview, see Bus & van IJzendoorn, 1999, and Ehri et al., 2001). In contrast, the contribution of semantics to orthography and the nature of this relationship has received considerable less attention and seems somewhat controversial.

Evidence for the role of semantics in word recognition has been provided by semantic-priming studies. Meyer and Schvaneveldt (1971) were the first who demonstrated the associative-priming effect. In their lexical-decision test, participants responded faster and more accurately to semantically related pairs of words (e.g., doctor and nurse) than to unrelated pairs of words (e.g., doctor and cloud). Since then, many researchers have replicated and extended the results to priming paradigms (presenting the word doctor prior to nurse, facilitates reading the latter). Semantic-priming studies have been carried out with both adult participants (e.g., Abad, Noguera, & Ortells, 2003; Spruyt, Hermans, De Houwer, & Eelen, 2004, see Neely, 1991, for a review) and children (e.g., Assink, Van Bergen, Van Teeseling, & Knuijt, 2004). In general, these studies have demonstrated that words preceded by a semantically related prime are recognized faster and with fewer errors than words preceded by an unrelated prime. This finding might indicate that the meaning of a word has been activated before the word has been completely recognized. This way, semantic activation facilitates word identification.

A second line of evidence for the role of semantics in word recognition has been provided by studies on isolated word recognition: Studies of ambiguity, that is, words associated with multiple meanings (Borowsky & Masson, 1996; Hino, Lupker, & Pexman, 2002; Locker, Simpson, & Yates, 2003; Pexman & Lupker, 1999), synonyms (Pecher, 2001), imageability (Cortese, Simpson, & Woolsey, 1997; de Groot, 1989; Raman & Baluch, 2001; Strain, Patterson, & Seidenberg, 1995), and number of features (Pexman, Lupker, & Hino, 2002). In general, these studies have demonstrated that words with rich semantic representations (i.e., ambiguous words, high-imageable words, and words with a large number of features) are recognized faster and with fewer errors than words with fewer enriched semantic features. By manipulating one of the above named variables, results of these studies have provided converging evidence for semantic contribution in word recognition. These results have been established in both lexical-decision tasks and naming tasks.

One way to explain semantic effects in word recognition is the connectionist account. Probably the most influential connectionist model is the distributed, developmental model of word recognition by Seidenberg and colleagues (Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland,1989). This model assumes that reading words involves the computation of three types of codes: Orthographic, phonological, and semantic codes. Each of these codes is a distributed representation. In this view, a word is not a local unit in memory, but a pattern of distributed activation. Based on the encounters with written words, a reader computes connections of varying strengths between orthographic, phonological, and semantic nodes. The translation from print to sound has been implemented in a computational model including grapheme units that send activation via hidden units to phoneme units. Interaction between orthographic, phonological and semantic units has not been implemented yet. Another example of these types of models is the Phonological Coherence Model (Bosman & Van Orden, 1997; Farrar & Van Orden, 2001; Van Orden, Pennington, & Stone; 1990; Van Orden et al., 1997). This model is fully recurrent, thus assuming interconnectivity between families of grapheme nodes, phoneme nodes and semantic nodes. Relations within the model are both excitatory (between node families) and inhibitory (within a family). In the model, the presentation of a printed word activates letter nodes that, in turn, activate phoneme and semantic nodes. Following initial activation, recurrent feedback dynamics begin among all these node families. The model has been successfully implemented in a computational model (Farrar & Van Orden, 2001).

It has been hypothesized that in ambiguous words, words with a large number of features or words with a large semantic neighborhood, multiple semantic nodes will be activated, which causes enhanced feedback strength from semantics to orthography. In contrast, the opposite is true for words with synonyms (Pecher, 2001). In synonyms, the string of letters (e.g., movie) will activate semantic nodes. These semantic nodes will send feedback activation to different strings of letters (movie, film, cinema). This inconsistent mapping from semantics to orthography interferes with word recognition and causes longer response times and more errors.

Taken together, it seems that semantics, at least in the English language, plays a role in word recognition and this effect can be attributed to feedback from semantics to orthography in a fully recurrent network with bi-directional activation flows. This feedback causes both facilitation and competition effects.

Several researchers have demonstrated that the effect of semantics is limited to low-frequency (irregular) words (e.g., Cortese, Simpson, & Woolsey, 1997; Raman and Baluch, 2001; Strain, Patterson, & Seidenberg, 1995, but see, Baluch & Besner, 2001; Jared, 1997). Thus, it is assumed that semantic effects are much greater when phonological coding is more slowly and error-prone (i.e., in low-frequency words). If this premise is correct, then it seems plausible that semantic activation will play a more important role in poor readers than in skilled readers. After all, poor readers generally exhibit slow phonological-coding abilities. It has been hypothesized that when recoding is inefficient, more time is left for semantics to intervene. Strain and Herdman (1999) considered this issue. They failed to find an interaction between imageability and regularity in participants with poor phonological skills: Imageability effects in regular words were as strong as in exception words. In medium to high-skill participants, however, significant interactions between these variables were found: Greater imageability effects in exception words than in regular words. The authors concluded that the imageability effect increased with decreasing phonological skill.

With respect to semantic priming, few studies have investigated priming effects in poor readers and good readers. Assink et al. (2004) addressed this issue in children from Grades 3 and 5 and found no differences in semantic priming effects between normal decoders and poor decoders.

To conclude, readers benefit from semantic activation during reading, and this advantage might be at least as effective for poor readers as for good readers. The goal of the present experiments was to examine whether semantic activation can be successfully implemented in a reading intervention program and whether the presumed semantic activation will support poor decoding skills. Few studies have investigated the effect of a semantic intervention program. We will discuss two examples in detail, one for the Dutch language (Berends & Reitsma, 2006) and one for the English language (Norbury & Chiat, 2000).

Berends and Reitsma (2006) investigated the effects of a semantically-oriented training program in Dutch first graders and second graders with poor decoding skills. The training was initiated after 6 months (Grade 1) and 16 months (Grade 2) of formal reading instruction. During four weeks, children practiced individually with 10 target words and 10 control words which were presented repeatedly, either in a semantic-based training or in an orthographic-based training. The semantic training included two types of exercises: Association and question. In the association task, the child had to determine whether two words belonged to the same semantic category or not. In the question format, a question was presented, followed by a target word. The child had to judge whether the target word was the correct answer to the question. A few days after the end of the training, the same target words were presented in a posttest. In Grade 1, no main effect of type of training, nor any interaction with type of training was significant. In Grade 2, results revealed a significant time of testing (pretest vs. posttest by word type (target vs. control) by type of training (semantic vs. orthographic) interaction in reaction times: Target words were read faster than control words and this difference was larger in the semantic training than in the orthographic training. Thus, only after the initial stage of learning to read, a semantic training facilitated word decoding more than a phonological training without semantics.

Norbury and Chiat (2000) investigated the effects of a semantic intervention in a single-case study. An 8.5 years-old boy, WS, with poor phonological-awareness skills but relatively good conceptual knowledge was provided with a semantic training to increase word-identification skills. Words were assigned to either a treatment condition (targets and foils) or a control condition (not presented during intervention). Targets were presented in semantic activities as matching written words with pictures, judging which words out of three are semantically related, free association, and quiz games. Foils appeared in the training as possible choices, but without direct focus to their semantic attributes. During five weeks, WS practiced twice a week for approximately one hour. Results showed significant reading gains for both target words, foils, and controls. The improvement of target words, however, was significantly greater than the improvement of foils and controls. An additional benefit of the semantic intervention was the positive experience for WS. He seemed very successful in semantic activities. The authors conclude that a semantically-based training contributes to the improvement of word-identification skills when explicit connections between orthography, phonology, and meaning have been made.

To conclude, the effects of a semantically-oriented intervention program have hardly been examined. Moreover, semantic interventions have been carried out over a relatively brief period of time and no study included children from the beginning of Grade 1, as soon as reading instruction begins.

Experiment 1 was conducted to investigate the effects of a semantically-oriented training program for children in Grade 1, who are at risk for reading disabilities. Already after two months of formal reading instruction, children who performed below average on a screening test for reading skills were presented either a phonologically-oriented training program or a semantically-oriented training program to increase word-identification skills. Both training programs were implemented on a computer. We started the intervention early in Grade 1, because there is general agreement among researchers that reading intervention is most effective early in childhood (e.g., Foorman, Breier, & Fletcher, 2003; Nicolson, Fawcett, Moss, Nicolson, & Reason, 1999).

In our phonologically-oriented training, word-identification skills were trained in the absence of a semantic context; adequate grapheme-phoneme mappings sufficed for a correct response. In the semantically-oriented training, additional to correct grapheme-phoneme mappings, semantic activation was necessary for a correct response. Semantic activation was elucidated by the inclusion of pictures or was stimulated by the type of response requested. Importantly, we aim at an automatic semantic compensation process rather than intentional compensation, in which the participant makes use of contextual cues and guessing strategies (e.g., Bowey, 1985; Kim & Goetz, 1994; Pring & Snowling, 1986). The semantic-training program was not aimed at increasing use of contextual cues and stimulation of semantic strategies, rather was designed to increase word-identification skills in the context of automatic-semantic activation. For example, words (e.g., rose, rope, tulip, tube) were successively presented on the screen and children had to judge whether each word belonged to the semantic category of flowers. Presentation of the semantic category (e.g., flowers) will probably activate exemplars of the category (e.g., rose, tulip, and narcissus), semantic associates (e.g., grass, vase, and smell), or other semantic attributes. But, guessing would lead to incorrect responses like rope or tube. Thus, proper manipulation of the distracters prevented students from guessing. Both training programs were additional to reading instruction in the classroom and were aimed at increasing word-identification skills. The goal of Experiment 1 was threefold.

The first aim was to evaluate the efficacy of both training programs in a natural school setting. Since several training studies have demonstrated positive effects of computer training for reading development, we expected that both training programs would be at least as effective as additional support by a remedial teacher. Blok, Oostdam, Otter, and Overmaat (2002) reviewed 42 studies over the past two decades and showed positive effects of computer-assisted instruction in beginning readers.

The second aim was determining which of the training programs - a training focused on semantic activation or a training without a semantic context - was the most effective one for the development of phonological skills, reading new words, pseudowords, and text. Based on the assumption that semantics are activated automatically in reading and given the assumption that poor readers will show a large influence of the semantic attributes, we hypothesized that children would profit more from a training in a semantically-oriented context. It was hypothesized that semantic activation would reduce the difficulty in linkage of the graphemes to the correct phonemes in children with reading difficulties. Recall that the program was not aimed at guessing words by means of context.

The third aim of the study was to examine the responsiveness to intervention. Responsiveness to intervention (RTI) is the extent to which a change in behavior or performance occurs as a result of an intervention. Although early identification and intervention reduces the number of poor readers at the end of first or second grade considerably, there are children who barely respond to intervention and still perform below standard after intervention (e.g., Berninger et al., 2002; Torgesen, 2000). These children are referred to as treatment resisters or slow responders. The number of treatment resisters reported in the literature varies considerably and depends on the selection criteria for intervening and the cut-off level for establishing success of the intervention. In Experiment 1, we selected children on the basis of actual reading scores at the beginning of Grade 1. We used a cut-off level of 25th percentile on a word-identification test at the end of the training, to judge success of intervention. The goal was to establish how many children who performed poorly on word-identification skills at the beginning of the training were still poor readers (scoring below the 25th percentile) at the end of Grade 1, after the intervention had taken place. The reason for carrying out Experiment 2 will be discussed later on.

Experiment 1

Method

Participants

Participants were 121 first graders who were at risk for reading disabilities. Risk status was defined by children's reading performance after two months of formal reading instruction. Characteristics of the participants are listed in Table 1.

Table 1. Characteristics of the Experimental Groups and Control Group

	Experimental groups				Control group	
	Semantic (*n*=32)		Phonological (*n*=59)		(*n*=30)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Number of schools	5		12		5	
Test assessment (days)						
test interval 1-2	121.5	7.3	116.8	15.4		
test interval 1-3	211.4	9.3	210.9	15.9	220.7	7.6
Age at time 1 (months)	78.9	5.1	79.0	3.7	77.9	4.8
Sex						
boys	62.5%		64.4%		63.3%	
girls	37.5%		35.6%		36.7%	
Instruction method						
Veilig Leren Lezen	100%		100%		30%	
other	0%		0%		70%	

All children in the experimental groups were instructed with 'Veilig Leren Lezen' ['Learning to Read Safely'], the most widely used reading program in Dutch schools (Mommers, 1979, 1994). The emphasis in this method is on the structure of the orthographic system and the relationship between letters and sounds (phonics). Initially only consistent words are used. After four months of instruction the children are familiar with the main grapheme-phoneme correspondence rules. It is a fairly rigid pre-programmed curriculum, which imposes a strict day-by-day and week-by-week progression.

Materials

Measures

Letter-knowledge

Graphemes Test 1 [Grafementoets] by Wentink and Verhoeven (2001) and Graphemes Test 2 [Grafementoets] by Verhoeven and Van Kuijk (1992).

Both tests consisted of all 34 graphemes of the Dutch language. Graphemes Test 1 was divided in two parts: Part 1 consisted of 16 graphemes, which were already taught at test assessment. Part 2 consisted of 18 graphemes, which were not yet directly instructed. The child was asked to read aloud the graphemes of Part 1 as quickly as possible. Then, the experimenter named the sounds of the graphemes, that were read correctly by the child. The child had to indicate the printed letter that matched the sound. Subsequently, Part 2 was administered in the same way, but this time in a non-speeded paradigm. The test score was the number of graphemes read correctly and the time to perform Part 1. In the Graphemes Test 2, the child was asked to read aloud all 34 graphemes as correctly and quickly as possible. The test score was the number of graphemes read correctly and the time to perform the test.

Phoneme Test 1 [Fonemendictee] by Wentink and Verhoeven (2001) and Phoneme Test 2 [Fonemendictee] by Verhoeven and Van Kuijk (1992).

Phoneme Test 1 consisted of 16 items. The experimenter read aloud a word and dictated one phoneme of the word that the child had to write down. The score was the number of phonemes correctly written. The maximum score was 16. Phoneme Test 2 consisted of 34 items. The score was the number of phonemes correctly written. The maximum score was 34.

Phonological skills

Auditory-blending Test [Toets voor Auditieve Synthese] by Verhoeven and Van Kuijk (1991).

This test consisted of 20 words. The experimenter named the phonemes of each word and the child was asked to blend the phonemes and to name the word. The maximum score was 20.

Auditory-segmentation Test [Toets voor Auditieve Analyse] by Verhoeven and Van Kuijk (1991).

This test consists of 20 words. The experimenter read aloud the words and the child was asked to segment the words into phonemes. The maximum score was 20.

Reading measures

Word-reading test [Toets Woorden lezen] by Wentink and Verhoeven (2001).

This test consisted of three parts, each of them including 10 (C)VC-words (C is consonant, V is vowel), referred to as Word Reading 1 to Word Reading 3. Word Reading 1 consisted of well-known words, which were taught in the classroom. Word Reading 2 consisted of new words, which differed one letter from the words taught in classroom. In this way, the items resembled the words, which were already taught. Word Reading 3 consisted of new words, which differed two letters from the words taught in classroom. In this condition, there was hardly any resemblance to the words that were already taught. The child was asked to read aloud the words as fast and as accurately as possible. The score on each test was the number of words read correctly (accuracy) and the time to perform the test (speed). For the analyses, we computed the mean number of items read correctly and the mean speed in Word Reading 2 and 3.

Nonword-reading test [Pseudowoordentest] by Van Leerdam (1996).

This test consists of two parts. Nonword-reading test 1 contains 60 CV/CVC-nonwords and nonword-reading test 2 consists of 60 CCVC/CVCC-nonwords. The nonwords are derived from the words of the standardized Three-Minutes-Test (Verhoeven, 1992). In each test, the child was asked to read aloud the nonwords as fast and as accurately as possible in one minute. The score was the number of items read correctly within one minute and thus reflected both speed and accuracy.

Three-minutes test [Drie-Minuten-Toets, DMT] by Verhoeven (1992).

This test consists of three cards. In the present study, Cards 1 and 2 were presented. Card 1 contains 150 CV/VC/CVC-words, Card 2 consists of 150 CCVC/CVCC-words. The child was asked to read aloud the words as correctly and as quickly as possible. The score was the number of items read correctly within one minute and thus reflected both speed and accuracy.

AVI-test [AVI-toets] by Visser, van Laarhoven and ter Beek (1998).

This test consists of nine cards with stories of increasing complexity. In the current study, only the first card was used. The story consists of short sentences with one-syllable CV/VC and

CVC-words. The child was asked to read aloud the story as correctly and as quickly as possible. The speed and the number of errors were registered.

Training

Experimental groups

Both experimental-training programs consisted of reading exercises at the level of graphemes, words, and short sentences. Each training session comprised three different types of exercises. The order in which new graphemes and word structures were presented was adjusted to the classroom instruction. The word stimuli in both programs were not identical, but were of comparable difficulty level. At grapheme level, the two training programs differed in the use of additional words in the spoken instruction; in the semantic training, a phoneme was orally presented with support of an instructed word that included the target phoneme. In the phonological training, a phoneme was presented in isolation. This was the only difference between both programs at the grapheme level. At word level, the training programs differed in the presence of a semantic context: In the semantic training, the meaning of a word had to be activated in order to provide a correct response. Words were often embedded in a relevant semantic context (e.g., pictures, semantic categories). In the phonological training, meaning activation was not necessary for a correct answer. In this program, phonological activation was relatively more important. Most words were consistent CVC-words. At sentence level, the distinction of both programs was created by the same principle: In the semantic training, the meaning of a sentence had to be activated in order to provide a correct response. Meaning activation was not strictly necessary in the phonological training.

In both the semantically-oriented training and the phonologically-oriented training, the child had to respond as accurately and as quickly as possible by means of a mouse click. Feedback on accuracy was given on each trial by means of a smiley. Time pressure was discernable by means of a picture of a glass of lemonade. The contents of the glass reduced either until an answer was provided or until the maximum responding time was exceeded. Notwithstanding the important difference (necessity of meaning activation in the semantic training group due to task characteristics), it is important to note that both programs were aimed at training word-identification skills. The reading interventions were supplemental to regular classroom reading instruction. For more detailed information about the content of both training programs, see the Appendix.

Control group

About 63% of the children in the control group received additional training in reading skills by a (remedial) teacher. Children practiced their reading skills either individually or in small groups. The frequency of this training ranged from 2 to 4 times a week (M = 2.9, SD =

.5). Time spent in each training session ranged from 15 to 30 minutes (*M* = 27.1, *SD* = 5.0). Different kinds of exercises were trained, for example, flash-card reading, tutor reading, letter knowledge, auditory segmentation and blending.

Procedure

After two months of formal reading instruction, 719 first graders from 26 regular primary schools in the Netherlands were tested on word-decoding skills (Word-reading test; Wentink & Verhoeven, 2001). Children who did not meet the standard criteria on this test were included in the study, provided that teachers agreed. This resulted in the selection of 185 children; 40 low achievers from five schools were assigned to the control group, 69 children from eight schools were assigned to the semantic group and 76 children from 13 schools were assigned to the phonological group. Subsequently, children from the experimental groups were subjected to the Phoneme Test, Grapheme Test, Auditory Segmentation Test, and Auditory blending Test. A few days after test assessment, the experimental training started. Children of the experimental groups trained three times a week during 5 to 15 minutes individually on a computer. Each training session included three different exercises. Participants were tested on two occasions to assess their progress in reading skills; February (Time 2), and May (Time 3). An overview of the test assessment is presented in Table 2.

Table 2. Test assessment in Experiment 1

Test	Time
Phonological skills	
Auditory segmentation	October, February, May
Auditory blending	October, February, May
Letter Knowledge	
Grapheme Test 1	October
Grapheme Test 2	February, May
Phoneme Test 1	October
Phoneme Test 2	February
Reading measures	
Word-reading test	October
Nonword-reading test	February
Three-minute test	February, May
AVI-test	May

The exact time of testing depended on the progress of reading instruction in the classroom and as a consequence, varied slightly among schools (see Table 1). Finally, we excluded 10 children from the control group because of too many missing data (e.g., due to migration or illness) and 54 children from the experimental groups because (I) the reading instruction method was not ‘Veilig Leren Lezen’ [‘Learning to Read Safely’], (II) teachers finished the training program premature, or (III) children had too many missing data. This resulted in the selection of 121 children. The exclusion criteria led to an imbalance of group sizes, because most children who were excluded from the analyses participated in the semantic group. As a consequence, group size of the semantic group was reduced from 69 to 32 students. In the analyses, the number of participants differs slightly because of missing data.

Results

October: Pretest

First, we examined whether word-identification skills of the experimental groups and the control group were statistically equal before training. In addition to a one-way Anova, Levene's test for homogeneity of variances was conducted, and if differences were found to be significant, the Welch correction for unequal variances was employed. Results of a One-way Anova with reader group (semantic vs. phonological vs. control) as between-subjects variable and speed and accuracy based on the mean scores on Word Reading Tests 2 and 3 as two separate dependent variables revealed that the groups did not differ significantly from each other in the pretest; speed, $F(2,111) = 1.20$, $p = .31$; accuracy, $F < 1$. Means and standard deviations are reported in Table 3.

To ascertain that both experimental groups had comparable levels of phonological skills, we compared performance of both groups on auditory-segmentation skills, auditory-blending skills, and on letter knowledge (Grapheme Test 1 and Phoneme Test 1). Results of a One-way Anova showed that performance of both experimental groups was statistically equal in all tests, F's < 1, except for Phoneme Test 1, $F(1,61) = 9.04$, $p < .01$ with Welch's correction. In the latter test, children in the semantic group ($M = 14.3$) wrote more letters correctly than those in the phonological group ($M = 12.3$).

Table 3. Mean Scores at all Test Moments in Experiment 1

		Semantic group (*n*=32)		Phonological group (*n*=59)		Control group (*n*=30)	
	max	*M*	*SD*	*M*	*SD*	*M*	*SD*
October							
Word-Reading Test							
speed (sec.)		9.5	2.3	9.1	3.2	8.3	3.0
accuracy	10	5.3	2.2	5.0	2.2	5.0	2.8
Auditory Blending	20	11.2	4.9	10.9	5.0		
Auditory Segmentation	20	7.7	3.9	7.9	4.3		
Phoneme Test 1	16	14.3	1.6	12.3	3.8		
Grapheme Test 1							
speed (sec.)		32.1	10.8	35.7	17.6		
accuracy	34	20.8	3.0	20.0	3.6		
February							
Auditory Blending	20	17.8	3.3	17.7	2.5		
Auditory Segmentation	20	15.8	4.4	15.8	3.9		
Phoneme Test 2	34	32.3	2.2	31.7	2.4		
Grapheme Test 2							
speed (sec.)		39.7	11.6	44.9	16.7		
accuracy	34	33.3	1.1	32.4	1.4		
DMT 1 (CVC-words)	150	20.2	6.2	16.5	5.2		
DMT 2 (CCVC/CVCC-words)	150	8.9	5.2	6.4	4.3		
Non-word reading test							
part 1	60	15.2	5.4	13.5	4.6		
part 2	60	7.5	4.7	5.9	3.5		
May							
Auditory Blending	20	19.5	.9	19.6	.9	18.3	2.0
Auditory Segmentation	20	19.1	1.2	18.4	2.2	18.1	3.7
Grapheme Test 2							
speed (sec.)		31.6	9.6	33.6	9.9	36.1	9.3
accuracy	34	33.6	.7	33.0	1.2	32.4	1.5
DMT 1 (CVC-words)	150	29.7	11.4	26.5	9.8	23.1	8.9
DMT 2 (CCVC/CVCC)	150	15.7	7.9	14.5	6.1	13.1	5.9
AVI-test							
speed (sec.)		136.1	55.5	164.7	94.4	137.2	60.8
errors		4.6	4.0	5.5	3.6	5.7	6.2

February

Four months after the pretest, phonological skills and reading skills of the experimental groups were re-assessed. Means and standard deviations of all tests are presented in Table 3. To investigate whether both experimental groups differed in their test performance, we carried out a One-way Anova with reader group (semantic vs. phonological) as between-subjects factor and scores of all tests (Phoneme Test 2, Grapheme Test 2, Auditory Segmentation and Auditory-Blending Test, Three-Minutes Test 1 and 2, and Nonword-Reading Test 1 and 2) as separate dependant variables.

Phonological skills. Both groups performed statistically equally in the Auditory-Blending Test and Auditory-Segmentation Test, both *F's* < 1. Test results clearly showed that both the phonological group and the semantic group had made considerable progress in their phonological skills since the beginning of first grade. Auditory segmentation and blending skills were not explicitly trained in the experimental intervention programs.

Letter knowledge. Whereas both groups did not differ significantly in performance in the Phoneme Test 2 ($F(1,86) = 1.62$, $p = .21$), accuracy in the Grapheme Test 2 differed significantly, $F(1,90) = 7.93$, $p < .01$; the semantic group ($M = 33.3$) outperformed the phonological group ($M = 32.4$). Recall, spelling graphemes (as measured in the Phoneme Test) was not practiced in the experimental training programs. Reading graphemes, however, was frequently practiced in both the semantic group and the phonological group.

Reading measures. Reading scores of the semantic group were significantly higher than those of the phonological group: DMT1 (CVC-words), $F(1,89) = 9.33$, $p < .01$ and DMT2 (CVCC- and CCVC-words), $F(1,81) = 5.15$, $p < .05$. Mean scores are graphically presented in Figure 1. In decoding nonwords, both groups performed statistically equally: Nonword-reading Test 1, $F(1,89) = 2.65$, $p = .11$, Nonword-reading Test 2, $F(1,81) = 3.24$, $p = .08$.

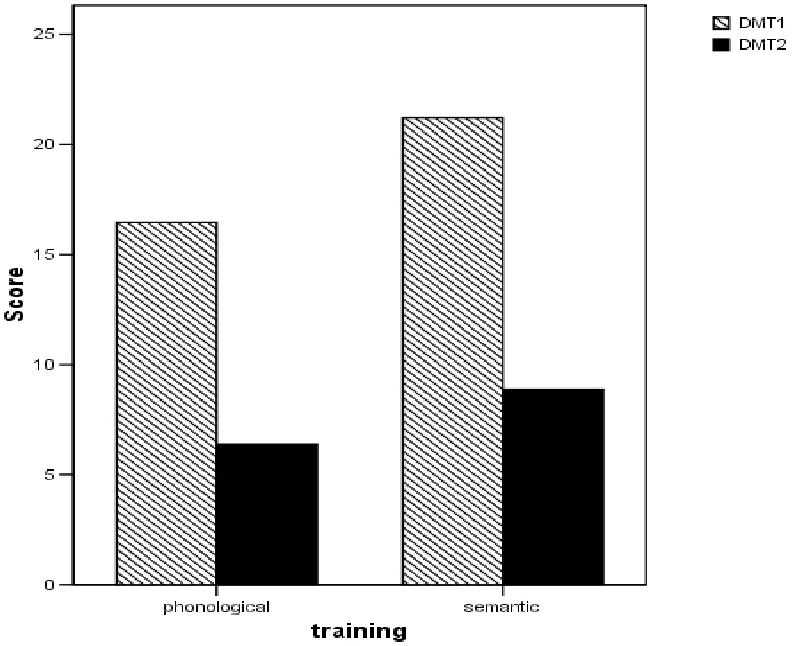

Figure 1. Mean Scores on the DMT1 and DMT2 at Time 2 for the Experimental Groups

May: Post-test

At the end of the training (seven months after the pretest had been carried out), phonological skills and reading skills of the experimental groups were re-assessed and reading skills of the experimental groups and control group were compared. Means and standard deviations of all tests are presented in Table 3. To investigate whether the experimental groups and the control group differed in their test performance, we carried out a One-way Anova with reader group (semantic vs. phonological vs. control) as between-subjects factor and scores on the Grapheme Test 2 (accuracy and speed), Auditory-

Segmentation Test, Auditory-Blending Test, Three-Minutes Test 1 and 2, and AVI-test (accuracy and speed) as separate dependent variables.

Phonological skills. All groups performed statistically equally in the Auditory-Segmentation Test $F(2,109) = 1.16$, $p = .32$. Table 3 shows that all groups performed at ceiling. In the Auditory-Blending test, there was a significant difference between the groups, $F(2,48) = 5.26$, $p < .01$ with Welch's correction. A post-hoc analysis with Dunnett T3 correction for unequal variances showed that the semantic group ($M = 19.5$) and the phonological group ($M = 19.6$) outperformed the control group ($M = 18.3$).

Letter knowledge. There were no significant differences between the groups in speed in the Grapheme Test 2, $F(2,108) = 1.49$, $p = .23$, whereas significant differences were found in accuracy, $F(2,59) = 9.79$, $p < .001$ with Welch's correction. A post-hoc analysis with Dunnett T3 correction for unequal variances showed that the semantic group ($M = 33.6$) outperformed the control group ($M = 32.4$) and the phonological group ($M = 33.0$). Accuracy scores showed that both children from the experimental groups and the control group had mastered nearly all letters at the end of Grade 1.

Reading measures. The only significant difference among groups was found in scores on the DMT1, $F(2,115) = 3.24$, $p < .05$. A Bonferroni post-hoc test revealed that the semantic group ($M = 29.7$) outperformed the control group ($M = 23.1$) Mean scores are graphically presented in Figure 2.

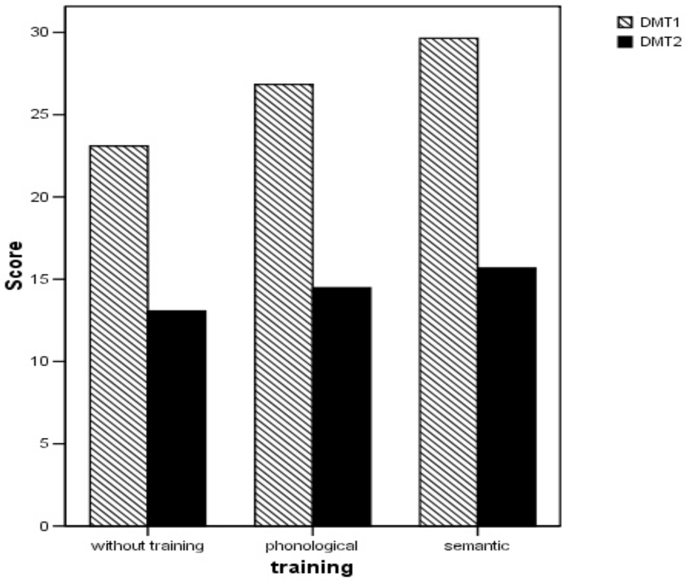

Figure 2. Mean Scores on the DTM1 and DMT2 at Time 3 for the Experimental Group and Control Group.

Responsiveness to intervention

Finally, it is interesting to know the percentage of slow responders to the intervention. In other words, how many at-risk children still had poor word-decoding skills at the end of the training? Poor word-decoding skills were identified by scores on the DMT1 (CVC-words) and DMT2 (CVCC- and CCVC-words). These standardized tests provide representative norms, based on a large sample of participants. To examine the response to intervention, we calculated the percentage of children who performed below the 25th percentile, relative to a norm-referenced basis. These percentages were computed for the experimental groups and control group separately. In the control group, 53.3% of all children scored below the 25th percentile on DMT1. In the experimental groups, 50.9% of the children in the phonological group and 32.3% of the children in the semantic group scored below the 25th percentile. On DMT2, these proportions are respectively 56.7%, 42.9%, and 45.2%.

Experiment 2

Experiment 2 was designed to replicate the results of Experiment 1 in a different sample of participants. In addition, we amplified the training sessions and changed the design on several factors, as described in the method section. The main goal was to re-evaluate the effectiveness of a semantic-oriented training program in beginning readers.

Method

Participants

Participants were 83 first graders who were at risk for reading disabilities. Risk status was defined by children's reading performance after two months of formal reading instruction. All children were instructed with 'Veilig Leren Lezen' ['Learning to Read Safely'] (Mommers, 1979, 1994). A brief description of this reading program was provided in Experiment 1. In contrast to Experiment 1, in which children from the semantic group and the phonological group attended different schools, in Experiment 2, children from both experimental groups attended the same schools: In each school, half of the children were randomly assigned to the phonological-training group and the other half were assigned to the semantic-training group. Table 4 provides participant characteristics.

Table 4. Characteristics of the Experimental Groups and Control Group in Experiment 2

	Experimental groups				Control group	
	Semantic (*n*= 24)		Phonological (*n*= 23)		(*n*= 36)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Number of schools	13		12		5	
Test assessment (days)						
test interval 1-2	92 .8	7 .6	93 .6	8 .7	97.0	12 .1
test interval 2-3	83 .4	11 .5	80 .6	12 .4	84 .4	9.0
test interval 3-4	71 .4	11 .1	72 .0	13 .1	56 .4	7 .7
test interval 4-5	200.0	6.4	200.3	6.4	217.3	5.4
Age (years)						
time 1	6.6	.5	6.7	.4	6.7	.4
time 5	7.8	.5	7.9	.4	7.9	.3
Sex						
boys	62.5%		52.2%		66.7%	
girls	37.5%		47.8%		33.3%	

Materials

Measures

Letter-knowledge tests

These tests were identical to those in Experiment 1.

Reading measures

The same reading measures as in Experiment 1 were applied in the present experiment. In addition, Card 3 of the Three-Minute Test (Verhoeven, 1992) was administered. This card consists of 2-syllable, 3-syllable, and 4-syllable words.

Receptive-vocabulary test

Language Test All Children [Taaltoets Alle Kinderen] by Verhoeven and Vermeer (2001).

This test consists of several subtests. In the current study, only the receptive-vocabulary test was used. The receptive-vocabulary test consists of 96 items. The experimenter read aloud a word and the child had to indicate the corresponding picture out of four pictures. The words were of increasing difficulty and the test was stopped when the child failed on five consecutive items. The score was the total number of pictures indicated correctly. The maximum score was 96.

Training

Experimental groups

The experimental-training programs resembled those of Experiment 1. However, the training was adapted and extended in the following ways. First, each training session comprised four different exercises, whereas in the first experiment, only three exercises were included. Second, the order in which new graphemes and word structures were presented was highly structured in the present experiment, and resulted in nine levels of increasing difficulty. Third, the program was made adaptive to children's reading progress: From the fifth difficulty level, children progressed to more difficult material after surpassing the accuracy criterion of 80% correct on each exercise.

Control group

About 85% of the children in the control group received additional training in reading skills by a (remedial) teacher. The majority of these children (76%) practiced their reading skills in small groups. The frequency of this training ranged from 1 to 5 times a week ($M = 4.2$, $SD = 1.2$). Time spent in each training session ranged from 10 to 60 minutes ($M = 31.0$, $SD= 13.2$). Different kinds of exercises were trained, for example, flash-card reading, tutor reading, letter knowledge, auditory segmentation and blending. Most children (86.2%) made additional use of a computer program.

Procedure

After two months of formal reading instruction, 540 first graders from 23 regular primary schools in the Netherlands were tested on word-decoding skills (Word-reading test; Wentink & Verhoeven, 2001). Children who did not meet the standard criteria on this test were subjected to the Phoneme Test, Grapheme Test and Receptive-Vocabulary Test. Subsequently, we selected 131 children for this study; 48 low achievers from 5 schools were assigned to the control group and 83 children from 18 schools were assigned to one of both experimental groups. After test assessment, the training started for both the experimental groups and control group. During the training, children were tested on three occasions to assess their progress in reading skills: February (Time 2), April (Time 3), and June (Time 4). Mid-Grade 2 (January), a follow-up was performed. Table 5 presents an overview of the tests at each test time.

Table 5. Test assessment in Experiment 2

Test	Time
Letter Knowledge	
Grapheme Test 1	November
Grapheme Test 2	February, April, June
Phoneme Test 1	November
Phoneme Test 2	February
Reading measures	
Word-reading test	November, February, April
Nonword-reading test	February, April, June, follow-up
Three-minute test	February, April, June, follow-up
AVI-test	February, June, follow-up
Vocabulary	
TAK	November, April

Finally, we excluded children who were not instructed with 'Veilig Leren Lezen'. In addition, we excluded participants with incomplete test results (e.g., because of migration or illness), and students who had few training sessions (e.g., due to computer breakdown or insufficient efforts of the teacher). This selection resulted in the inclusion of 83 children. In the analyses, the number of participants differs slightly because of missing data.

Results

November: Pretest

First, we examined whether word-identification skills of the experimental groups and the control group were statistically equal at the pretest. In addition to a one-way Anova, Levene's test for homogeneity of variances was conducted, and if differences were found to be significant, the Welch correction for unequal variances was employed. Results of a One-way Anova with reader group (semantic vs. phonological vs. control) as between-subjects variable and scores on the Word Reading Tests 1, 2 and 3 (speed and accuracy), Phoneme Test, Graphemes Test and Vocabulary Test as separate dependent variables revealed that the groups did not differ significantly from each other in Word-reading tests, speed, all *F*'s < 1, Word-reading tests 1 and 3 accuracy, *F*'s < 1, Word-reading test 2 accuracy, $F(2,80) = 2.02$, $p > .10$, Graphemes Test, $F(2,79) = 1.2$, $p > .10$, Receptive-vocabulary Test, $F < 1$. The only difference between the groups was found in performance on the Phonemes Test, $F(2,34) = 9.73$, $p < .001$ with Welch's correction. Children in the control group ($M = 15.4$) wrote significantly more letters correctly than those in the phonological group ($M = 13.6$) and semantic group ($M = 13.5$). Mean scores on the Grapheme Test and Phoneme Test are reported in Table 6. Table 7 shows the results on the Word-Reading Test.

Table 6. Mean Scores and Standard Deviations on the Grapheme Test (max. = 34) and Phoneme Test (max. = 16 at Time 1, and 34 at Time 3)

	November		**February**		**April**		**June**	
Grapheme Test	*M*	SD	*M*	SD	*M*	SD	*M*	SD
Accuracy								
phonological	21.0	3.1	32.5	2.1	33.5	.7	33.7	.6
semantic	21.7	3.6	32.7	1.7	33.4	.9	33.6	.5
Control	22.4	3.6	33.3	1.5	33.6	.8	33.4	.9
Speed								
phonological			39.7	14.1	31.5	6.9	31.6	9.3
semantic			44.7	13.8	30.9	7.8	30.1	6.5
Control			43.4	15.7	34.7	9.4	31.9	8.2
Phoneme Test								
phonological	13.6	2.6	30.8	3.7				
semantic	13.5	2.6	30.4	3.2				
Control	15.4	1.0	32.5	1.8				

Table 7. Mean Scores and Standard Deviations on the Word-Reading Test

	November		February		April	
	M	*SD*	*M*	*SD*	*M*	*SD*
Word-reading test 1						
Accuracy						
phonological	9.7	.6	9.7	.6	9.9	.2
semantic	9.8	.5	9.7	.7	9.9	.3
control	9.8	.5	9.9	.3	9.9	.2
Speed (sec.)						
phonological	17.3	10.3	10.6	3.4	8.3	4.3
semantic	17.9	8.2	10.7	3.6	8.2	3.2
control	16.0	9.5	8.9	3.2	8.0	2.9
Word-reading test 2						
Accuracy						
phonological	5.8	2.4	8.6	1.6	9.0	1.5
semantic	6.3	1.7	8.6	1.6	9.2	1.0
control	6.9	2.0	9.4	.9	9.4	.9
Speed (sec.)						
phonological	84.5	32.3	26.4	9.4	17.0	7.6
semantic	76.7	22.8	31.1	8.7	18.6	7.5
control	86.2	33.3	27.3	13.5	17.9	8.4
Word-reading test 3						
Accuracy						
phonological	5.5	2.8	8.6	1.5	8.7	1.2
semantic	5.7	2.6	8.3	2.2	9.0	1.3
control	5.8	2.3	9.1	21.6	9.2	1.1
Speed (sec.)						
phonological	95.6	30.2	32.5	13.7	19.5	10.4
semantic	86.5	29.2	37.4	12.7	21.4	7.4
control	92.5	37.0	32.9	13.3	21.5	9.9

Intervention

Letter-knowledge test 2.

Speed. A 3 (Group: semantic vs. phonological vs. control) x 3 (Time: February, April, June) multivariate analysis of variance was performed with repeated measures on the time factor. No significant group by time interaction was found, $F(4,142)= 1.11$, $p = .36$. There was no main effect of group, $F < 1$. Results indicated a significant main effect of time, $F(2,71)= 29.68$, $p < .001$, *partial* $\eta^2 = .46$. Subsequent tests of within-subjects contrasts revealed that the effect of time was only significant between February and April, $F(1,72) = 57.90$, $p < .001$, *partial* $\eta^2 = .47$. From April to June, there was no significant increase in speed, $F(1,72) = 2.07$, $p = .15$. Mean scores are reported in Table 6.

Accuracy. A 3 (Group: semantic vs. phonological vs. control) x 3 (Time: February, April, June) multivariate analysis of variance was performed with repeated measures on the time factor. No significant group by time interaction was found, $F(4,142) = 1.74$, $p = .15$. There was no main effect of group, $F < 1$. Results indicated a significant main effect of time, *F*(2,71)= 7.85, *p* < .001, *partial* η^2 = .18. Subsequent tests of within-subjects contrasts revealed that the effect of time was only significant between February and April, *F*(1,72) = 14.25, *p* < .001, *partial* η^2 = .17. From April to June, there was no significant increase in accuracy, *F* < 1.

Reading measures.

Results of the Word-Reading test (Wentink & Verhoeven, 2001) are reported in Table 7. Table 8 shows the results of the Three-Minutes Test and the Nonword-Reading Test.

Table 8. Mean Scores and Standard Deviations on the Three-Minutes Test and Nonword-Reading Test

	February		April		June		follow-up	
Three-Minutes Test	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Card 1								
phonological	17.8	6.7	28.8	12.2	35.6	14.6	49.9	20.3
semantic	14.0	5.4	24.1	9.9	31.5	15.5	44.4	16.9
control	17.2	7.2	25.6	11.2	31.9	14.4	43.7	19.3
Card 2								
phonological	5.1	3.6	14.9	9.1	21.3	12.2	35.1	19.6
semantic	3.9	3.2	12.6	7.6	18.2	12.6	31.4	19.8
control	5.6	4.6	12.8	9.2	17.9	12.9	31.8	20.8
Card 3								
phonological					12.2	8.8	20.4	14.9
semantic					8.7	7.3	17.0	10.8
control					9.2	8.0	19.1	15.5
Nonword-Reading Test								
Part 1								
phonological	14.0	7.0	20.9	8.6	24.7	11.2	32.3	13.1
semantic	10.7	4.6	18.7	7.1	20.8	9.6	30.2	10.5
control	13.1	6.3	20.0	8.6	22.4	11.0	30.1	13.2
Part 2								
phonological	5.6	3.8	11.5	5.0	14.3	7.0	22.9	12.5
semantic	4.9	3.1	9.4	3.9	12.3	7.4	19.4	11.0
control	5.4	3.8	10.7	6.7	13.7	8.0	20.2	12.7

Word-reading test.

A 3 (Group: semantic vs. phonological vs. control) x 3 (Time: November, February, April) multivariate analysis of variance was performed on speed and accuracy of Word-reading tests 1, 2, and 3 with repeated measures on the time factor. No significant group by time interaction was found, *F*(24,132) = 1.02, *p* = .45. There was no main effect of group, *F* < 1. Results indicated a significant main effect of time, *F*(12,66)= 51.10, *p* < .001, *partial* η^2 = .90. Subsequent tests of within-subjects contrasts revealed that the effect of time on speed was significant for all three tests, both between November and February and between February and April (all *p*'s < .001). Subsequent tests of within-subjects contrasts on accuracy revealed a marginally significant effect for time on Word-reading test 1 between February and April,

$F(1,77) = 4.08$, $p = .05$, *partial* $\eta^2 = .05$, on Word-reading test 2 for both periods, respectively $F(1,77) = 122.80$, $p < .001$, *partial* $\eta^2 = .62$, and $F(1,77) = 5.33$, $p < .05$, *partial* $\eta^2 = .07$, and on Word-reading test 3 between November and February, $F(1,77) = 128.14$, $p < .001$, *partial* $\eta^2 = .63$.

Three-Minutes Test 1.

A 3 (Group: semantic vs. phonological vs. control) x 3 (Time: February, April, June) multivariate analysis of variance was performed on scores on the DTM1 with repeated measures on the time factor. No significant group by time interaction was found, $F < 1$. There was no main effect of group, $F < 1$. Results indicated a significant main effect of time, $F(2,71) = 89.59$, $p < .001$, *partial* $\eta^2 = .72$. Subsequent tests of within-subjects contrasts revealed that the effect of time was significant both between February and April, $F(1,72) = 137.79$, $p < .001$, *partial* $\eta^2 = .66$ and between April and June, $F(1,72) = 87.61$, $p < .001$, *partial* $\eta^2 = .55$.

Three-Minutes Test 2.

A 3 (Group: semantic vs. phonological vs. control) x 3 (Time: February, April, June) multivariate analysis of variance was performed on scores on the DTM2 with repeated measures on the time factor. No significant group by time interaction was found, $F < 1$. There was no main effect of group, $F < 1$. Results indicated a significant main effect of time, $F(2,71) = 69.33$, $p < .001$, *partial* $\eta^2 = .66$. Subsequent tests of within-subjects contrasts revealed that the effect of time was significant both between February and April, $F(1,72) = 120.84$, $p < .001$, *partial* $\eta^2 = .63$ and between April and June, $F(1,72) = 68.21$, $p < .001$, *partial* $\eta^2 = .49$.

Nonword-reading test 1.

A 3 (Group: semantic vs. phonological vs. control) x 3 (Time: February, April, June) multivariate analysis of variance was performed on scores on the nonword-reading test 1 with repeated measures on the time factor. No significant group by time interaction was found, $F < 1$. There was no main effect of group, $F < 1$. Results indicated a significant main effect of time, $F(2,71) = 96.41$, $p < .001$, *partial* $\eta^2 = .73$. Subsequent tests of within-subjects contrasts revealed that the effect of time was significant both between February and April, $F(1,72) = 177.29$, $p < .001$, *partial* $\eta^2 = .71$ and between April and June, $F(1,72) = 20.35$, $p < .001$, *partial* $\eta^2 = .22$.

Nonword-reading test 2.

A 3 (Group: semantic vs. phonological vs. control) x 3 (Time: February, April, June) multivariate analysis of variance was performed on scores on the nonword-reading test 2 with repeated measures on the time factor. No significant group by time interaction was found, $F < 1$. There was no main effect of group, $F < 1$. Results indicated a significant main effect of time, $F(2,70) = 79.81$, $p < .001$, *partial* $\eta^2 = .70$. Subsequent tests of within-subjects contrasts revealed that the effect of time was significant both between February and April, $F(1,71) = 99.66$, $p < .001$, *partial* $\eta^2 = .58$ and between April and June, $F(1,71) = 20.58$, $p < .001$, *partial* $\eta^2 = .23$.

AVI-test.

We only analyzed accuracy and speed on the AVI-test in June, because too many children failed on the test in February. A one-way ANOVA with speed and accuracy as dependent variables and group (semantic vs. phonological vs. control) as between-subjects factor resulted in a non-significant main effect of group, speed, $F(2,70)= 1.42$, $p > .05$, accuracy, $F < 1$. The mean overall speed was 164.7 sec. ($SD = 116$), the mean number of errors was 6.1 ($SD = 6.0$).

Receptive-vocabulary test

A 3 (Group: semantic vs. phonological vs. control) x 2 (Time: November, April) multivariate analysis of variance was performed on scores on the receptive-vocabulary test with repeated measures on the time factor. No significant group by time interaction was found, $F < 1$. There was no main effect of group, $F < 1$. Results indicated a significant main effect of time, $F(1,73)= 42.69$, $p < .001$, *partial* $\eta^2 = .37$: Accuracy in November ($M = 71.05$) was significantly lower than in April ($M = 77.26$).

Responsiveness to intervention

Finally, the percentage of children with poor word-decoding skills at the end of the training was calculated for the experimental groups and control group separately. The procedure was identical to that in Experiment 1. Recall, the cut-off score of the 25th percentile did not represent the lowest quartile of scores in the present sample. Rather, it was a norm-referenced cut-off, based on an appropriate comparison population. To examine the response to intervention, we calculated the percentage of children who performed below the 25th percentile on the DMT1 and DMT2 in June, relative to a comparison population. In the control group, 27.7% of all children scored below the 25th percentile on DMT1. In the experimental groups, 26.1% of the children in the phonological group and 29.2% of the children in the semantic group scored below the 25th percentile. On DMT2, these proportions are respectively

30.6%, 26.1%, and 29.1%. Note that these percentages are quite similar and smaller than those in Experiment 1. The latter might be caused by differences in test time, that is, May (Experiment 1) vs. June (Experiment 2).

Follow-up

A one-way ANOVA with group (semantic vs. phonological vs. control) as between-subjects factor and scores on the Three-Minutes test, Nonword-Reading Test, and AVI-test as dependent variables was performed at follow-up assessment (mid-Grade 2). No significant effect of group was found, all *F's* < 1.

In sum, across all outcome measures, no significant group by time interaction was found: The development of letter knowledge, reading skills, and vocabulary during Grade 1 was statistically equal for all groups. In addition, no main effect for group (treatment) was found, implying that the scores of the semantic group were similar to those of the phonological group and control group.

General discussion

The first aim of the study was to evaluate the efficacy of both a phonological-training program and a semantic-training program in a natural school setting. Results of Experiments 1 and 2 indicated that the progress in reading performance of both experimental groups at the end of the training was at least as strong as the reading development of the control group. Moreover, in Experiment 1, the semantic group outperformed the control group in reading CVC-words after seven months of reading intervention. In addition to reading instruction and practice in the classroom, both experimental groups trained their reading skills in an experimental computer training, whereas the majority of the control group practiced their reading skills with the help of a remedial teacher. Obviously, this difference in intervention method did not negatively affect reading outcomes of the experimental group. This finding implicates that computers are a useful supplemental tool for practicing reading skills, which is an encouraging result for teachers who deal with many children in need for reading support. The efficacy of computer programs in beginning readers was already demonstrated in a meta-study by Blok et al. (2002). In both experiments, we demonstrated the ecological validity of the training and showed that effective intervention can be implemented early in Grade 1.

The second aim of the study was to evaluate which of the two experimental training programs –either a semantically-oriented training or a phonologically-oriented training - was the most effective one for phonological skills, reading new words, nonwords, and text. In Experiment 1, both experimental groups revealed a similar developmental pattern in phonological awareness. At the beginning of Grade 1, segmentation and blending skills were insufficient, but during the first school year, performance on these tests reached its ceiling. Previous research has demonstrated that phonological awareness develops quickly once literacy instruction begins, especially in a relatively transparent orthography like Dutch (e.g., Anthony & Francis, 2005; Wesseling & Reitsma, 2001). Results of Experiment 1 demonstrated that even poor beginning readers made considerable progress in their phonological skills once reading instruction had started.

Interestingly, after four months of training, children from the semantic group outperformed those from the phonological group in letter knowledge, reading CVC-words and reading CVCC/CCVC-words. During the first part of the intervention (from November to February), the majority of the exercises concerned practicing graphemes and CVC-words. These results demonstrated that readers benefited most from a training if these types of stimuli were presented in a semantic context during the training. Performance differences between both groups were not found in reading nonwords. Obviously, training words in semantic context does not facilitate decoding nonwords to a higher or lesser degree than a phonological training does. At the end of the training, children from the phonological group and the semantic group performed statistically equally on all tests. During the second part of the training (from February to May), the majority of exercises concerned reading CVC-words and CCVC/CVCC-words. During this period, all graphemes were mastered by most of the children and skills that had been learned previously had to be automatized. The results indicated that the superiority of the semantic training after four months of reading intervention disappeared at the end of the training. It seems that a semantic training has additional benefits over a phonological training in the initial phase of learning to read. After that, the reading gains of both a semantic program and a phonological program are similar. Results of the second experiment validate the latter result: Repeated-measures analyses yielded similar outcomes for the semantic training and the phonological training across several reading measures. Thus, in the end, semantic instruction in a training program did not increase reading skills to a higher degree than phonological instruction did. Stated differently, phonological instruction was not superior to semantic instruction. This result implies that the emphasis on phonological skills in a training program is not necessary if the training is additional to phonics instruction in the classroom: Semantic exercises were as effective as phonological exercises and might even be more motivating for children.

When comparing the two training methods, it is important to note that the labels of the experimental programs (phonological and semantic) may be misleading. In both training programs, it is assumed that reading will activate phonology, semantics, and orthography. However, the types of exercises in the phonological program biased attention to the phonological information, whereas the exercises in the semantic program required to pay relatively more attention to semantic information. In the semantic-training program, guessing strategies could not be applied to provide a correct response and the training was not aimed at increasing use of contextual cues. Rather, it is assumed that semantic information is activated automatically, and children in the semantic training had to focus on this source of information to a larger degree than focusing on phonological information.

To summarize, a word-identification training focusing on the semantics of words seems to be an effective intervention for poor beginning readers. The question is, in what way could the semantically-oriented training have been beneficial? The underlying assumption in the semantic training is that semantic information is retrieved in a cascaded way; even the slightest activation of orthographic or phonological information will flow to the semantic level. Thus, the meaning of a word can be activated before a word is recognized. This view is accepted by many other researchers (e.g., Balota, Ferraro, & Connor, 1991; Coltheart et al., 2001). Direct evidence for cascaded processing was provided by Rodd (2004) and Bowers (2005). A recurrent model of word identification with bi-directional activation between semantic, orthographic, and phonological units (e.g., Van Orden, Pennington, & Stone, 1990) offers a comprehensive explanation for this process. If a letter sting is presented, activation will immediately flow from orthographic units to phonemic units and semantic units, which in turn will activate each other and provide feedback information to the grapheme nodes. In poor readers, the correspondence between the grapheme units and phoneme units is assumed to be slow, error-prone, or damaged. This weakness may be compensated by activation that is fed back from the semantic units to the orthographic units. Thus, in the semantic training, we assume an interplay between all three types of information (orthographic, phonemic, and semantic), which will enhance word-identification skills. It is not clear whether semantic support strengthens grapheme-phoneme correspondences. According to the Phonological Coherence model, grapheme-phoneme linkages in poor readers lack complete reciprocity and therefore, these connections cannot be improved at the grapheme-phoneme level. Norbury and Chiat (2000), however, suggest that semantic instruction might indirectly strengthening the links between graphemes and phonemes. Finally, due to the interplay of different types of activation (semantic, orthographic, and phonological) and different skills (e.g., phonological skills and semantic skills), it is not possible to establish the unique independent contribution of semantics in the study.

The third aim of the first study was to examine the responsiveness to intervention. The percentage of children performing below the 25th percentile on standardized reading tests at the end of Grade 1 ranged from 32% to 56% in Experiment 1 and from 26% to 31% in Experiment 2. For these children, reading intervention could not increase their word-identification skills to satisfying levels. In the literature, a failure rate of 30% to 40% is often cited (Torgesen, 2000). The overall number of slow responders would probably have been smaller if the intervention had been more intensive. In the experimental training, children trained three times a week during 10 to 15 minutes and even this intensity was not realized in some schools.

Finally, a few remarks and recommendations deserve some attention. First, the present experiments only included children who were at risk for reading disabilities because of poor word-identification skills at the beginning of Grade 1. It would be interesting to study the effects of a semantically-oriented program in average or good readers. To our knowledge, no studies have been carried out so far to test the effects of a semantic training in beginning readers with adequate word-identification skills.

Second, this study was conducted in Dutch, a relatively transparent language. Most studies concerning the role of semantics are accomplished in English, a relatively opaque language, including many irregular words and exception words. In contrast, the spelling-sound consistency in Dutch is relatively large. De Groot (1989) demonstrated significant, though small effects of imageability in Dutch word recognition. In addition, with respect to Turkish, a very transparent orthography, Raman and Baluch (2001) reported no significant effect of imageability in previously skilled Turkish readers. Gijsel and Bosman (submitted) studied the effect of imageability in Dutch children across Grade 2 to Grade 6 and failed to find an imageability effect in naming results of young readers. Thus, in a deep orthography, the impact of a semantic training might even be more pronounced than in a transparent orthography like Dutch.

To conclude, the present findings suggest that a semantic-training program is effective for improving word-identification skills of poor beginning readers, if the program is additional to phonics instruction in the classroom. These findings are consistent with connectionist theories on word recognition in which bi-directional activation flows between semantic, orthographic, and phonological units.

REFERENCES

Abad, M. J. F., Noguera, C., & Ortells, J. J. (2003). Influence of prime-target relationship on semantic priming effects from words in a lexical-decision task. *Acta Psychologica, 113,* 283-295.

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14,* 255-259.

Assink, E. M., Van Bergen, F., Van Teeseling, H., & Knuijt, P. P. (2004). Semantic priming effectds in normal versus poor readers. *Journal of Genetic Psychology, 165, 67-79.*

Baluch, B., & Besner, D. (2001). Basic processes in reading: semantics affects speeded naming of high-frequency words in an alphabetic script. *Canadian Journal of Experimental Psychology, 55(1),* 63-69.

Berends, I. E., & Reitsma, P. (2006). Addressing semantics promotes the development of reading fluency. *Applied Psycholinguistics, 27,* 247-265.

Berent, I., & Perfetti, C. A. (1995). A Rose is a REEZ: The two-cycles model of phonology assembly in reading English. *Psychological Review, 102,* 146-184.

Berninger, V. W., Abbott, R. D., Vermeulen, K., Ogiers, S., Brooksher, R., Zook, D., & Lemos, Z. (2002). Comparison of faster and slower responders to early intervention in reading: Differentiation features of their language profiles. *Learning Disability Quarterly, 25, 59-76.*

Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research, 72,* 101-130.

Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 63-85.

Bosman, A. M. T., & van Hell, J. G. (2002). Orthography, phonology, and semantics: Concerted action in word perception. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 165-187). Amsterdam: John Benjamins.

Bosman, A. M. T., & Van Orden, G. C. (1997). Why spelling is more difficult than reading. In C. A. Perfetti, L. Rieben, Fayol, M. (Eds.), *Learning to spell: Research, theory, and practice across languages* (p. 173-194). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bowey, J. A. (1985). Contextual facilitation in children's oral reading in relation to grade and decoding skills. *Journal of Experimental Child Psychology, 40,* 23-48.

Bus, A. G., & Van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91,* 403-414.

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review, 4*, 226-231.

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

Farrar, W. T., & Van Orden, G. C. (2001). Errors as multistable response options. *Nonlinear Dynamics, Psychology, and Life Sciences, 5,* 223-265.

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

Frost, R. (1995). Phonological computation and missing vowels: Mapping lexical involvement in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 398-408.

Gonzalez, M. R. O., Espinel, A. I. G., & Rosquete, R. G. (2002). Remedial interventions for children with reading disabilities: Speech perception – An effective component in phonological training? *Journal of Learning Disabilities, 35,* 334-342.

de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 824-845.

Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language, 36, 505-529.*

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120,* 358-394.

Kim, Y. H., & Goetz, E. T. (1994). Context effects on word recognition and reading comprehension of poor and good readers: A test of the interactive-compensatory hypothesis. *Reading Research Quarterly, 29,* 179-187.

McGuinness, C., McGuinness, D., & McGuinness, G. (1996). Phonographix™: A new method for remediating reading difficulties. *Annals of Dyslexia, 46*, 73-96.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227-234.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.

Nicolson, R. I., Fawcett, A. J., Moss, H., Nicolson, M. K., & Reason, R. (1999). Early reading intervention can be effective and cost-effective. *British Journal of Educational Psychology, 69, 47-62.*

Norbury, C. F., & Chiat, S. (2000). Semantic intervention to support word recognition: a single-case study. *Child Language Teaching and Therapy, 16,* 141-163.

Pecher, D. (2001). Perception is a two-way junction: Feedback semantics in word recognition. *Psychonomic Bulletin & Review, 8, 545-551.*

Pexman, P. M., & Lupker, S. J. (1999). Ambiguity and visual word recognition: Can feedback explain both homophone and polysemy effects? *Canadian Journal of Experimental Psychology, 53,* 323-334.

Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: Number-of-features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review, 9, 542-549.*

Pring, L. & Snowling, M. (1986). Developmental changes in word recognition: An information-processing account. *The Quarterly Journal of Experimental Psychology, 38A,* 395-418.

Raman, I., & Baluch, B. (2001). Semantic effects as a function of reading skill in word naming of a transparent orthography. *Reading and Writing, 14,* 599-614.

Seidenberg, M., & McClelland, J. (1989). A distributed developmental model for word recognition and naming. *Psychological Review, 96, 523-568.*

Spruyt, A., Hermans, D., De Houwer, J., & Eelen, P. (2004). Automatic non-associative semantic priming: episodic affective priming of naming responses. *Acta Psychologica, 116,* 39-54.

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: an individual differences analysis. *Canadian Journal of experimental psychology, 53,* 347-359.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of experimental psychology: Learning, Memory, and Cognition, 21,* 1140-1154.

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice, 15,* 55-64.

Van Leerdam, M. (1996). *Pseudowoordentest* [Nonword-reading test]. Unpublished materials.

Van Orden, G. C., Bosman, A. M. T., Goldinger, S. D., & Farrar, W. T. (1997). A recurrent network account of reading, spelling, and dyslexia. In J. W. Donahoe & V. Packard Dorsel (Eds.), *Neural network models of cognition: A biobehavioral foundation* (pp. 522-538). Amsterdam, the Netherlands: Elsevier Science Publishers.

Van Orden, G.C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97,* 488-522.

Verhoeven, L. (1992). *Drie-Minuten-Toets* [Three-minutes test]. Arnhem: CITO.

Verhoeven, L. (1991). *Fonemendictee* [Phoneme Test] Arnhem: CITO.

Verhoeven, L. (1992). *Grafementoets* [Grapheme Test]. Arhem: CITO.

Verhoeven, L., & Vermeer, A. (2001). *Taaltoets Alle Kinderen.* [Language Test All Children]. Arhem: CITO.

Visser, J., van Laarhoven, A., & ter Beek , A. (1998). *AVI-toetspakket* [AVI-test]. 's Hertogenbosch : KPC Groep.

Wentink, H., & Verhoeven, L. (2001). *Protocol voor leesproblemen en Dyslexie.* Nijmegen: Expertisecentrum Nederlands.

Wesseling, R., & Reitsma, P. (2000). The transient role of explicit phonological recoding for reading acquisition. *Reading and Writing, 13,* 313-336.

APPENDIX

Description of the experimental training programs

Difficulty level	semantic group	phonological group
Graphemes	right or wrong	right or wrong
	letter worm	letter worm
	matching	matching
	phoneme in words	phoneme in words
Words	right or wrong	right or wrong
	antonyms	word recognition
	semantic decision	lexical decision
	matching	balloons
	odd one out	odd one out
	the lost grapheme	the lost grapheme
	semantic categorization	phonological categorization
Sentences	Word hunt	Word hunt
	Strange story	Lexical check
	Right or wrong	Right or wrong

Graphemes

Right or wrong.

A grapheme was presented visually on the screen. Subsequently, a phoneme was presented verbally. In some trials, these stimuli corresponded, whereas in other trials there was an incorrect mapping between the grapheme and the phoneme. The child had to decide whether the phoneme and grapheme corresponded or not. The number of stimuli per exercise was 10.

Letter worm.

Ten graphemes were visually presented. Then, from one of these graphemes, the corresponding phoneme was presented verbally. The child had to select the right grapheme. The number of stimuli per exercise was 10.

Matching.

Ten graphemes were visually presented in two columns on the left side of the screen. On the right-hand of the screen, 10 loudspeakers were depicted. By moving the mouse across one of the loudspeakers, a phoneme could be heard. One of the graphemes on the left-hand of the screen was marked. The correct phoneme behind the loudspeakers had to be identified. The number of stimuli per exercise was 10.

Phoneme in words.

Nine CVC-words were visually presented in colored circles at random position. Then, a phoneme was presented verbally. The child had to select each word that included the requested phoneme. The number of words to be selected varied from 3 to 5. The number of stimuli per exercise was 5.

Words

Semantic training

Right or wrong.

A picture was presented, together with a word that represented the picture. In some trials, the spelling of the word was correct (e.g., a picture of a book was presented with the word ‘book’), whereas in others, the spelling of the word was incorrect (for example, ‘hook’). If the spelling was incorrect, the word (distracter) was perceptually confusing with the target. The child had to decide whether the spelling of the word was correct (corresponding to the picture) or not. The number of stimuli per exercise was 10.

Antonyms.

On the left-hand of the screen, 5 to 10 target words were presented in one column. On the right-hand, an equal number of words were presented in a second column. These words were the antonym of the targets. One word of the first column was marked. Then, the child had to select the corresponding antonym in the second column.

Semantic decision.

The name of a semantic category was verbally presented (e.g., ‘food’, ‘animals’, or ‘furniture’). Then, a word was presented visually on the screen. In some trials, this word was an example of the semantic category that was presented (for example, ‘deer’ belonged to animals), whereas in others this was not the case (for example, ‘beer’ and animals). The child had to decide whether the word belonged to the presented semantic category or not. The number of stimuli per exercise was 10.

Matching.

A picture was presented on the screen with 4 words. One of these words corresponded to the picture, the other 3 words were perceptually confusing distracters. The child had to select the correct word. The number of targets per exercise was 10.

Odd one out.

Four words were presented visually in one row. Three of these words were semantically related to each other. The words were related by means of (examples of) a semantic category (for example, ‘trousers’, ‘skirt’, and ‘jacket’) or by means of a script (for example, ‘sea’, ‘boat’, ‘sail’). One word was not semantically related to the other words. The child had to select the semantically unrelated word (odd one out). The number of targets per exercise was 10.

The lost grapheme.

A picture was presented on the screen. A word that corresponded to this picture was visually presented. However, one (or two) grapheme(s) of the word was (were) excluded. The position of this ‘lost grapheme’ varied from initial, medial to final. In addition to the incomplete string of graphemes (partial word), 4 graphemes were presented. One of these graphemes was the correct one that fitted the string. The other three graphemes were distracters. The child had to select the right grapheme. The number of targets per exercise was 10.

Semantic categorization.

Two or three pictures were presented, each of them representing a semantic category (e.g., buildings, cloth) or a script (e.g., school, garden). Together with these pictures, 5 to 12 words belonging to one of the semantic categories or scripts were presented. The child had to drag each word to the right semantic category/script. This procedure was repeated three times with different stimuli.

Phonological training

Right or wrong.

A word was presented visually. Subsequently, another word was presented verbally. This word at times corresponded with the string of graphemes, at other times this was not the case. The spelling of the verbal distracter was perceptually confusing with the target word. The child had to decide whether the verbally-presented word was the correct one. The number of stimuli per exercise was 10.

Word recognition.

Two words were presented visually. Subsequently, one of the words was presented verbally. The child had to select the correct word that corresponded to the verbal stimulus. The number of stimuli per exercise was 10.

Lexical decision.

A string of letters was presented visually. In some trials, this letter string constituted a word, whereas in others, the letter string constituted a pseudoword. The child had to decide whether the presented stimulus was a word or not. Pseudowords were defined as letter strings that were orthographically and phonologically legal in the Dutch orthography. The number of stimuli per exercise was 10.

Balloons.

Ten words were presented visually at random positions on the screen. Subsequently, five of these words were presented verbally. The other five words were perceptual distracters. After a word was presented verbally, the child had to select the correct visual representation. This procedure was repeated five times with different stimuli.

Odd one out.

Four words were presented visually in one row. Three of these words were phonologically related to each other, most of the time by means of end rime (for example, 'snail', 'trail', and 'tail' or 'fear', 'beer', and 'near'). One word did not share the same rime. The child had to select this incorrect word (odd one out). The number of targets per exercise was 10.

The lost grapheme.

A string of letters was presented visually. To form a word, one grapheme was missing in the string. The position of this 'lost grapheme' varied from initial, medial to final. In addition to the incomplete string of letters, four graphemes were presented. One of these graphemes was the correct one that fitted the string to create a word. The other three graphemes were distracters. The child had to select the right grapheme. The number of targets per exercise was 10.

Phonological categorization.

Two or three distinct categories were presented visually, each of them representing a phonological (rime) category. If a child selected a domain (by mouse click), a word was presented verbally. This word was the target rime. Together with these domains and target rimes, 5 to 12 words belonging to one of the phonological categories were presented. A word belonged to a particular phonological category, if it shared the end rime with the word that was presented verbally in that category. For example, the word 'cat' was the correct category for words that rime with cat, like 'hat', or 'rat'. The child had to drag each word to the right phonological domain. This procedure was repeated for three times with different stimuli.

Sentences

Semantic training

Word hunt.

The name of a semantic category was presented verbally (e.g., 'food'), together with a picture of that category. Subsequently, a story consisting of 4 or 5 short sentences was presented. The child had to select those words in the story, that were members of the semantic category (for example, 'carrots'). The number of words to be selected ranged from 2 to 5.

Strange story.

Three to five short sentences were presented visually (one sentence per line). These sentences formed a story. However, the sentences were placed in a wrong order. The child had to arrange the sentences in order to create a logical story.

Right or wrong.

A sentence was presented visually. Subsequently, a question about this sentence was presented on the screen. This question concerned the contents of the sentence and made use of synonyms and antonyms of words in the stimulus. For example: the stimulus was ‘Simon doesn’t like beans’. Question: ‘Is Simon fond of beans?’ The question could always be answered with ‘yes’ or ‘no’.

Phonological training

Word hunt.

The name of a semantic category was presented verbally (e.g., ‘food’), together with a picture of that category. Subsequently, a story consisting of 4 or 5 short sentences was presented. However, these sentences were presented in the wrong order to prevent context use as much as possible. The child had to select those words in the sentences, that were members of the semantic category (e.g., ‘carrots’). The number of words to be selected ranged from 2 to 5.

Lexical check.

Three to five short sentences that were semantically unrelated were presented on the screen. The number of words in a sentence ranged from 4 to 8. Some sentences comprised a misspelled word; in other sentences, no spelling errors occurred. The child had to decide whether the sentence was spelled correctly or not. Additionally, if an error had been noticed, the child had to select the misspelled word.

Right or wrong.

A sentence was presented visually. Subsequently, a question about this sentence was presented visually. This question concerned the contents of the sentence and either made use of exactly the same words as in the stimulus, or included perceptual similar words. For example,: the stimulus was ‘Simon loves his bear. Question: ‘Does Simon love his bear?’ (identical words) or ‘Does Simon love his dear?’ (perceptual similar word). The question could always be answered with ‘yes’ or ‘no’.

Chapter 4: Semantic skills in relation to word decoding in primary grades[1]

Abstract

In the present study, the relationship between semantic skills and word-decoding skills was examined. In Experiment 1, 99 first graders participated in a semantic-categorization task, a word-association task and a word-decoding test. Results revealed no differences between poor decoders and good decoders in word-association skills, whereas poor readers were more error prone in a semantic-categorization test. In Experiment 2, children from Grades 1 to 6 participated in two types of semantic-categorization tasks and a word-decoding test. The categorization tasks were performed in different modalities: Concepts were presented by means of printed words, spoken words, or pictures. Response options were always pictures. It turned out that poor readers showed longer reaction times on both types of categorization tasks than average readers and good readers. This difference did neither vary across grades nor across different modalities of the stimuli. The results suggest that semantic skills are related to reading difficulties.

---

[1] An adapted version of this chapter has been submitted for publication as: Gijsel, M. A. R., Ormel, E. A., Hermans, D., Bosman, A. M. T., & Verhoeven, L. Semantic-categorization skills in relation to word decoding in primary grades. We thank Anke van der Velden, Kim van Pol, Marjolein Wennekers, Maartje van Ijzendoorn, Jikke Planting, Annelies Leechburg-Auwers, and Corina Michielsen who collected the data. We are also grateful to all children who participated in the study and to the teachers who gave their permission to run the experiments.

Introduction

It is generally assumed that reading involves phonological, orthographic, and semantic activation. The empirical evidence with respect to the role of orthography and phonology in reading has received a great deal of attention (e.g., Bradley & Bryant, 1983; Bryant, MacLean, Bradley, & Crossland, 1990; Mann & Liberman, 1984; Wagner & Torgesen, 1987). The role of semantics in reading and learning to read, on the other hand, is recognized, but has received less attention. This study investigates the relationship between semantic skills and word-decoding skills in Dutch students.

Evidence for a positive relationship between phonological skills and word decoding has been provided predominantly by training and intervention studies (see Ehri et al., 2001 for a review), and concurrent or longitudinal prediction studies (e.g., Elbro, Borstrøm, & Petersen, 1998; O'Connor & Jenkins, 1999). In general, these studies have demonstrated that training phonological skills (e.g., phonemic awareness) improves reading skills and that phonological abilities turn out to be good predictors of reading performance. The strength of the connection between phonological skills and reading skills was found to be dependent on developmental time, that is, the time at which moment both skills were measured (e.g. Wagner et al., 1997; Wesseling & Reitsma, 2000), and on the kind of phonological-processing skills being involved: phonological awareness, phonological memory, or phonological-naming skills. In the majority of studies, phonological awareness turned out to be the best predictor of reading performance. Moreover, in a phonological-awareness task, correlations are higher if the test involves phonemes, rather than syllables (Wagner & Torgesen, 1987). Thus, although the strength of the connection between phonological skills and reading skills depends on a variety of factors, previous research has shown that poor phonological skills can be seen as a major cause in the development of reading difficulties.

It is, however, unlikely to assume that phonological difficulties are the sole cause for low reading performance. Because reading involves orthographic, phonological, and semantic activation, semantic skills may play a role as well in reading performance. Unfortunately, semantic skills have been defined rather loosely so far. It includes the ability to recognize and define words (receptive and expressive vocabulary tests), to generate words in an association task (semantic fluency), to make synonym judgment, to detect common features in semantic concepts, to generate exemplars of a semantic category, and to verify whether a word belongs to a certain semantic category (categorization test). In addition, semantic skills also refer to higher levels of processing, that is, the ability to predict the plot of a story, extract the meaning of a story, and to discuss a book after reading it. The highly variable operationalization of semantic skills demonstrates the versatility of the concept and the difficulty in making statements about its relationship with reading skill.

Research in the area of semantic skills of poor readers has generated mixed results. Several researchers have studied the longitudinal relationship between early language skills and later reading development (e.g., Catts, Fey, Zhang, & Tomblin, 1999; Frost, Madsbjerg, Niedersøe, Olofsson, & Møller Sørensen, 2005; Gillon & Dodd, 1994; Scarborough, 1989, 1990, 1991). These studies have demonstrated that semantic skills (e.g., vocabulary, language comprehension, and narrative skills) and syntactic skills may play a role in reading difficulties. Other researchers have tried to establish to what extent semantic-categorization skills and word-association skills are related to word decoding. Semantic-categorization tasks and word-association tasks have been used frequently to obtain insight into the organization of the semantic-memory system. We will now turn to a brief overview of studies that included such tasks.

Word association. In a word-association task, the child is generally presented with a word and is subsequently asked to generate all the words he or she can think of in a certain amount of time (continued free-word association) or to produce the first word that comes to mind (discrete free-word association). Responses have been analysed in both a quantitative (number of responses) and qualitative way (type of responses). Nation and Snowling (1998, 2004) included in their study a slightly different version of the word-association task, referred to as semantic-fluency task. Children had to generate as many exemplars of category members as possible when provided a category name, and performance was related with reading outcomes. Nation and Snowling (1998, Experiment 2) showed that children with poor reading-comprehension skills produced significantly fewer words than children with good comprehension skills. They also investigated the relationship with word recognition and found that oral-language skills (including semantic fluency) accounted for a significant proportion of variance in word-decoding skills. Because three semantic variables were grouped into one variable, it was impossible to investigate the unique contribution of semantic fluency.

With respect to the type of responses, a distinction has been made between syntagmatic responses, referring to all responses that indicate properties or descriptions of a stimulus, and paradigmatic responses, referring to responses of the same word class (e.g., synonyms and category names). It is assumed by some researchers (e.g., Cronin et al., 1986) that word associations shift from syntagmatic to paradigmatic responses with increasing reading skills. In short, the number and type of responses in a word-association task may reflect the underlying structure of the semantic-memory system and this structure might be different for poor readers and good readers.

Semantic categorization. In a semantic-categorization test, participants are generally asked to relate a category name (e.g., animals) to semantic exemplars that belong to the category (e.g., dog, cat). In a few studies, a positive relationship was found between categorization skills and word identification in primary grades (e.g., Ben-Dror, Bentin & Frost, 1995; Howell & Manis, 1986; Vellutino, Scanlon, & Spearing, 1995). Howell and Manis (1986) presented participants with pictures and printed words that were exemplars of one of four semantic categories. Participants had to decide whether the stimulus (e.g., dog) was presented with either the correct superordinate label (i.e., animal) or the correct basic label (i.e., dog). Performance of participants differed as a function of age (Grades 2 and 3 vs. Grades 5 and 6) and word-identification skills (normal readers versus poor readers). Poor readers were significantly slower in making decisions than controls and this difference was larger in younger readers than in older readers. The slower semantic performance of poor readers was apparent in both the pictures and printed words, suggesting that semantic skills are not specific to written words. Ben-Dror et al. (1995) addressed the relationship between semantic categorization, morphological skills, phonological skills, and reading in Hebrew. Participants were both normal readers and children with reading difficulties from Grade 5 and a control group of normal readers from Grade 3, matched on vocabulary. Results demonstrated that poor readers performed worse (more errors and longer reaction times) than normal readers at the same age level in the semantic-categorization test. These studies seem to show that poor semantic skills, as measured by means of a semantic-categorization task, are associated with poor word-decoding skills.

However, some researchers failed to demonstrate a relationship between semantic-categorization skills and word decoding (e.g., Silva-Pereyra et al., 2003; Vellutino, Scanlon, & Tanzman, 1990). Vellutino et al. (1990) presented both poor readers and normal readers from Grades 1, 2, 4, and 6 triads of words and participants were asked to ‘put together that go together’. Words were semantically/syntactically related or phonologically/orthographically related. Performance differences between reading groups were attributed to difficulties in identifying all words in the sets: After the analysis was corrected for the number of items identified correctly, poor and normal decoders did not differ with respect to semantic categorizations. Silva-Pereyra et al. (2003) also challenged the claim that poor readers exhibit semantic deficits and attributed poor readers’ lower performance in semantic tasks to deficiencies in the processing of words, rather than to poor semantic skills per se. In their experiment, poor readers differed from control readers in a word-categorization task, but not in a picture-categorization task.

To summarize, the evidence for a positive relationship between semantic skills and word decoding is not fully convincing. Moreover, few studies have been carried out with beginning readers. Therefore, in Experiment 1, we performed a study with readers in Grade 1, who differed in word-decoding skills. All children performed a semantic-categorization task and a word-association task to evaluate semantic skills. In the association task, children were asked to generate as many words as possible to a given stimulus. We manipulated the imageability ratings of the words, to examine differences in responses to low-imageability words and high-imageability words. Imageability is defined as the extent to which the referent of the word evokes a mental image (de Groot, 1989). Results of studies on reading (e.g., Cortese, Simpson, & Woolsey, 1997; Strain, Patterson, and Seidenberg, 1995) and memory (e.g., Vellutino et al., 1995) have demonstrated a processing advantage for high-imageability words over low-imageability words. Therefore, we expected more associative responses to high-imageability words. In the semantic-categorization task, children were presented a target picture and were asked to select those pictures that belonged to the same semantic category or script as the target.

In short, in Experiment 1, we were interested whether (a) poor readers generated fewer associations and fewer typical associations in a word-association task than average or good readers did, (b) poor readers demonstrated longer reaction times and more errors in a semantic-categorization task and (c) semantic skills contribute to the prediction of decoding speed. The reason for carrying out Experiment 2 will be discussed later on.

Experiment 1

Method

Participants

In this study, 99 students, 53 boys (53.5%) and 46 girls (46.5%), from four regular-primary schools participated. These schools (three rural schools and one urban school) were situated in two different regions in the Netherlands, that is, Noord-Brabant and Gelderland. All children were recruited from Grade 1. At the first test time, the mean age was 7.1 years (*SD* = .41). The majority of children (83.8%) were native Dutch speakers.

Materials

Reading skills

Three-Minutes Test [Drie-Minuten Toets] by Verhoeven (1992).

This standardized word-reading test consists of three cards, each containing five columns of words with specific word structures. In the current study, only Card 1 and Card 2 were administered. Card 1 consists of 150 CVC-, CV-, or VC-words (C is consonant, V is vowel). Card 2 consists of 150 CCVC-, (C)CVCC-, CCCVC-, and CVCCCC-words. The child was asked to read as many words as possible and as accurately as possible within one minute. The score is the number of items read correctly.

Klepel by Van den Bos, Lutje Spelberg, Scheepstra, & de Vries (1994) .

This standardized nonword-reading test consists of 116 pseudowords of increasing difficulty. The child was asked to read the pseudowords as fast and as accurately as possible within two minutes. The score is the number of items read correctly within two minutes.

Semantic skills

Semantic-categorization task.

This test was developed for the present study and addressed the ability to categorize objects. The test consisted of five experimental trials. In each trial, a target picture was presented at the top of a sheet of paper. In addition, nine pictures were presented beneath the target picture in 3x3 columns. The child had to select as fast and as correctly as possible three pictures that belonged to the same semantic category or script as the target. Only common categories were used (see Appendix A for all stimuli in the categorization task). Children marked their responses by means of pencil marks. Finally, the child was asked to name the semantic category (‘why do these pictures belong together?’). With respect to the analyses, we calculated the number of pictures correctly identified (item accuracy), the number of categories correctly named (description accuracy), and the time to complete the task (speed).

Word-association task.

This test was developed for the current study and addressed the ability of fluent association. The experimenter verbally presented 10 words, one-by-one, and the child had to produce as many words as possible that came to mind within 30 seconds. All words were nouns and were selected from the word familiarity ratings or frequency counts by Kohnstamm, Schaerlaekens, de Vries, Akkerhuis, and Froonincksx (1981). According to this source, 6-year-old students are familiar with the selected words (criterion ≥ 80%). Half of the items were high-imageability words (HI-words) and half of the items were low-imageability words (LI-words). The imageability ratings were based on the ratings from van Loon-Vervoorn (1985), which were collected by means of a seven-point scale (1= low imageability; 7= high imageability). Words with imageability ratings between 6.0 and 7.0 were classified as HI-words; words with ratings between 2.5 and 4.0 were classified as LI-words. Words with imageability ratings beneath 2.5 were excluded; these items probably have very few associates for 7-year-old children. The words were of different semantic categories and scripts. Finally, the orthographic structure of the HI-words was matched with the orthographic structure of the LI-words. Word characteristics are shown in Table 1.

Table 1. Descriptives of the Words in the Association Task in Experiment 1

	Imageability rating	Word frequency (Grades 1, and 2)	Familiarity rating (%)
HI-words			
soldier [soldaat]	6.53	24	90
pastry [gebak]	6.37	14	98
closet [kastje]	6.60	124	92
work-book [schrift]	5.97	65	98
school [school]	6.39	382	100
LI-words			
story [verhaal]	3.93	267	100
secret [geheim]	3.27	25	98
trick [kunstje]	3.83	19	94
fright [schrik]	3.98	58	96
fart [scheet]	4.03	0	82

With respect to the analyses, we calculated the number of adequate generated associations (i.e., repetitions excluded) and the typicality of the first association, that is, the proportion of children that produced the same association on the first trial. Thus, a high percentage of typical responses indicated that the child produced an associate that was generated by many others.

Procedure

All tests were administered individually and the order of the tests varied among participants. The children were classified as poor readers, average readers or good readers. We defined poor readers as students with poor word-recognition skills, that is, reading scores on the DMT1 are below the 25th percentile. Average readers performed between the 25th and 75th percentile. Good readers were defined as students with good word-recognition skills, that is, reading scores on DMT1 are above the 75th percentile.

Results

Three children were excluded from the analyses because of missing data. Table 2 shows the results of the reading tests for poor readers, average readers, and good readers. Note that the percentages of errors on DMT2 and Klepel are relatively high. This high number of errors is probably caused by the early assessment: At the time both tests were performed, children had not yet mastered the skills that are assessed by these tests.

Table 2. Means and Standard Deviations of the Naming Times per Word (in msec) and Percentages of Errors[2] in Word Reading and Pseudoword Reading in Experiment 1

	Reader group							
	Poor (*n* = 14)		Average (*n* = 48)		Good (*n* = 34)		Total (*n* = 96)	
Reaction times								
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Word reading								
DMT 1	3120	805	1681	313	942	144	1630	801
DMT 2	6550	2510	3464	1059	1547	533	3235	2045
Pseudoword reading								
Klepel	6446	2540	3866	1623	1976	661	3573	2123
% errors								
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Word reading								
DMT 1	19.7	13.9	11.4	8.8	4.9	4.1	10.3	9.7
DMT 2	28.0	19.3	22.7	14.5	10.5	10.3	19.2	15.4
Pseudoword reading								
Klepel	49.6	17.4	44.6	14.3	26.2	10.9	38.8	16.6

To ensure that the reading scores of the poor, average, and good readers differed significantly from each other, we performed a one-way ANOVA with reaction times and percentages of errors on the DMT1 as dependent variables and reader group (poor vs. average vs. good) as between-subjects factor. The main effect of reader group was significant, error percentages, $F(2,93) = 15.56$, $p < .001$; reaction times, $F(2,93) = 160.41$, $p < .001$. A post-hoc Bonferroni test revealed that all groups differed significantly from each other, $p < .01$.

Word-Association skills

The first question was whether poor readers generated fewer associations and fewer typical associations in a word-association task than average or good readers did. Mean scores on the Association Test are shown in Table 3.

Both number of responses and typicality of responses are summarized for poor readers, average readers, and good readers. Mean number of associations and mean percentage of typical responses were submitted to separate ANOVA's. In all analyses, significant effects are based on a .05 alpha level.

---

[2] The commonly used outcome measure in the DMT and Klepel is a score which reflects both speed and accuracy. In this study, however, the mean naming times per word and percentages of errors were calculated separately.

Table 3. Mean Number of Meaningful Associations and Standard Deviations and Typicality of Responses in the Association Test

	Reader group							
	Poor (*n* = 14)		Average (*n* = 47)		Good (*n* = 33)		Total (*n* = 94)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Number								
LI-words	2.63	1.04	2.69	1.03	2.22	1.34	2.52	1.16
HI-words	3.56	1.53	3.68	1.12	3.65	1.41	3.65	1.28
Typicality								
LI-words	10.54	4.49	9.62	4.13	9.38	4.88	9.67	4.42
HI-words	17.14	8.14	20.89	7.40	20.27	8.37	20.12	7.87

A 3 (Reader group: poor, average, good) x 2 (Imageability: low, high) ANOVA was performed on typicality with reader group as between-subjects factor and imageability as within-subjects factors. Recall, typicality reflected the number of non-unique responses. The main effect of imageability was significant, $F(1,92) = 105.48$, *partial* $\eta^2 = .53$; the number of typical responses was larger in high-imageability words than in low-imageability words. Neither the main effect of reader group, nor the interaction between reader group and imageability reached a significant level. Thus, poor readers, average readers, and good readers generated identical proportions of non-unique responses.

A 3 (Reader group: poor, average, good) x 2 (Imageability: low, high) ANOVA was performed on number of associations with reader group as between-subjects factor and imageability as within-subjects factor. The main effect of imageability was significant, $F(1,91) = 25.91$, *partial* $\eta^2 = .22$; the mean number of associations was larger in high-imageability words than in low-imageability words. Neither the main effect of reader group, nor the interaction between reader group and imageability was significant. Thus, poor readers, average reader, and good readers generated a similar number of associations.

Semantic-categorization skills

The second question was whether poor readers demonstrated longer reaction times and more errors in a semantic-categorization task. In the categorization task, extremely short (< 3 sec.; 2 observations) and long (> 130 sec., 2 observations) reaction times were removed from the analyses. Mean scores on the Categorization Test for poor readers, average readers and good readers are shown in Table 4.

Table 4. Mean Scores and Standard Deviations in the Categorization Test in Experiment1

	Reader group							
	Poor (n = 14)		Average (n = 48)		Good (n = 34)		Total (n = 96)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Wind instruments								
% errors	28.6	20.3	27.3	21.1	17.0	22.6	23.8	21.9
speed (sec.)	17.0	7.0	17.2	7.6	18.8	12.9	17.7	9.6
description	35.7		41.7		61.8		47.9	
Motor vehicles								
% errors	11.1	16.9	5.3	10.5	5.2	9.6	6.1	11.4
speed (sec.)	18.7	14.0	14.9	7.7	11.8	5.9	14.4	8.6
description	35.7		62.5		41.2		51.0	
Vegetables								
% errors	33.3	24.3	19.9	21.6	20.6	23.2	22.1	22.8
speed (sec.)	23.6	18.9	24.1	15.3	20.4	9.6	22.8	14.2
description	28.6		35.4		55.9		41.7	
Farm animals								
% errors	22.2	19.5	23.4	20.0	15.7	16.0	20.5	18.8
speed (sec.)	24.1	13.3	24.3	14.9	18.1	9.2	22.2	13.2
description	21.4		27.1		26.5		26.0	
Electronic devices								
% errors	34.9	14.4	25.9	17.4	26.8	17.1	27.5	17.0
speed (sec.)	23.6	12.8	27.5	14.7	26.4	14.2	26.5	14.2
description	7.1		10.4		14.7		11.5	
Total								
% errors	26.0	10.3	20.4	9.9	17.1	8.6	20.0	9.9
speed (sec.)	21.4	8.1	23.1	12.7	19.9	9.9	21.7	11.2
description	25.7		35.4		40.0		35.6	

Note. % errors = % pictures incorrectly identified/judged

Mean reaction times, error rates, and description accuracy were submitted to separate ANOVA's (GLM repeated measures). In all analyses, significant effects are based on a .05 alpha level.

Item accuracy. A 3 (Reader group: poor, average, good) x 5 (Category: wind instruments, motor vehicles, vegetables, animals on the farm, electronic devices) ANOVA was performed on error percentages with reader group as between-subjects factor and category as within-subjects factors. The main effect of category was significant, *F*(4,90) = 31.29, *partial* η^2 = .58. Pairwise comparisons revealed that error percentages in motor vehicles were significantly lower than error percentages in all other categories; wind instruments, *F* (1,93) = 43.16, *partial* η^2 = .32; vegetables, *F* (1,93) = 39.52, *partial* η^2 = .30; farm animals, *F* (1,93) = 28.66, *partial* η^2 = .24; electronic devices, *F* (1,93) = 88.78, *partial* η^2 = .49. In addition, error percentages in electronic devices were significantly higher than those in farm animals, *F*(1,93) = 10.59, *partial* η^2 = .10.

Importantly, the main effect of reader group was significant, *F*(2,93) = 4.46, *partial* η^2 = .09. Post-hoc Bonferroni tests revealed that poor readers made significantly more errors (*M*

= 26.0%) than good readers did (*M* = 17.1%). The interaction between category and reader group was not significant.

Description accuracy. A 3 (Reader group: poor, average, good) x 5 (Category: wind instruments, motor vehicles, vegetables, animals on the farm, electronic devices) ANOVA was performed on description accuracy with reader group as between-subjects factor and category as within-subjects factors. The main effect of category was significant, *F*(4,90) = 13.11, *partial* η^2 = .37. Pairwise comparisons revealed that in farm animals, description accuracy was significantly worse than in Wind instruments, *F*(1,93) = 6.92, *partial* η^2 = .07, motor vehicles, *F*(1,93) = 9.07, *partial* η^2 = .09, and electronic devices, *F*(1,93) = 4.11, *partial* η^2 = .04. In electronic devices, description accuracy was significantly worse than accuracy in all other categories; wind instruments, *F*(1,93) = 29.46, *partial* η^2 = .24, motor vehicles, *F*(1,93) = 26.22, *partial* η^2 = .22, and vegetables, *F*(1,93) = 19.82, *partial* η^2 = .18. The main effect of reader group was not significant (*p* = .09), nor was the interaction between category and reader group.

Speed. A 3 (Reader group: poor, average, good) x 5 (Category: wind instruments, motor vehicles, vegetables, animals on the farm, electronic devices) ANOVA was performed on reaction times with reader group as between-subjects factor and category as within-subjects factors. The main effect of category was significant, *F*(4,86) = 10.0, *partial* η^2 = .32. Pairwise comparisons revealed significantly longer reaction times in vegetables than in wind instruments, *F*(1,89) = 8.61, *partial* η^2 = .09, and motor vehicles, *F*(1,89) = 16.14, *partial* η^2 = .15. Reaction times in farm animals were also significantly longer than reaction times in wind instruments, *F*(1,89) = 6.35, *partial* η^2 = .07, and motor vehicles, *F*(1,89) = 17.92, *partial* η^2 = .17. Finally, reaction times in electronic devices were significantly longer than those in wind instruments, *F*(1,89) = 22.55, *partial* η^2 = .20 and motor vehicles, *F*(1,89) = 30.56, *partial* η^2 = .26. The main effect of reader group was not significant, neither was the interaction between reader group and category.

To examine the relationship between accuracy and speed in the semantic-categorization test, correlation analyses were performed. These analyses yielded mixed results, wind instruments, *r* = -.23; motor vehicles, *r* = .28; vegetables; *r* = -.02; farm animals, *r* = .30; electronic devices, *r* = .05. A significant negative correlation (demonstrated in wind instruments) indicated a speed-accuracy trade off: In this category, longer reaction times corresponded with a smaller error percentage. In the categories vegetables and electronic devices, the correlation was not significant. These results show that speed analyses have to be interpreted carefully.

Third, to explore the concurrent predictive value of categorization and association skills for reading, we performed a stepwise hierarchical regression analysis. Results of these analyses are shown in Table 5.

Table 5. Hierarchical Regressions predicting Word Decoding Speed (DMT 1)

	% R^2 change	F change	p
Step 1			
speed DMT 2	74.5	269.33	< .001
Step 2			
Categorization accuracy	5.0	22.51	< .001
Step 3			
speed Klepel	1.3	6.22	.01

We choose DMT1 as a standardized measure of word-recognition skills. Moreover, we decided to report only regression analyses on speed, because speed is the most important factor in Dutch reading. A moderate correlation ($r = .60$) between error percentages and speed on the DMT1, that is, few errors corresponding with short reaction times, indicated that there was no speed-accuracy trade off. Table 5 shows that the most important concurrent predictor of speed in DMT1 was reading speed on DMT2, which accounted for 75% of the variance. On step 2, accuracy in the Categorization Test accounted for a further 5% of variance. Finally, on step 3, speed on the Klepel added 1.3% unique variance. Thus, individual differences in categorization skills contributed significantly to differences in reading speed, even more than speed in a pseudoword-reading test.

To conclude, in the association task, no reader group differences (not in number of associations, nor in typicality of associations) emerged. The only significant effect was the main effect of imageability. High-imageability words evoked more associations and more non-unique responses (i.e., higher typicality of responses) than low-imageability words. In the semantic-categorization task, relatively large differences in reaction times, percentages of errors, and description accuracy between categories indicated that categorization performance strongly depended on the type of category and/or category distracters that were used. Importantly, poor readers made more errors in the semantic-categorization task than good readers did. Another indication for a contribution of semantic skills in reading performance was found in results of the regression analyses, in which accuracy in the semantic-categorization test accounted for unique variance in word-decoding speed.

Experiment 1, however, was limited in several ways. First, the semantic-categorization task included a small number of trials. Moreover, it was a rather complex task and the scores showed a speed-accuracy trade off. Results might have been different if the test required only one response per trial instead of three responses per trial. Second, we only included children from Grade 1 with relatively little reading experience. Vellutino et al. (1995, Experiment 1A) provided evidence for a developmental shift in the relationship between semantic skills and word decoding. They used a test battery of semantic development, which included a vocabulary test, a similarities test, a word-fluency test, a category-fluency test, and a category-verification test. In the category-verification test, a word was presented verbally and the child had to decide whether it belonged to a specific semantic category. Both normal readers and poor readers of Grades 2 and 6 performed the tasks. It turned out that normal readers were significantly faster in these decisions than poor readers, albeit this difference was only significant at Grade 6, and only on items that required a no-response. Results of the vocabulary tests and similarities test showed the same pattern of results: Reader group differences were smaller at second-grade than at sixth-grade level. To address these issues, we decided to run a second experiment.

Experiment 2

Experiment 2 was designed to investigate the development of semantic skills and its relationship with word decoding in time. Semantic categorization and word-decoding skills were studied in Grades 1 to 6. We used a semantic-categorization task in which speed and accuracy of taxonomic knowledge were assessed. Taxonomic knowledge refers to the organization of concepts in a hierarchical structure. It concerns the knowledge, for example, that a poodle belongs to the category ‘dogs’ and that a cat and a dog share similarities in meaning; they both belong to the semantic category ‘animals’. It is generally assumed that knowledge in semantic memory in adults is organized in a hierarchical structure, and even early in childhood, children are able to use taxonomically relations between concepts (e.g., Blewitt & Toppino, 1991; Nguyen & Murphy, 2003). We varied the presentation form of the stimuli by presenting pictures, written letter strings and spoken words as targets. This way, we could examine the relative ease of accessing semantic memory in different input modes. All response options were pictures, for purposes of comparison between reaction times in all tests. Moreover, the use of pictures precludes the problem of young readers with limited word-recognition skills (e.g., Vellutino, Scanlon, DeSetto, & Pruzek, 1981; Vellutino et al., 1990) and enables us to investigate effects of semantic skills independent on task-specific skills underlying the processing of written materials. The type of distracters was manipulated to examine the effect of phonological, semantic, and perceptual characteristics. Several researchers have argued that perceptual similarity is important in learning word meanings in a categorization task (e.g., Gentner & Namy, 1999; Imai, Gentner, & Uchida, 1994). We wanted to investigate whether perceptual similarity of concepts distracts children from correct responding. Phonological similarity was manipulated to investigate whether children use phonological cues in categorizing objects, albeit the instruction is taxonomically oriented. In short, the purpose of the study was to investigate whether (a) poor readers have lower categorization skills - reflected by longer reaction times and more errors in the categorization test - than average or good readers, (b) the difference between poor readers and good readers depends on the type of stimuli (spoken vs. written vs. pictures), (c) beginning readers perform worse in a categorization task than advanced readers, (d) the relationship between decoding skills and categorization skills changes over time, and (e) poor readers are distracted by other characteristics of concepts (phonology vs. semantics vs. shape) than good readers are.

Method

Participants

In this study, 141 students participated: 66 boys (46.8%) and 75 girls (53.2%) from two regular primary schools in the Netherlands. The children were recruited from Grades 1 to 6. The mean age ranged from 7.6 (*SD* = .7) in the lower grades (Grades 1 and 2) to 10.5 (*SD* = 1.2) in the higher grades (Grades 3 to 6). The majority of the children (93%) were native Dutch speakers. Children were assigned to a poor-readers group, average-readers group, or good-readers group. Group characteristics are listed in Table 6.

Table 6. Group Characteristics of the Participants in Experiment 2

		Reading level		
		poor	average	good
Lower grades				
n		12	28	7
age		7.7 (.6)	7.6 (.7)	7.5 (.8)
sex				
	boys	33.3%	46.4%	71.4%
	girls	66.7%	53.6%	28.6%
Higher grades				
n		21	54	19
age		9.8 (.9)	10.7 (1.3)	10.8 (1.0)
sex				
	boys	52.4%	38.9%	63.2%
	girls	47.6%	61.1%	36.8%

Materials

The materials comprised five forced-choice tests to assess taxonomic knowledge and a standardized word-reading test.

Forced-choice semantic tests.

Three Exemplar-stimuli tests (including a picture, spoken word or written word as target stimulus) and two Category-stimuli tests (including a spoken word or written word as target stimulus) were included. In all tests, the response options comprised four pictures for each target stimulus. Children were instructed to choose the picture that matched best with the target stimulus. All pictures originated from 'Leesladder' [Reading Ladder] from Irausquin and Mommers (2001), a computer program for children with reading disabilities. The pictures were colored line-drawings and represented nouns which were highly likely to be known by six-years-old Dutch children (Schaerlaekens, Kohnstamm, Lejaegere, & de Vries, 1999; familiarity rating ≥ .80 on a scale from 0 to 1). In addition, only high-imageability nouns were selected (van Loon-Vervoorn, 1985; imageability rating > 5.5 on a seven-point scale). All stimuli were presented on a laptop. A detailed description of each test is presented below.

Exemplar-stimuli test, pictures.

This test consisted of 20 experimental trials, preceded by three practice trials. A target picture was presented (e.g., orange), followed by four pictures, one target response and three distracters. The target response represented a concept from the same taxonomic category (e.g., cherry). Categories that were included were insects, predators, mammals, rodents, reptiles, sense organs, furniture, transport, clothes, jewels, tools, toys, vegetables, fruit, parts of the body, and buildings. Each trial comprised a different semantic category, except for the latter four categories (vegetables, fruit, parts of the body, and buildings): Each of these categories had two trials. A first distracter picture, which was included in half of the trials (10 out of 20 trials), was a semantic distracter: A concept (e.g., egg) that belonged to the superordinate category 'food', but was no exemplar of the subcategory 'fruit', represented by the target stimulus (orange). A second distracter picture was a phonological distracter (11 out of 20 trials, e.g., in English, 'beer' and a stimulus 'ear') or a perceptual distracter (9 out of 20 trials, e.g., ball and a stimulus 'orange'). The criterion for phonological similarity was sharing the same end-rime with the stimulus. Perceptual similarity was created by similar contours or perceptual features or colors to the stimulus. A third distracter picture, included in all 20 trials, was an unrelated picture (e.g. chair and a stimulus 'orange'); trials without a semantic distracter included two unrelated response options. Criterion for the unrelated concept was the absence of semantic (taxonomic or associative), perceptual, or phonological similarity. Both accuracy and reaction time were measured. The child was asked to choose the picture that matched best with the target stimulus.

Exemplar-stimuli test, spoken words.

This test consisted of 20 experimental trials, preceded by three practice trials. The target stimuli were members from the same categories as those in the pictures test. However, the target stimuli in this test were presented verbally. Response options were presented by pictures. The identity of the target stimuli and response options differed from those in the pictures test but were of comparable difficulty, as indicated by the familiarity ratings of Schaerlaekens et al. (1999). The main body of the target stimuli were one-syllable CVC-, CCVC-, CVCC-words, six stimuli consisted of two syllables. There were no semantically ambiguous words included. The types and distribution of the distracters were identical to the pictures test, except for the distribution of phonological and perceptual distracters, which was now exactly 10 trials for each distracter type. Both accuracy and reaction time were measured.

Exemplar-stimuli test, written words.

This test consisted of 20 experimental trials, preceded by three practice trials. The target stimuli were derived from the same categories as in the pictures test and spoken-word test. However, target stimuli in this test were presented as written letter strings. Response options were presented by pictures. Again, the identity of the target stimuli and response options differed from the other exemplar-stimuli tests, but were of comparable difficulty. The main body of the target stimuli were one-syllable CVC-, CCVC-, CVCC, CCCVC- words, two stimuli consisted of two syllables. The types and distribution of the distracters were identical to the pictures test. Both accuracy and reaction time were measured.

Category-stimuli test, spoken words.

This test consisted of 25 experimental trials, preceded by two practice trials. The name of a semantic category was presented verbally, followed by four pictures, one target response and three distracters. The target response represented a member of the semantic category that was presented. Categories that were included were residence, toys, drinks, cutlery, jobs, transport, musical instruments, tools, animals, clothes, sport, pets, birds, fruit, dairy products, headgear, insects, furniture, vegetables, limbs, mammals, writing tools, sense organs, numbers, and flowers. Additionally, 15 trials included two semantic distracters and one unrelated picture in the response options. Ten trials included three unrelated pictures in the response options. Criteria for these distracters have been described earlier. Both accuracy and reaction time were measured.

Category-stimuli test, written words.

This test consisted of 25 experimental trials, preceded by two practice trials. Categories that were included were identical to the spoken-word test. However, the stimuli in this subtest were presented by written words. Response options were presented by pictures. The identity of the target stimuli and response options differed from the spoken-word test, but were of comparable difficulty. The types and distribution of distracters was the same as those included in the spoken-word test. Both accuracy and reaction time were measured.

Word Decoding

Three-Minutes Test [Drie-Minuten Toets] by Verhoeven (1995).

This test has been described in Experiment 1. In Grades 1 and 2, Card 1 was administered. In Grades 3 to 6, Card 3 was administered. This card consists of 2-syllable, 3-syllable, and 4-syllable words.

Procedure

All forced-choice tests were designed in E-prime (Schneider, Eschman, & Zuccolotto, 2002). Words were recorded in 'Spraak' (Boersma, & Weenink, 2004). The forced-choice tests were performed on a laptop. First, a fixation stimulus (a plus sign, 50-point Times New Roman font) was presented during 1000 ms. Immediately at the offset, the target stimulus (picture, spoken word or letter string) was presented. This stimulus remained visible until a response was provided. Subsequently, four response options (pictures) appeared centrally on the screen and the child had to decide which picture matched best with the stimulus. Participants indicated their responses by means of four keys on the keyboard, corresponding to the position of the pictures on the screen (keys 'c', 'b', 'm', and '.' at the 'qwerty keyboard'). These keys were marked by white tags. The participants were asked to put their hands in front of the keyboard. Word stimuli appeared in white on a black background centrally on the screen in lowercase 24-point Courier New font. There was a 1500 ms delay between the response and the onset of the next trial. Each participant was presented a different random order. Reaction times were measured from the offset of the stimulus (the time that the space-bar was pressed) until a response was given.

Children performed the forced-choice tests in groups of 8 students. All students started with the exemplar-stimuli tests. Subsequently, the category-stimuli tests were performed. The order of the subtests in the exemplar-stimuli tests (pictures test, spoken-word test and written-word test) and category-stimuli tests (spoken-word test and written-word test) was varied. Children who were tested simultaneously received the subtests in identical orders. Trials within a subtest were randomized. All forced-choice tests were performed by all children, except for the Category-stimuli written test in Grades 1 and 2, because of insufficient reading experience in these grades. Children from Grades 1 and 2 were tested in two different sessions of 30 minutes each. Children from Grades 3 to 6 performed all tests in one single session. The DMT (word identification) was administered in March (Grades 2 to 6) and May (Grade 1).

Based on the DMT-scores, three groups of readers were distinguished: Poor readers, average readers, and good readers. Poor readers were students with poor word-recognition skills, that is, reading scores below the 25th percentile. Average readers had reading scores between the 25th and 75th percentile and good readers performed above the 75th percentile. To examine the effect of grade, we discriminated between the lower grades (Grades 1 and 2) and higher grades (Grades 3, 4, 5, and 6). We distinguished between these groups because explicit reading instruction ends in Grade 2. After that time, the focus shifts towards reading comprehension.

Results

For each participant, mean reaction times and accuracy percentages were calculated for each semantic test, aggregated over items. For each test, incorrect trials and reaction times more than two standard deviations below or above participant's mean reaction time were excluded from the latencies analyses[3]. The percentages of outliers in each test are listed in Table 7.

Table 7. Percentages of Outliers in the Semantic Tests removed from the Analyses

		Reaction Times			Accuracy		
Test	*n*	< 2 *SD*	> 2 *SD*	total	practice items	errors	total
Exemplar-stimuli							
pictures	141	.4	7.3	7.6	13.0	19.7	32.7
spoken	141	.4	8.2	8.6	13.0	28.0	41
written	141	.4	9.3	9.8	13.0	22.3	35.3
Category-stimuli							
spoken	141	.1	8.4	8.5	7.4	16.0	23.4
written	94	0	8.4	8.4	7.4	10.1	17.5

---

[3] We realize that the large number of outliers might lead to spurious results. Therefore, we reanalyzed the reaction times with the inclusion of all outliers. The same pattern of results was obtained.

Table 8 shows the results of the Categorization Test for poor readers, average readers, and good readers.

Table 8. Mean Error Percentages and Standard Deviations in the Semantic-Categorization Test

	Reader group							
	Poor (*n* =33)		Average (*n* = 82)		Good (*n* = 26)		Total (*n* = 141)	
Accuracy	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Exemplar-stimuli								
written	28.0	14.3	24.9	13.6	24.8	12.8	25.6	13.6
spoken	35.3	18.9	31.5	14.0	30.0	14.4	32.1	15.3
pictures	25.5	13.9	21.7	14.0	21.7	14.8	22.6	14.1
Category-stimuli								
written	15.4	8.4	9.5	7.1	9.7	8.5	10.9	8.0
spoken	19.3	9.4	16.6	9.9	17.1	15.7	17.3	11.1
Reaction times	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Exemplar-stimuli								
written	3181	1049	2809	848	2473	531	2834	878
spoken	3682	1376	3245	1161	2963	850	3295	1183
pictures	3407	1507	2771	957	2690	637	2905	1095
Category-stimuli								
written	2149	581	1848	481	1863	432	1918	506
spoken	2334	599	2097	675	2041	521	2142	637

Before we address the research questions, three related issues will be discussed. First, we wanted to examine whether the reading level of the participants was reflected in the time each participant needed to decode the written stimuli in the Exemplar-stimuli written test and Category-stimuli written test. Therefore, we performed a correlation analysis on the decoding time in the written subsets of the semantic-categorization tests and the scores on DMT1 and DMT3, which reflected word-decoding skills. In the lower grades, scores on the DMT1 significantly correlated ($p < .01$) with decoding time in the Exemplar-stimuli written test, $r = -.64$. In the higher grades, scores on the DMT3 significantly correlated with decoding time in the Exemplar-stimuli written test ($r = -.37$) and Category-stimuli written test ($r = -.68$). Thus, poor readers read target stimuli in the semantic-categorization tests slowly, whereas good readers demonstrated relatively fast decoding times in the semantic tests. Subsequently, we excluded decoding times of target stimuli from further analyses. This way, results of the semantic-categorization tests could not be affected by decoding times.

Second, in view of the observed pattern of the mean scores in the semantic-categorization tests, we decided to compare the mean reaction times and error percentages on the Exemplar-stimuli tests with the means on the Category-stimuli tests. In addition, we compared the mean reaction times and error percentages of the subtests. A 2 (Grade: lower vs. higher) x 3 (Reading level: poor vs. average vs. good) x 4 (Test: Exemplar-stimuli written, Exemplar-stimuli spoken, Exemplar-stimuli picture, Category-stimuli spoken) ANOVA was performed on reaction times and error rates with grade and reading level as between-subjects factor and test as within-subjects factors.

Reaction times. The main effect of test was significant, *F*(3,133) = 53.56, *partial* $\eta^2 = .55$. Subsequent analyses revealed that reaction times in the Exemplar-stimuli tests (*M* = 3011) were significantly longer than those in the Category-stimuli test (*M* = 2030), *F*(1,135) = 154.79, *partial* $\eta^2 = .53$. In the Exemplar-stimuli tests, reaction times in the spoken subtest (*M* = 3295) were significantly longer than those in the written subtest (*M* = 2834), *F*(1,135) = 25.80, *partial* $\eta^2 = .16$, and in the pictures subtest (*M* = 2905), *F*(1,135) = 12.40, *partial* $\eta^2 =$ .08.

Errors. The main effect of test was significant, *F*(3,133) = 27.03, *partial* $\eta^2 = .38$. Subsequent analyses revealed that error percentages in the Exemplar-stimuli tests (*M* = 28.4%) were significantly higher than those in the Category-stimuli test (*M* = 20.0%), *F*(1,135) = 47.17, *partial* $\eta^2 = .26$. In the Exemplar-stimuli tests, error percentages on all tests differed significantly from each other; error percentages in the spoken subtest (33.9%) were significantly higher than those in the written subtest (27.8%), *F*(1,135) = 17.20, *partial* $\eta^2 =$ .11, which in turn were significantly higher than those in the pictures subtest (23.6%) , *F*(1,135) = 7.70, *partial* $\eta^2 = .05$.

Third, we performed a correlation analysis on speed and error percentages in all semantic-categorization tests. Results revealed significant correlations for all tests, except for the Category-stimuli written test, Exemplar-stimuli pictures, *r* = .22; Exemplar-stimuli spoken, *r* = .38; Exemplar-stimuli written, *r* = .24; Category-stimuli spoken; *r* = .30, Category-stimuli written, *r* = .06. These results indicate that there was no speed-accuracy trade off: Longer reaction times corresponded with higher error percentages.

To answer our research questions, we performed multivariate analyses. Mean reaction times and error rates were submitted to separate ANOVA's (GLM repeated measures). A 2 (Grade: lower vs. higher) x 3 (Reading level: poor vs. average vs. good) x 4 (Test: Exemplar-stimuli written, Exemplar-stimuli spoken, Exemplar-stimuli picture, Category-stimuli spoken) ANOVA was performed on reaction times and error rates with grade and reading level as between-subjects factor and test as within-subjects factors. In all analyses, significant effects are based on an alpha level of .05. Because the Category-stimuli written test was not administered in the lower grades, we excluded this test from the analyses.

(a) Poor readers vs. good readers

With respect to reaction times, a significant main effect of reading level was demonstrated, $F(2,135) = 5.0$, *partial* $\eta^2 = .07$. A post-hoc Bonferroni test revealed that poor readers (M = 3233 ms) had significantly longer reaction times than average readers (M = 2829 ms) and good readers (M = 2658 ms). Although poor readers appeared to make more errors (M = 27.9%) than average readers (M = 25.7%) and good readers (M = 25.3%), this difference was not significant. We are aware of the small difference in mean age between the reader groups. The mean age of the poor readers in the higher grades was lower than that the mean age of the average readers and good readers. Although this might appear to be a confound for the results, this is highly unlike for the following reason: The difference in semantic performance between the reader groups in the higher grades revealed exactly the same pattern as semantic-performance differences between the reader groups in the lower grades (see research question d). In the lower grades, the mean age of the reader groups was comparable. Thus, reader group differences in semantic performance are presumably not caused by other factors like the age of the participants, at least not in the lower grades.

(b) Reading level differences related to the type of stimuli

No significant interaction effect between test and reading level was found, neither in the reaction times, $F(6,266) = 1.12$, $p > .10$ nor in error percentages, $F < 1$. Thus, differences in semantic-categorization performance between poor readers, average readers, and good readers were similar for spoken words, written words, and pictures.

(c) beginning readers vs. advanced readers

The main effect of grade was significant, both for reaction times, $F(1, 135) = 15.82$, *partial* $\eta^2 = .11$, and errors, $F(1,135) = 24.90$, *partial* $\eta^2 = .16$. Reaction times in the lower grades (M = 3195 ms) were significantly longer than those in the higher grades (M = 2619 ms). Error percentages in the lower grades (M = 30.9%) were significantly higher than those in the higher grades (M = 21.7%). In the error percentages, the main effect was qualified by a significant test by grade interaction, $F(3,133) = 2.85$, *partial* $\eta^2 = .06$. On each test, performance differences between the lower grades and higher grades were significant. However, performance difference between the lower grades and higher grades in the Exemplar-pictures test was smaller than performance differences between these grades in all other tests. This interaction is shown in Figure 1.

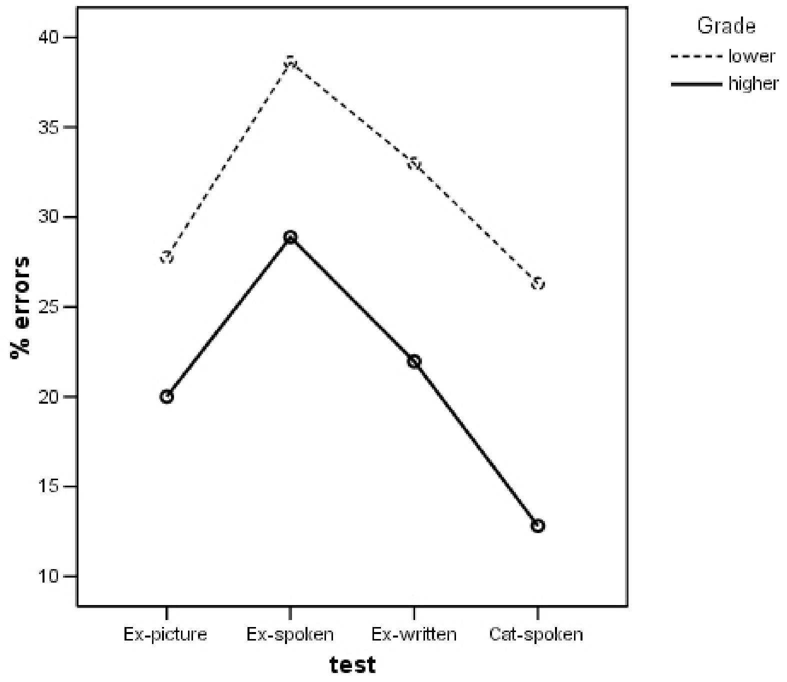

Figure 1. Mean Percentage of Errors of the lower Grades and higher Grades in the Semantic Tests in Experiment 2

(d) changing relationship over time

Neither in reaction times (*F* < 1), nor in error percentages (*F*(2,135) = 1.34, *p* > .10), a significant interaction effect between grade and reading level was found. This means that differences in semantic performance between poor, average, and good readers in the lower grades were similar to semantic performance differences in the higher grades.

The preceding analyses did not include results of the Category-stimuli written test. Next, to examine performance in this test, we carried out a 3 (Reading level: poor vs. average vs. good) x 5 (Test: Exemplar-stimuli written, Exemplar-stimuli spoken, Exemplar-stimuli picture, Category-stimuli spoken, and Category-stimuli written) ANOVA on reaction times and errors with reading level as between-subjects factor and test as within-subjects factor. Because the Category-stimuli written test was not administered in the lower grades due to the difficulty of this test, the analyses are restricted to Grades 3 to 6. We only report the results with respect to the Category-stimuli written test, and its' relationship with all other tests.

The main effect of test was significant for reaction times, *F*(4,88) = 45.51, *partial* η^2 = .67, and errors, *F*(4,88) = 26.65, *partial* η^2 = .55. Reaction times in the Exemplar-stimuli tests were significantly longer than those in both Category-stimuli tests, *F*(1,91) = 174.71, *partial* η^2

= .66 and significantly more errors were made in the Exemplar-stimuli tests than in both Category-stimuli tests, *F*(1,91) = 82.48, *partial* $\eta^2 = .48$. No significant difference was found between reaction times in the Category-stimuli written test and those in the Category-stimuli spoken test, *F* < 1. The analysis on the number of errors in the Category-stimuli tests was marginally significant, *F*(1,91) = 3.45, *partial* $\eta^2 = .04$, p = .07: More errors were made in the Category-stimuli spoken test than in the Category-stimuli written test.

(e) distracter options

The final research question addressed differences between poor readers and good readers in the type of responses. To examine whether poor readers are distracted by other characteristics of concepts (phonology vs. semantics vs. shape) than good readers, we analyzed how often each of the distracters was chosen by poor readers, average readers, and good readers. To test for differences between the distracter responses, we submitted the frequencies (in percentages) of all distracter responses to a GLM repeated-measures analyses for both the Category-stimuli tests and the Exemplar-stimuli tests, and for all possible item series. Item series refer to the combination of response options in a trial, that is, the combination of a correct response, perceptual distracter, semantic distracter, and an unrelated distracter. In the Category-stimuli tests, two different item series were included; in the Exemplar-stimuli tests, four different item series were included.

Both in the Category-stimuli tests and the Exemplar-stimuli tests, results failed to demonstrate significant interaction effects between reading level and response, indicating that poor readers were not differently affected by distracters than average or good readers were. Table 9 shows the mean frequencies for each of the response options in the Category-stimuli tests. Mean frequencies in the Exemplar-stimuli tests are shown in Table 10.

Table 9. Frequencies (%) of Correct and Distracter Choices in the Category-Stimuli Tests

Distracter	spoken		written	
	M	*SD*	*M*	*SD*
Item series 1				
correct	74.7	14.2	83.6	11.1
unrelated	4.6	6.2	3.0	5.5
semantic	11.2	8.8	6.1	5.6
semantic	9.6	8.3	7.3	8.1
Item series 2				
correct	92.9	10.1	96.2	7.1
unrelated	1.5	3.5	1.5	4.3
unrelated	2.6	4.8	1.0	2.8
unrelated	3.0	6.1	1.4	4.0

Table 10. Frequencies (%) of Correct and Distracter Choices in the Exemplar-Stimuli Tests

Distracter	Spoken		Pictures		Written	
	M	*SD*	*M*	*SD*	*M*	*SD*
Item series 1						
correct	63.3	19.4	84.4	16.5	57.3	21.1
phonological	8.2	16.4	3.9	10.5	15.1	19.0
unrelated	3.1	7.7	3.9	8.1	.7	3.4
semantic	25.4	15.7	7.8	10.3	26.8	16.2
Item series 2						
correct	83.6	22.8	90.4	17.5	90.2	19.1
phonological	8.5	16.6	5.5	14.2	4.1	12.8
unrelated	3.7	8.5	3.4	7.9	2.6	7.5
unrelated	4.3	10.4	.7	3.7	3.1	8.0
Item series 3						
correct	55.0	25.8	67.0	26.3	84.8	20.9
perceptual	19.9	22.4	13.3	21.0	6.6	14.2
unrelated	2.6	6.7	1.6	6.1	1.1	5.1
semantic	22.6	18.3	18.0	18.5	7.6	14.0
Item series 4						
correct	69.7	22.0	64.4	22.6	70.6	19.9
perceptual	26.0	19.7	28.2	20.3	23.1	17.1
unrelated	1.3	5.5	2.4	6.5	1.4	5.7
unrelated	3.1	8.7	5.0	10.2	4.8	9.5

A discussion of all significant effects is beyond the scope of this chapter. The interested reader is referred to Appendix A, in which F-values of all significant effects of the GLM-analyses are listed.

General discussion

In the present study, the relationship between semantic skills and word-decoding skills was examined. In Experiment 1, children from Grade 1 performed a word-association task and a semantic-categorization task. Task performances were related to word-decoding performance. In Experiment 2, readers from Grades 1 to 6 performed a set of semantic-categorization tasks. Results indicated that in the word-association task, no reader group differences -neither in number of associations nor in typicality of associations - emerged. Thus, poor readers generated as many associations as average readers and poor readers did and there was no evidence for more idiosyncratic responses in poor readers. Observation of the responses revealed mainly syntagmatic responses in both poor readers and good readers. This result corresponds with Cronin et al. (1986), who demonstrated that reading skill was a predictor of the number of paradigmatic responses; the better reading skills were developed, the more paradigmatic responses were observed. In our study, only beginning readers who had relatively little reading experience were included, and as a consequence, participants demonstrated mainly syntagmatic responses. The only significant effect was the main effect of imageability; in high-imageability words, more associations were generated and more similarity in responses was observed than in low-imageability words. The imageability advantage is in line with results of Brysbaeart, Van Wijnendaele, and De Deyne (2000), who demonstrated shorter reaction times to high-imageability words than low-imageability words in a word-association task in adults. They also found that participants reported more frequently the same associate for high-imageability words, a result that corresponds with our data on children. The results are also according to studies on memory (e.g., Vellutino et al., 1995, Experiment 1B) and reading performance (e.g., Cortese, Simpson, & Woolsey, 1997; de Groot, 1989; Strain, Patterson, & Seidenberg, 1995).

Although poor, average, and good readers did not differ in their word-association skills, both experiments revealed performance differences between poor readers and good readers in semantic categorization. In Experiment 1, we performed a semantic-categorization task which required insight in similarities and differences across concepts of related categories. Results revealed that poor readers made more errors than good readers did. These results perhaps indicate that for poor readers, it was more difficult to classify exemplars in a specific category (e.g., vegetables) and exclude exemplars of a related category (e.g. fruit). Boundaries between categories become fuzzy if different categories share a lot of characteristics (Jerger & Damian, 2005). Seemingly, this difficulty affected poor readers' performance more than good readers' performance. These results suggest a positive relationship between categorization skills and decoding skills. This suggestion was supported by results of a regression analysis in Experiment 1, which demonstrated unique contribution of categorization accuracy in the prediction of decoding speed. The contribution was even higher than the contribution of performance in the pseudoword test, which might be due to the relative difficulty of this test. However, in Experiment 2, in which a different type of semantic-categorization test was assessed across all grades in primary school, accuracy of poor readers was statistically equal for poor readers, average readers, and good readers. The percentage of errors for all readers ranged between 11% and 32%. In contrast, results demonstrated longer reaction times in poor readers than in average or good readers and there was no speed-accuracy trade off. These results –poor readers had relatively high accuracy, but long reaction times- are in line with the results of Howell and Manis (1986). In their study, participants had to decide whether a stimulus (e.g., dog) was presented with either the correct superordinate label (i.e., animal) or the correct basic label (i.e., dog). Poor readers were significantly slower in making decisions than controls. In addition, low-typical items took longer than high-typical items and the magnitude of this effect was similar for poor readers and normal readers. Based on these findings - low error-rates and large typicality effects in poor readers-, the authors suggested that the structure of semantic memory of poor readers might be intact, whereas differences in speed cause lower performance on semantic-processing tasks.

Second, in Experiment 2, no significant interaction effect between test and reading level was found, neither in reaction times nor in error percentages: Differences in semantic-categorization performance between poor readers, average readers, and good readers were similar for spoken words, written words, and pictures. Thus, lower performance in the semantic-categorization test in the current study cannot be attributed to difficulties in processing written stimuli: Even in a test including pictures and spoken words, poor readers responded slower than good readers. This result corresponds with the studies of Howell and Manis (1986) and Ben-Dror et al. (1995). They presented a semantic-categorization test with respectively picture stimuli and verbal stimuli and demonstrated lower semantic-categorization performance in poor readers than in normal readers (but see, Silva-Pereyra, 2003 for absence of reader-group differences in a picture-categorization task). Whereas reader-group differences in semantic performance were statistically equal for pictures, spoken words, and written words, overall semantic performance varied across the different types of stimuli: Results revealed that semantic-categorization performance was worst (longer reaction times and more errors) if target stimuli were presented verbally. This might be partly caused by the temporal character of spoken words related to visual stimuli, which might be difficult for young children to draw attention to.

Third, we investigated whether beginning readers performed worse in a semantic-categorization task than advanced readers. Results demonstrated longer reaction times and more errors in the lower grades than in the higher grades. This proves the sensitivity of the test to developmental changes. Increasing accuracy in semantic categorization with age is in line with previous studies (e.g., Nguyen & Murphy, 2003).

Fourth, we wanted to know whether the relationship between decoding skills and semantic-categorization skills changes over time. Absence of an interaction effect in reaction times between grade and reading level indicated that differences in semantic performance between poor, average, and good readers in the lower grades were similar to performance differences between the reader groups in the higher grades. This result is in contrast with the results of Vellutino et al. (1995, Experiment 1A). They presented second graders and sixth graders several tests of semantic development (e.g., a vocabulary test and a category-verification test) and found reader-group differences on semantic performance (reflected in both accuracy and speed, depending on the type of semantic test) to be smaller at the second-grade level than at the sixth-grade level.

Finally, we examined whether poor readers are distracted by other characteristics of concepts (phonology vs. semantics vs. shape) than good readers. Absence of an interaction effect between reading level and response indicated that poor readers were not differently affected by the phonological, perceptual, and semantic distracters than average readers or good readers were.

Taken together, these results indicate that semantic skills are related to reading performance in a specific way. On the basis of the results of Experiment 2, we suggest that poor readers have slower access to semantic information than average readers or good readers, regardless of the modality of the stimuli. Although reading performance and semantic-categorization accuracy were significantly correlated in Experiment 1, we reject the notion of a qualitative difference between the reader groups. Identical word-association skills in Grade 1 and identical error patterns of all reader groups and similar preferences for different types of distracters in Experiment 2 support this view. Differences in task instruction might be responsible for different outcomes across studies.

Two issues related to the current study deserve some attention. First, our findings do not imply that semantic processing per se is affected in poor readers. A distinction between offline procedures, which require conscious processing, and online procedures, which address automatic components of semantic organization, needs to be considered. Several researchers (e.g., Assink, Van Bergen, Van Teeseling, & Knuijt, 2004; Nation & Snowling, 1999) have argued that offline tasks like expressive vocabulary and semantic categorization are not adequate to examine implicit semantic knowledge, which is probably more directly linked to word decoding than conscious processes. Thus, several researchers have suggested that semantic-priming studies might be a better alternative. Assink et al. (2004) carried out a semantic-priming experiment with poor readers and controls and found no differences between poor decoders and normal decoders in sensitivity to semantic cues. Waterman and Lewandowski (1993) verbally presented sets of words and participants had to decide whether the word had been heard before in the list. The test included groups of antecedent words (bag), rhyming targets (rag), rhyming controls (dab), semantic targets (sack), and semantic controls (mess). All readers made significantly more errors on targets than on controls. To illustrate, both poor readers and good readers more often incorrectly reported that they had heard a semantic target (e.g., sack) than a semantic control (e.g. mess). Importantly, results showed that this difference between responses on semantic targets and semantic controls was significantly larger for poor readers than for good readers. Thus, poor readers made more semantic errors than good readers. On the basis of these results, the authors suggested that the semantic-processing system of poor readers is intact and might even be the preferred coding strategy in poor readers.

Thus, semantic processes can be assessed either in a direct way (conscious, offline processing) or in a rather indirect way (e.g., semantic priming) and both methods address different components. We think that both online procedures and offline procedures might be important in the study of semantic processing skills of poor readers. Deficiencies in conscious semantic processing do not automatically imply deficits in unconscious semantic processing and vice versa.

A second issue concerns the kind of relationship between semantic skills and reading. If there is indeed a positive relationship between (offline) semantic processing and reading, what does this relationship look like? Are semantic deficits causal to reading difficulties or is it just the opposite - lower reading performance is responsible for lower semantic skills -, or are both skills the result of an underlying process or do they affect each other? Vellutino et al. (1995) found that reader-group differences were larger in sixth graders than in second graders and concluded that semantic deficits are more often a consequence than a cause of reading problems. However, the present study does not support this conclusion. We failed to find an interaction effect between grade and reading level, suggesting that differences in semantic performance between poor, average, and good readers in the lower grades were similar to semantic-performance differences between reader groups in the higher grades. The present study cannot answer the complex question of causality, rather is an attempt to contribute to the knowledge of reading difficulties and its relationship to other skills. Longitudinal studies are needed to investigate the direction of the relationship between word-decoding skills and semantic skills. Results of the present study fit with the idea that other skills than phonological skills (e.g., semantic skills), are related to reading difficulties, albeit possibly to a smaller degree than phonological skills are (e.g., Catts et al., 1999; Frost et al., 2005; Gillon & Dodd, 1994).

REFERENCES

Assink, E. M. H., Van Bergen, F., Van Teeseling, H., & Knuijt, P. P. N. A. (2004). Semantic priming effects in normal versus poor readers. *The Journal of Genetic Psychology, 165, 67-79.*

Ben-Dror, I., Bentin, S., & Frost, R. (1995). Semantic, phonologic, and morphologic skills in reading disabled and normal children: Evidence from perception and production of spoken Hebrew. *Reading Research Quarterly, 30,* 876-893.

Blewitt, P., & Toppino, T. C. (1991). The development of taxonomic structure in lexical memory. *Journal of Experimental Child Psychology, 51,* 296-319.

Bradley, L. & Bryant, P. E. (1983).Categorizing sounds and learning to read: a causal connection. *Nature, 301,* 419-421.

Bryant, P. E., MacLean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology, 26,* 429-438.

Brysbaert, M., Lange, M., & Van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology, 12, 65*-85.

Catts, H. W., Fey, M.E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3,* 331-361.

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

Elbro, C., Borstrøm, I., & Petersen, D. K. (1998). Predicting dyslexia from kindergarten. The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly, 33*, 36-60.

Frost, J., Madsbjerg, S., Niedersøe, J., Olofsson, Å., & Møller Sørensen, P. (2005). Semantic and phonological skills in predicting reading development: From 3-16 years of age. *Dyslexia, 11,* 79-92.

Gillon, G., & Dodd, B. J. (1994). A prospective study of the relationship between phonological, semantic and syntactic skills and specific reading disability. *Reading and Writing, 6,* 321-345.

Howell, M.J., & Manis, F. R. (1986). Developmental and reader ability differences in semantic processing efficiency. *Journal of Educational Psychology, 78,* 124-129.

Jerger, S., & Damian, M. F. (2005). What's in a name? Typicality and relatedness effects in children. *Journal of Experimental Child Psychology, 92, 46-75.*

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities, 17,* 592-599.

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language, 39,* 85-101.

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition, 70,* B1-B13.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: broader language skills contribute to the development of reading. *Journal of Research in Reading, 27,* 342-356.

Nguyen, S. P., & Murphy, G. L. (2003). An apple is more than just a fruit: Cross-classification in children's concepts. *Child Development, 74,* 1783-1806.

O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading,* 3, 159-197.

Scarborough, H. S. (1989). Prediction of reading disability from familial and individual differences. *Journal of Educational Psychology, 81,* 101-108.

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

Scarborough, H. S. (1991). Antecedents to reading disability: preschool language development and literacy experiences of children from dyslexic families. *Reading and Writing, 3,* 219-233.

Schaerlaekens, A., Kohnstamm, G. A., & Lejaegere, M. (1999). *Streeflijst woordenschat voor zesjarigen* [Target list vocabulary for 6-year-old children]. Lisse, The Netherlands: Swets & Zeitlinger.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime.* Pittsburgh, PA: Psychology software tools.

Silva-Pereyra, J., Rivera-Gaxiola, M., Fernández, T., Díaz-Comas, L., Harmony, T., Fernández-Bouzas, A., Rodríguez, M., Bernal, J., & Marosi, E. (2003). Are poor readers semantically challenged? An event-related brain potential assessment. *International Journal of Psychophysiology, 49,* 187-199.

Van Loon-Vervoorn, W. A. (1985). *Voorstelbaarheidswaarden van Nederlandse Woorden: 4600 substantieven, 1000 verba en 500 adjectieven* [Imageability ratings of Dutch words: 4600 substantives, 1000 verbs and 500 adjectives]. Lisse: Swets & Zeitlinger.

Vellutino, F.R., Scanlon, D.M., DeSetto, L., & Pruzek, R.M. (1981). Developmental trends in the salience of meaning versus structural attributes of written words. *Psychological Research, 43,* 131-153.

Vellutino, F.R., Scanlon, D.M., & Spearing, D. (1995). Semantic and phonological coding in poor and normal readers. *Journal of Experimental Child Psychology, 59,* 76-123.

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1990). Differential sensitivity to the meaning and structural attributes of printed words in poor and normal readers. *Learning and Individual Differences, 2,* 19-43.

Van den Bos, K. P. Lutje Spelberg, H. C., Scheepstra, A. J. M., & de Vries, J. R. (1994). *De Klepel, een test voor de leesvaardigheid van pseudowoorden.* [Klepel, a test for decoding skills of pseudowords]. Nijmegen: Berkhout.

Verhoeven, L. (1992). *Drie-Minuten-Toets* [Three-minute test]. Arnhem: CITO.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192-212.

Waterman, B., & Lewandowski, L. (1993). Phonologic and semantic processing in reading-disabled and non-disabled males at two age levels. *Journal of Experimental Child Psychology, 55,* 87-103.

Wesseling, R., & Reitsma, P. (2000). The transient role of explicit phonological recoding for reading acquisition. *Reading and Writing, 13,* 313-336.

Appendix A

Concepts in the Semantic-Categorization Task in Experiment 1

			Category		
	wind instruments	motor vehicles	vegetables	animals on the farm	electronic devices
Items					
example	flute	car	broccoli	cow	flat-iron
target	German flute	bus	white cabbage	sheep	hoover
	tooter	tractor	carrots	chicken	coffee-maker
	cornet	truck	French beans	pig	washing machine
distracter	piano	step	cherries	elephant	teapot
	guitar	bicycle	citron	swan	cup
	violin	carriage	banana	hedgehog	scissors
	harp	barrow	grapes	cat	broom
	drum	go-card	apple	fish	saucepan
	triangle	sailing boat	pear	monkey	sponge

Appendix B

Results of GLM-Analysis on Response Options in the Semantic-Categorization Tests in Experiment 2

Test and effect	*F*	*p*	partial η^2
Category-stimuli			
Item series 1			
test	*F*(1,91) = 6.40	< .05	.07
response	*F*(2,90) = 22.03	< .001	.33
reading level	*F*(2,91) = 4.77	< .05	.10
test x response	*F*(2,90) = 4.36	< .05	.09
Exemplar-stimuli			
Item series 1			
test	*F*(2,137) = 87.84	< .001	.56
response	*F*(2,137) = 266.48	< .001	.80
text x response	*F*(4,135) = 37.91	< .001	.53
Item series 2			
test	*F*(2,137) = 6.42	< .01	.09
response	*F*(2,137) = 6.86	< .01	.09
text x response	*F*(4,135) = 4.33	< .01	.11
Item series 3			
test	*F*(2,137) = 62.62	< .001	.48
response	*F*(2,137) = 148.80	< .001	.69
text x response	*F*(4,135) = 18.76	< .001	.36
Item series 4			
test	*F*(2,137) = 2.87	.06	.04
response	*F*(2,137) = 174.62	< .001	.72

Chapter 5: Imageability Effects in Printed-Word Perception in Primary Grades[1]

Abstract

This study focused on imageability effects in isolated-word reading in Dutch children from Grades 2 to 6. Word-reading skills were assessed by lexical decision (Experiment 1) and naming (Experiment 2). In both experiments, a speeded task as well as a non-speeded task were administered. Results of Experiment 1 revealed an imageability advantage in the non-speeded task in latencies and an imageability advantage in low-frequency words in accuracy; in low-frequency words, more errors were made in low-imageability (LI-) words than in high-imageability (HI-) words. Moreover, accuracy analyses revealed a significant grade by instruction by imageability interaction; in the higher grades, only in the speeded test more errors were made in LI-words than in HI-words. Results of Experiment 2 indicated that in naming, no imageability advantage was demonstrated for accuracy scores. Unexpectedly, children in the lower grades made more errors in HI-words than in LI-words, but only in high-frequency words and only in the speeded task. Results are discussed within different theoretical accounts of reading.

---

[1] Submitted for publication as Gijsel, M. A. R., Bosman, A. M. T., & Verhoeven, L. Imageability Effects in Printed-Word Perception in Primary Grades. We thank Sanne Basten, Christa Wintjes, Bernice Bolscher, and Moniek Sap who collected the data. We are also grateful to all children who participated in the study and to the teachers who gave their permission to run the experiment.

Introduction

There is considerable agreement in the literature about the assumption that reading involves the contribution of orthography, phonology, and semantics. A large number of researchers have emphasized the prominent role of phonology in printed-word perception (for overviews, see Berent & Perfetti, 1995; Bosman & van Hell, 2002; Frost, 1995; Jared & Seidenberg, 1991; Van Orden, Pennington, & Stone, 1990). In contrast, the role of semantics in (isolated) printed-word perception has been neglected for a long time. Most researchers assumed that word meanings were activated only after recognition of a letter string, implying no mediating role for semantics in the word-reading process. However, the last decades, the impact of semantics in word recognition has received much more attention (e.g., see Balota, Ferraro, & Connor, 1991 for an overview; Locker, Simpson, & Yates, 2003; Pexman & Lupker, 1999) and it is now argued that semantics also facilitates isolated word reading under certain circumstances. The increased focus on semantics is reflected in more recent (connectionist) models of printed-word perception (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996; Taft & van Graan, 1998; Van Orden, Bosman, Goldinger, & Farrar, 1997). In this study, semantic contribution in isolated word reading was examined by means of imageability manipulations. We investigated whether word imageability affected lexical-decision performance and naming performance of Dutch children from Grades 2 to 6. The focus of the present study is not on the structure of semantic memory and mechanisms to explain imageability, rather on its relationship with word-decoding processes.

Semantics effect in printed word perception

Evidence for semantic activation in reading isolated words has been found in studies in which a semantic variable has been manipulated and demonstrated to affect naming latencies, reaction times and/or reading accuracy in lexical decision. One example of such a semantic variable is ambiguity. It has been argued that words associated with multiple meanings (Borowsky & Masson, 1996; Hino, Lupker, & Pexman, 2002; Locker, Simpson, & Yates, 2003; Pexman & Lupker, 1999) are responded to faster in a lexical-decision task or naming task than words with unambiguous meanings. Locker et al. (2003) further explored the ambiguity effect and showed that the ambiguity advantage in a lexical-decision task only occurred within small-set low-connectivity words, that is, words with a low number of meaningfully related responses and a low number of associative connections.

The opposite occurs with synonyms, that is, reaction times or reading times are longer and accuracy is lower when reading synonyms than reading words without a synonym (Hino et al., 2002; Pecher, 2001). Another example of a semantic manipulation is the number of features (Pexman, Lupker, & Hino, 2002). Semantic features are defined by attributes or characteristics that describe word meanings. Pexman et al. showed in both lexical decision and naming that concrete words with a high number of features, like 'lion' are recognized faster than concrete words with a low number of features, like 'lime'. Finally, we mention the manipulation of word imageability, the variable we will concentrate on in the present study (e.g., Cortese, Simpson, & Woolsey, 1997; de Groot, 1989; Raman & Baluch, 2001; Strain, Patterson, & Seidenberg, 1995).

Imageability

The focus of the present study is on imageability effects in reading performance in Dutch children. We choose to investigate semantics by means of imageability manipulations, following the argument of Strain, Patterson, and Seidenberg (1995) that "..it is one of the best predictors of oral-reading performance in certain acquired disorders of reading".

Imageability is probably the most widely investigated semantic feature (Balota et al., 1991) and is defined as the extent to which the representation of a word's meaning has sensori-motor properties (Strain et al., 1995) or the extent to which the referent of the word evokes a mental image (de Groot, 1989). The term imageability is often confounded with concreteness, as is reflected in the first definition. In the current study, no explicit distinction is made between these concepts. Most researchers investigated imageability effects in printed-word perception in skilled adult readers. Strain et al. (1995) used a word-naming task to study the effect of imageability on word recognition in adults. Effects of regularity, frequency, and imageability were investigated. The results showed shorter latencies and fewer regularization errors for high-imageability words than low-imageability words. However, error analyses revealed a significant interaction between imageability and frequency, indicating that the imageability effect was only present in low-frequency words. In a second experiment, increasing the number of stimuli and participants, words were matched for positional-bigram frequencies and only low-frequency words were included. Both error and latency data indicated an interaction between regularity and imageability; the imageability effect was only significant in exception words. These findings of Strain et al. encouraged several researchers to carry out similar studies or to replicate one or more of their experiments. Cortese et al. (1997) replicated and extended the results of Strain et al. In their experiment, related and unrelated word pairs were presented in a naming task. Results revealed that among irregular words, high-imageability stimuli were responded to faster and with fewer errors. However, this was not the case in regular words. These findings are consistent with the results of Strain et al. and suggest that imageability effects are greater when phonological coding is slow.

Imageability and confounding variables

Other researchers, however, have questioned unique contribution of the imageability effect in reading performance (e.g., Brysbaert, Lange & Van Wijnendale, 2000; Monaghan & Ellis, 2002; Shibahara et al., 2003). Monaghan and Ellis questioned the role of imageability and replicated Experiment 2 of Strain et al. (1995). They found the same imageability effect on naming. However, this effect disappeared when age of acquisition (AoA) was entered as a covariate, that is, the age at which a word has been learned (but, see Strain, Patterson & Seidenberg, 2002, for a commentary). These results are supported by Brysbaert et al. (2000) who demonstrated for Dutch that both in naming and in lexical decision, no imageability effect emerged if the words were matched for AoA and frequency. Shibahara et al. (2003) also suggested that the imageability effect of Strain et al. may have been the result of a confounding variable, that is, age of acquisition. They used the stimuli of Seidenberg et al.'s Experiments 1 and 2 and included age of acquisition as covariate. They found a main effect of imageability, and again, the main effect of imageability disappeared with age of acquisition entered as covariate, but the interaction of regularity by imageability in low-frequency words was still significant. Baluch and Besner (2001) questioned the specificity of the imageability effect for low-frequency words. They studied the imageability effect in the Persian orthography, which includes opaque words (words without specified diacritics) and transparent words (word in which the vowel is a fixed part of the spelling). Both high- and low-frequency unambiguous opaque words and transparent words were presented in a speeded naming task. Results demonstrated an imageability effect for opaque words, but not for transparent words. Moreover, in opaque words, both high-frequency and low-frequency words were affected by imageability. In sum, there is evidence that suggests that a semantic factor like imageability influences word recognition. However, the evidence is not fully convincing and is affected by other variables.

Language: English versus non-English studies

For example, the imageability effect may be different for languages other than English. The orthography in English is rather inconsistent, which makes a correct grapheme to phoneme mapping more difficult than in a transparent language. Thus, the weights between semantics, orthography and phonology might be different in deep orthographies and shallow orthographies, resulting in contradictory findings concerning the role of imageability. As said earlier, in the study of Baluch and Besner (2001), latencies for transparent words were not affected by imageability, whereas in opaque words, imageability effects were demonstrated.

Shibahara et al. (2003) showed that imageability effects in reading Japanese Kanji (deeper orthography than English) were more obvious than in reading English. Raman (2000) and Raman and Baluch (2001) studied the effect of imageability in Turkish, a highly transparent orthography. The alphabet consists of 29 letters, all corresponding to only one phoneme. In one study (Raman & Baluch, 2001), adult participants were presented words in a speeded-naming task, manipulated on imageability and frequency. There was no main effect of imageability. In a second experiment, the results of skilled readers were contrasted with those of very skilled readers. When reading skill was taken into consideration, latency analyses yielded significant interactions between skill and imageability and a significant skill by frequency by imageability interaction; in very skilled readers, low-frequency high-imageability words were named faster than low-frequency low-imageability words. Thus, semantics in the Turkish orthography only seemed involved in very skilled readers. Barca, Burani, and Arduino (2002) investigated word-naming times for native adult speakers of Italian, a relatively shallow orthography. Thirty students had to name as quickly and as accurately as possible words in several blocks. It turned out that a semantic factor including age of acquisition, imageability, and concreteness did not have any effect on naming latencies when other factors (frequency, length, and mean bigram frequency) were controlled.

De Groot (1989) investigated the role of imageability in Dutch readers in both lexical decision and word pronunciation. Targets were words, varied by imageability and frequency. In lexical decision, the effect of imageability was significant (shorter reaction times for high-imageability words) and interacted significantly with frequency on the subjects analysis: Imageability effects were only present in low-frequency words. In the naming task, only a very small effect of imageability was demonstrated. In addition, this imageability effect did not interact with frequency. Thus, in a transparent orthography, imageability effects seem to be smaller than in a deeper orthography. Moreover, effects seem dependent on the type of task that has been used (lexical decision versus naming).

Participants: Adults versus Children

In the present study, we wanted to extend the findings on imageability for Dutch, a relatively transparent language. Specifically, we were interested in imageability effects in young readers. This way, we could examine semantic effects while the reading system is still developing. Studies of imageability in children have been carried out less frequently.

Schwanenflugel and Akin (1994) ran a lexical-decision task and found reliable imageability effects with respect to latencies for both second or third graders and adults. Moreover, there was a significant age by imageability interaction, indicating that the imageability effect was much more pronounced for children than it was for adults. Baddeley, Ellis, Miles, and Lewis (1982) also studied semantic effects in word recognition in children and found imageability effects in both children with dyslexia, chronological-age controls, and reading-age controls. The authors concluded that the imageability effect is a characteristic of normal reading, rather than of dyslexic reading.

However, Coltheart, Laxon, Keating, and Pool (1986) studied the imageability effect in naming accuracy in children in Grade 2 and found the effect only to be significant in average or poor readers. A study by Coltheart, Laxon, and Keating (1988) reported similar results in 47 children, performing an oral-reading task. Imageability effects were only significant on accuracy for poor readers and the effect of imageability on speed was significantly greater for poor readers than for good readers (participants were assigned to a better-readers group or a poor-readers group, based on results of a reading comprehension test). These results are in line with the results of Strain and Herdman (1999), who extended the findings of Strain et al. (1995) and investigated whether imageability effects were the same in different types of adult readers. They distinguished three groups of readers on the basis of their phonological skills: A high-skill, medium-skill, and low-skill group. All participants read aloud as quickly as possible a set of words, manipulated on imageability and regularity. In the high-skill and medium-skill group, significant effects existed in naming latencies for imageability and regularity and an interaction effect between these variables was demonstrated, indicating a larger effect of imageability for exception words. However, in the low-skill group, only a main effect of imageability was found, which indicates that imageability effects were equally strong for both types of words. The authors concluded that the effect of imageability increased with decreasing phonological skills.

Other researchers found no effect of imageability in children at all. In a pilot-study (Raman, 2000), beginning readers of Turkish received a speeded-naming task, including high-imageability and low-imageability words. Accuracy scores revealed no significant effect of imageability. Porpodas, Pantelis, and Hantziou (1990) investigated the imageability effect in Greek. First graders performed a naming task with HI-nouns and LI-nouns. Again, results revealed no significant difference in both groups between reading HI-words and LI-words. Schwanenflugel and Noyes (1996) investigated imageability effects in naming and lexical decision in second graders (non-speeded task), third graders (speeded task), and fifth-graders (speeded task) and found only very limited effects of word imageability once word length and frequency were controlled for. In the current study, we explored imageability effects in both lexical decision and naming in Dutch children from Grades 2 to 6.

To conclude, imageability effects seem dependent on word characteristics, population, type of reading task (naming versus lexical decision), and type of outcome measure (accuracy versus latencies). We were interested in the impact of these variables on the magnitude of imageability effects in a developmental perspective. In Experiment 1, we conducted a lexical-decision task and manipulated word frequency and imageability. Readers from Grades 2 to 6 performed both speeded and non-speeded lexical decisions on these words and on nonwords. In Experiment 2, the same participants as in Experiment 1 performed a speeded and a non-speeded naming task on the items. The study is an attempt to answer the following questions: (a) Is visual word perception in primary grades affected by imageability? (b) What is the effect of word frequency? We hypothesized that the imageability effect would only emerge in low-frequency words, a finding that is most common in the literature with adult participants (e.g., de Groot, 1989; Kroll & Merves, 1986; Raman & Baluch, 2001; Shibahara et al., 2003; Strain et al., 1995). (c) Do the results differ for a speeded task and a non-speeded task? We assumed that the effect would be larger in a non-speeded task than in a speeded task, because in a non-speeded task, more time is available for semantics to affect reading performance. (d) Are the effects different across grades? Effects were studied across grades, because we were interested in developmental changes and because of equivocal findings concerning semantic effects in beginning readers. (e) Do effects of imageability vary across tasks (lexical decision vs. naming)? We expected the imageability effect to be more pronounced in the lexical-decision task, because lexical decision most likely involves minimal knowledge of word meanings (e.g., Bosman & de Groot, 1996; Chumbley & Balota, 1984), a factor that is not required in naming.

Experiment 1: Lexical Decision

Method

Participants

In this study, 233 students, 107 boys (45.9%) and 126 girls (54.1%), from four regular-primary schools participated. The schools were localized in three different regions in the Netherlands, that is, Brabant (Oss), Overijssel (Oldenzaal), and Gelderland (Bergharen and Afferden). The group consisted of a heterogeneous sample of children of middle-class socio-economic status. The children were recruited from Grades 2 to 6. The mean age was 10.3 years (SD = 1.5), ranging from 7.5 years in Grade 2 to 13.3 years in Grade 6.

Materials

The stimuli in this experiment comprised 120 words and 120 nonwords. Half of the words were high-imageability words (HI-words) and the other half constituted a set of low-imageability words (LI-words). The imageability ratings were based on the ratings from van Loon-Vervoorn (1985), which were collected by means of a seven-point scale (1= low imageability; 7= high imageability). Words with imageability ratings between 6.0 and 7.0 were classified as HI-words; words with ratings between 1.0 and 4.5 were classified as LI-words. We decided to omit words in the middle range, to maximize the contrast between HI-words and LI-words.

In addition, to control for frequency effects, each set of HI- and LI-words comprised high-frequency words (HF-words) and low-frequency words (LF-words). Frequencies were defined by the ratings of Schrooten and Vermeer (1994). These ratings are based on a wide range of words from both a written corpus (illustrated books, reading books, and school-related books) and a corpus of spoken language (language of teachers in the classroom) reflecting language input in the age of 4 to 12 years. Ratings were transformed to log frequencies. Half of the items below the median log frequencies (1.84) were designated LF-words. The set of items above the median was designated HF-words.

All words were controlled for word length, familiarity, number of syllables, and positional trigram frequency and letter frequency. Familiarity was based on the ratings of Schaerlaekens, Kohnstamm, and Lejaegere (1999). These ratings reflect the estimates of a representative sample of teachers, who were asked whether a 6-year-old child should know the words (receptive vocabulary). Morrison, Chappell, and Ellis (1997) have suggested that adult ratings provide a valid estimate of actual age of acquisition. To ensure that the majority of children was familiar with the items, only words with ratings up or above 75% were included in this study. Note, that children in our study were at least one year older than those, for which the ratings were estimated. Positional-trigram frequencies and positional letter frequencies were based on the ratings of Rolf and Van Rijnsoever (1984). All four different word types (HI-HF, HI-LF, LI-HF, and LI-LF) were evenly distributed over experimental blocks. Word characteristics are presented in Table 1.

For detailed information, we refer to Appendix A, which includes word characteristics for all words separately. To evaluate the correctness of the matching procedure, we performed a one-way ANOVA with word type as independent factor and word characteristics (trigram frequency, letter frequency, number of syllables, length, age of acquisition, frequency, and imageability) as dependent factor. Results demonstrated both in the speeded task and non-speeded task no differences between word types in mean trigram frequency, positional letter frequency, number of syllables, length, and age of acquisition; all F's < 1.5. In addition, imageability ratings and frequency ratings differed significantly in the intended direction; speeded task: imageability $F(3,59) = 208.19$, $p < .001$, frequency, *F*(3,59) = 30.56, *p* < .001; non-speeded task: imageability *F*(3,59) = 215.70, *p* < .001, frequency, *F*(3,59) = 40.29, *p* < .001.

Table 1. Descriptive Statistics of the Words in the Lexical-Decision and Naming Tasks

	Trigram frequency	Positional Letter frequency	Length (letters)	Number of syllables	AoA	Imageability	Frequency (log)
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Speeded							
HI-HF words	197.61 (208.70)	5702 (2633)	5.73 (1.67)	1.73 (0.59)	89.67 (5.04)	6.41 (0.25)	2.23 (0.26)
HI-LF words	525.07 (1131.65)	7233 (3922)	5.47 (1.36)	1.60 (0.51)	85.87 (8.68)	6.50 (0.18)	1.40 (0.28)
LI-HF words	223.52 (146.14)	6467 (5634)	5.60 (1.40)	1.67 (0.49)	88.0 (7.94)	3.72 (0.65)	2.40 (0.47)
LI-LF words	97.68 (106.83)	4831 (1948)	5.53 (1.60)	1.47 (0.52)	88.18 (7.32)	3.93(0.39)	1.35 (0.46)
Non-speeded							
HI-HF words	140.53 (191.05)	5225 (2331)	5.53 (1.41)	1.67 (0.49)	87.93 (5.80)	6.44 (0.27)	2.28 (0.32)
HI-LF words	154.59 (178.83)	4364 (2408)	5.80 (1.70)	1.73 (0.59)	88.60 (8.13)	6.48 (0.26)	1.31 (0.39)
LI-HF words	206.92 (202.55)	6306 (4029)	5.67 (1.68)	1.67 (0.49)	89.0 (7.23)	3.43 (0.59)	2.39 (0.35)
LI-LF words	264.02 (381.03)	6294 (4086)	5.73 (1.49)	1.60 (0.51)	85.90 (7.05)	3.72 (0.54)	1.31 (0.39)

The nonwords were created by changing one or more graphemes of the target words at initial, medial, or final position, maintaining Dutch orthographically legal spellings that were pronounceable. The nonwords were matched on word length and trigram frequencies with the words (see Table 2), and no homophones or pseudohomophones were included. Appendix B provides an overview of the characteristics of all nonwords.

Table 2. Descriptive Statistics of the Words and Nonwords in the Lexical-Decision and Naming Tasks

		Instruction	
		speeded	non-speeded
		M (SD)	*M (SD)*
trigram frequency			
	words	253(551)	193 (254)
	nonwords	209 (532)	114 (138)
length			
	words	5.6 (1.5)	5.7 (1.5)
	nonwords	5.6 (1.4)	5.7 (1.6)

Procedure

Two different lexical-decision tasks were employed, a speeded and a non-speeded one. Both tasks comprised 60 words and 60 nonwords. A word or nonword that appeared in the speeded task did not appear in the non-speeded task and vice versa, to preclude possible repetition effects, that is, the improvement of speed or accuracy as a result of repeated exposures to a stimulus. In both conditions, the stimuli were visually presented on the screen of a laptop using E-prime (Schneider, Eschman, & Zuccolotto, 2002). Words appeared in white on a black background centrally on the screen in lowercase bold 18-point Courier New font. Each stimulus item was preceded by a fixation stimulus (a plus sign) during 500 ms. Immediately after the offset of the fixation stimulus, the target stimulus was presented. In the non-speeded condition, the stimulus remained visible until a response was given. In the speeded condition, the stimulus disappeared after 2500 ms in Grade 2 and after 2000 ms in Grades 3 to 6. There was a 1000 ms delay between the response and the onset of the next trial. Each participant was presented with a different random order of three blocks of 40 items each. There was a break between two blocks as long as was required by the participant. Participants indicated their responses by pressing one of two keys (z and slash keys on the computer keyboard). These keys were marked by tags with a green or blue circle. Participants used their dominant hand to signal 'word', and the non-dominant hand to signal 'nonword'. Reaction times were measured from the onset of each word until a response was given. In the speeded task, the instruction stressed speed over accuracy. In the non-speeded task, accuracy was stressed over speed. Five practice items preceded the experiment proper.

Results

Reaction times shorter than 100 ms (4 observations) and longer than 5000 ms (0.2% in words and 0.8% in nonwords) were excluded from the analyses. In addition, incorrect trials (9.2% in the speeded lexical-decision task and 5.8% in the non-speeded task) and missing values, that is, no response within the restricted time period (3.4%) in the speeded task were considered errors and were also excluded from the latencies analyses. Two children were excluded from the analyses because of extreme high error rates for words in the non-speeded task (43.3% and 91.7%) and for nonwords in the speeded task (71.7% and 75%). Mean lexical-decision latencies for correct responses and mean percentage of errors were calculated across subjects, and are shown in Table 3. In all analyses, significant effects are based on a .05 alpha level.

Table 3. Mean Reaction Times (ms) and Error Rates (%) in the Lexical-Decision Task

	All children (*n* =231)		Lower grades (*n* = 87)		Higher grades (*n* = 144)	
	Mean RT (*SD*)	% errors	Mean RT (*SD*)	% errors	Mean RT (*SD*)	% errors
Speeded						
HI-HF words	961 (202)	6.1 (7.5)	1101 (202)	9.1 (9.4)	877 (149)	4.3 (5.3)
HI-LF words	988 (223)	9.9 (10.2)	1141 (233)	14.9 (10.9)	895 (155)	6.9 (8.5)
LI-HF words	937 (204)	6.4 (6.1)	1072 (202)	7.9 (6.6)	856 (157)	5.5 (5.6)
LI-LF words	1000 (226)	11.5 (9.5)	1163 (227)	15.2 (10.9)	901 (158)	9.3 (7.9)
Non-speeded						
HI-HF words	1176 (355)	2.7 (4.3)	1368 (385)	3.7 (4.6)	1061 (280)	2.2 (3.9)
HI-LF words	1284 (460)	5.4 (7.3)	1562 (502)	7.9 (9.1)	1116 (336)	3.8 (5.4)
LI-HF words	1201 (410)	2.5 (4.6)	1427 (451)	3.8 (5.4)	1065 (313)	1.7 (3.8)
LI-LF words	1314 (441)	6.0 (8.5)	1584 (472)	9.5 (10.7)	1150 (328)	3.9 (5.9)

First, we excluded performance on nonwords from further analyses. For each participant, mean reaction times and percentages of errors were calculated for all four word-stimuli conditions in both the speeded and non-speeded condition (F_1). Mean latencies and error rates based on the experimental word trials were submitted to separate ANOVA's (GLM repeated measures). Correlation analyses indicated a significant relationship between reaction times and accuracy; the smaller the reaction time, the smaller the percentage of errors. Because of our interest in development, we distinguished between lower grades (Grades 2 and 3) and higher grades (Grades 4, 5, and 6) separately. We distinguished between these groups because explicit reading instruction ends in Grade 3. After that time, the focus of reading instruction has been shifted towards reading comprehension. Also, for each of the 120 words, mean reaction times and error percentages were calculated, collapsed across participants (F_2).

Reaction times

A 2 (Grade: lower vs. higher) x 2 (Instruction: speeded vs. non-speeded) x 2 (Imageability: low vs. high) x 2 (Frequency: low vs. high) ANOVA was performed on reaction times with grade as between-subjects factor and instruction, imageability, and frequency as within-subjects factors (subject analysis). Also, a 2 x 2 x 2 x 2 repeated measures analysis was performed on reaction times with frequency, imageability, and instruction as between-subjects factor and grade as within-subjects factor (item analysis).

A three-way interaction between instruction, frequency and grade was significant in the subject analysis, $F_1(1,229) = 9.55$, *partial* $\eta^2 = .04$, and marginally significant in the item analysis, $F_2(1,112) = 3.50$, *partial* $\eta^2 = .03$, $p = .06$. Subsequent analyses on the interaction effect for the lower grades and higher grades separately showed that the interaction effect between frequency and instruction was significant in both grade levels, however, the effect was larger for the lower grades, $F_1(1,86) = 24.14$, *partial* $\eta^2 = .22$, than for the higher grades, $F_1(1,143) = 10.33$, *partial* $\eta^2 = .07$.

The interaction between instruction and frequency was significant in the subject analysis, $F_1(1,229) = 40.90$, *partial* $\eta^2 = .15$, and marginally significant in the items analysis, $F_2(1,112) = 3.48$, *partial* $\eta^2 = .03$ ($p = .06$); in both tasks, HF-words were responded to significantly faster than LF-words. However, effects were more pronounced in the non-speeded task ($F_1(1,229) = 139.80$, *partial* $\eta^2 = .38$) than in the speeded task ($F_1(1,229) = 108.48$, *partial* $\eta^2 = .32$).

The instruction by grade interaction appeared to be significant, $F_1(1,229) = 16.48$, *partial* $\eta^2 = .07$; $F_2(1,112) = 79.16$, *partial* $\eta^2 = .41$. Subsequent analyses on the effect of instruction performed for each grade level separately showed that reaction times in the non-speeded task were longer than those in the speeded task. This effect, however, was smaller (though significant) in the higher grades ($F_1(1,143) = 138.82$, *partial* $\eta^2 = .49$; $F_2(1,112) = 192.84$, *partial* $\eta^2 = .63$; difference is 216 ms) than in the lower grades ($F_1(1,86) = 99.88$, *partial* $\eta^2 = .54$; $F_2(1,112) = 203.39$, *partial* $\eta^2 = .65$; difference is 366 ms).

The interaction effect between frequency and grade was also significant, $F_1(1,229) = 38.53$, *partial* $\eta^2 = .14$; $F_2(1,112) = 18.33$, *partial* $\eta^2 =.14$. Separate analyses showed that in both grades, reaction times for HF-words were significantly shorter than those for LF-words. However, children in the lower grades were more affected by word frequency ($F_1(1,86) = 132.52$, *partial* $\eta^2 = .61$; $F_2(1,112) = 23.09$, *partial* $\eta^2 = .17$; difference is 121 ms) than children in the higher grades ($F_1(1,143) = 73.54$, *partial* $\eta^2 = .34$; $F_2(1,112) = 11.22$, *partial* $\eta^2 = .09$; difference is 51 ms).

With respect to the semantic effects, the four-way interaction between instruction, frequency, imageability, and grade was significant in the subject analysis ($F_1(1,229) = 4.53$, *partial* $\eta^2 = .02$, though this effect was not significant by items ($F_2(1,112) = 2.21$, $p = .14$). Subsequent analyses on the interaction effect for the lower grades and higher grades separately showed that in the lower grades, the three-way interaction between instruction, frequency, and imageability was significant ($F_1(1,86) = 4.56$, *partial* $\eta^2 = .05$). In the higher grades, however, the interaction did not reach significance, $F_1 < 1$.

The interaction between instruction, frequency, and imageability was marginally significant in the subjects analysis ($F_1(1,229) = 3.99$, *partial* $\eta^2 = .02$, $p = .047$), though this interaction was not significant by items ($F_2 < 1$). The analysis on the frequency by imageability effect performed separately for the speeded task and the non-speeded task showed that the interaction effect was significant in the speeded task ($F_1(1,229) = 14.61$, *partial* $\eta^2 = .06$), but was not significant in the non-speeded task, $F_1 < 1$. In the speeded task, the imageability effect was significant in HF-words ($F_1(1,230) = 14.08$, *partial* $\eta^2 = .06$): In contrast to the expected pattern, HI-words were responded to slower than LI-words. In LF-words, the imageability effect was not significant ($F_1(1,230) = 2.71$).

Also, there was a significant instruction by imageability interaction in the subject analysis, $F_1(1,229) = 15.86$, *partial* $\eta^2 = .07$, though this interaction was not significant by items ($F_2 < 1$). The analysis on imageability effect performed separately for the speeded task and the non-speeded task showed that the imageability effect was significant in the non-speeded task, ($F_1(1,229) = 12.56$, *partial* $\eta^2 = .05$), but not in the speeded task ($F_1(1,229) = 1.55$, $p = .22$). In the non-speeded task, HI-words were responded to faster than LI-words.

Among the main effects, there was a significant effect of instruction ($F_1(1,229) = 248.11$, *partial* $\eta^2 = .52$; $F_2(1,112) = 221.58$, $\eta^2 = .66$), frequency ($F_1(1,229) = 235.77$, *partial* $\eta^2 = .51$; $F_2(1,112) = 20.08$, *partial* $\eta^2 = .15$), and grade ($F_1(1,229) = 90.21$, *partial* $\eta^2 = .28$; $F_2(1,112) = 1468.18$, *partial* $\eta^2 = .93$. The main effect of imageability was significant in the subject analysis, $F_1(1,229) = 5.40$, *partial* $\eta^2 = .02$, but not in the item analysis, $F_2 < 1$. Reaction times were faster for the speeded task than for the non-speeded task, faster for HF-words than for LF-words, faster for the higher grades than for the lower grades, and finally, faster for HI-words than for LI-words.

To summarize, the present findings demonstrated a main effect of imageability in reaction times: Reaction times in HI-words were shorter than those in LI-words. A significant interaction effect, however, between instruction and imageability, revealed that the imageability advantage was only significant in the non-speeded task. These results are in accordance with our assumption that in a non-speeded task, more time is available for semantics to affect reading performance. Finally, a significant 3-way interaction between instruction, frequency, and imageability and a significant 4-way way interaction between instruction, frequency, imageability and grade demonstrated an unexpected finding in the lower grades. In these grades, there was a significant interaction between frequency and imageability in the speeded task; surprisingly, in high-frequency words, reaction times were longer for HI-words than for LI-words. Thus, the hypothesis that imageability effects would only occur in LF-words was not supported by the present findings.

Accuracy

A 2 (Grade: lower vs. higher) x 2 (Instruction: speeded vs. non-speeded) x 2 (Imageability: low vs. high) x 2 (Frequency: low vs. high) ANOVA was performed on percentages of errors with grade as between-subjects factor and instruction, imageability and frequency as within-subjects factor (subject analysis). Also, a 2 x 2 x 2 x 2 repeated measures analysis was performed on percentages of errors with frequency, imageability, and instruction as between-subjects factor and grade as within-subjects factor (item analysis).

The instruction by grade interaction was significant, $F_1(1,229) = 7.57$, *partial* $\eta^2 = .03$; $F_2(1,112) = 5.42$, *partial* $\eta^2 = .05$. Subsequent analyses of the effect of instruction performed for each grade level separately showed that in both grade levels, significantly more errors were made in the speeded task than in the non-speeded task. The effect, however, was smaller in the higher grades ($F_1(1,143) = 74.93$, *partial* $\eta^2 = .34$; $F_2(1,112) = 9.76$, *partial* $\eta^2 = .08$, difference is 3.6%) than in the lower grades ($F_1(1,86) = 79.46$, *partial* $\eta^2 = .48$; $F_2(1,112) = 16.01$, *partial* $\eta^2 = .13$, difference is 5.5%).

The interaction effect between frequency and grade was also significant, $F_1(1,229) = 22.39$, *partial* $\eta^2 = .09$; $F_2(1,112) = 13.70$, *partial* $\eta^2 = .11$. Separate analyses showed that in both grade levels, significantly more errors were made in LF-words than in HF-words. However, children in the lower grades were more affected by word frequency ($F_1(1,86) = 75.70$, *partial* $\eta^2 = .47$; $F_2(1,112) = 17.13$, *partial* $\eta^2 = .13$, difference is 5.7%) than children in the higher grades ($F_1(1,143) = 62.63$, *partial* $\eta^2 = .31$; $F_2(1,112) = 5.16$, *partial* $\eta^2 = .04$, difference is 2.6%).

The instruction by frequency interaction was significant in the subject analyses ($F_1(1,229) = 7.06$, *partial* $\eta^2 = .03$), but not in the item analyses ($F_2 < 1$). Separate analyses on the effect of frequency for the speeded task and non-speeded task showed that frequency effects were significant in both tasks. However, the effect was more pronounced in the speeded task ($F_1(1,229) = 100.17$, *partial* $\eta^2 = .30$, difference is 4.9%) than in the non-speeded task ($F_1(1,229) = 92.89$, *partial* $\eta^2 = .29$, difference is 3.4%).

With respect to the semantic effects, there was a significant three-way interaction between instruction, imageability and grade in the subject analysis, $F_1(1,229) = 7.91$, *partial* $\eta^2 = .03$, and a marginally significant interaction effect in the item analysis, $F_2(1,112) = 3.87$, $p = .05$, *partial* $\eta^2 = .03$; in the lower grades, no significant interaction between instruction and imageability was revealed, whereas in the higher grades, the interaction effect between instruction and imageability reached significance in the subject analysis, $F_1(1,143) = 11.61$, *partial* $\eta^2 = .08$, but not in the item analysis, $F_2 < 1$; In the speeded task, fewer errors were made in HI-words than in LI-words ($F_1(1,143) = 13.99$, *partial* $\eta^2 = .09$), whereas in the non-speeded task, the imageability effect was not significant.

The interaction between imageability and frequency was significant in the subject analysis, $F_1(1,229) = 4.50$, *partial* $\eta^2 = .02$, but not in the item analysis, $F_2 < 1$. Separate analyses for LF-words and HF-words showed that the imageability effect was only significant in LF-words, $F_1(1,229) = 5.20$, *partial* $\eta^2 = .02$; less errors were made in low-frequency HI-words than in low-frequency LI-words.

Among the main effects, there was a significant effect of instruction, $F_1(1,229) = 161.29$, *partial* $\eta^2 = .41$; $F_2(1,112) = 14.49$, *partial* $\eta^2 = .12$, frequency, $F_1(1,229) = 157.69$, *partial* $\eta^2 = .41$, $F_2(1,112) = 12.11$, *partial* $\eta^2 = .10$, and grade, F_1 $(1,229) = 63.09$, *partial* $\eta^2 = .22$, $F_2(1,112) = 102.83$, *partial* $\eta^2 = .48$. In the speeded task, 4.6% more errors were made than in the non-speeded task, in HF-words, 4.2% fewer errors were made than in LF-words, and mean percentages of errors in the higher grades was 4.3% less than those in the lower grades. The main effect of imageability was not significant, $F_1(1,229) = 2.60$, $p = .11$; $F_2 < 1$.

To summarize, in contrast to results in the latencies analyses, no significant main effect of imageability was found in the error analyses. However, a significant interaction effect between imageability and frequency demonstrated that in low-frequency words, more errors were made in LI-words than in HI-words, whereas in high-frequency words, no imageability effect was found. This result is consistent with our prediction that imageability effects would only occur in low-frequency words. Finally, there was a significant 3-way interaction between instruction, imageability and grade, indicating that only in the higher grades, a significant interaction between instruction and imageability was found. In these grades, more errors were made in LI-words than in HI-words, but only in the speeded task.

Experiment 2: Naming

Method

Participants and stimulus materials

The participants and materials were identical to those of Experiment 1. Recall, each participant was presented a speeded task and a non-speeded task, with different items in both tasks. Items that were instructed in a speeded paradigm in the lexical-decision task were also instructed in a speeded paradigm in the naming task.

Procedure

Eight different orders of the items were created. The items of each list were printed in four columns on a single sheet of A4-paper, in black on a white background in lowercase 12-point Times New Roman. The participants were asked to read the items as accurately and as fast as possible in the speeded-naming task. The instruction stressed speed over accuracy. In the non-speeded task, accuracy was stressed over speed. The order of the tasks was varied. Children were tested twice. At each test moment, they performed a lexical-decision task (Experiment 1) and a naming task (Experiment 2). One of the tasks was a speeded task, the other was a non-speeded task. This way, children did not see the same words twice a day, because the speeded task and non-speeded task comprised different items.

We did not assess reaction times per item, because words were not presented on the computer. Moreover, a latencies analysis would require to control for speech onset times and thus needs pairing all words by phonetic onset, which causes additional problems to the matching of the stimuli. As a result, we only analyzed error scores.

Results

Mean percentages of errors were calculated across subjects and are shown in Table 4.

Table 4. Mean Error Rates (%) and Standard Deviations in the Naming Task

	All children (*n* = 233)	Lower grades (*n* = 89)	Higher grades (*n* = 144)
Speeded			
HI-HF words	3.2 (6.6)	6.6 (9.0)	1.0 (3.1)
HI-LF words	*2.9 (6.2)*	*5.6 (8.5)*	*1.2 (3.1)*
LI-HF words	1.3 (3.9)	2.0 (5.5)	.8 (2.4)
LI-LF words	*3.3 (6.7)*	*5.5 (8.7)*	*1.9 (4.6)*
Non-speeded			
HI-HF words	1.6 (3.9)	2.9 (5.3)	.8 (2.4)
HI-LF words	2.1 (5.3)	4.2 (7.5)	.8 (2.5)
LI-HF words	2.0 (5.6)	4.3 (8.3)	.7 (2.0)
LI-LF words	2.3 (5.9)	4.6 (8.6)	.9 (2.5)

In all analyses, significant effects are based on a .05 alpha level. First, we excluded performance on nonwords from further analyses. For each participant, percentages of errors were calculated for all four word-stimuli conditions in both the speeded and non-speeded condition.

Accuracy

A 2 (Grade: lower vs. higher) x 2 (Instruction: speeded vs. non-speeded) x 2 (Imageability: low vs. high) x 2 (Frequency: low vs. high) ANOVA was performed on percentages of errors with grade as between-subjects factor and instruction, imageability, and frequency as within-subjects factor (subject analysis). Also, a 2 x 2 x 2 x 2 repeated measures analysis was performed on percentages of errors with frequency, imageability, and instruction as between-subjects factor and grade as within-subjects factor (item analysis).

There was a significant 4-way interaction between instruction, frequency, imageability, and grade in the subject analysis, $F_1(1,231) = 10.27$, *partial* $\eta^2 = .04$, though the analysis did not reach significance in the item analysis, $F_2(1,112) = 2.92$, $p = .09$; Analyses performed for the lower grades and higher grades separately showed that in the lower grades, the interaction of instruction by frequency by imageability was significant, F_1 (1,88) = 11.04, *partial* $\eta^2 = .11$, whereas in the higher grades it was not.

Also, there was a significant three-way interaction between instruction, frequency and imageability in the subject analysis, $F_1(1,231) = 16.43$, *partial* $\eta^2 = .07$, but not in the item analysis, $F_2(1,112) = 2.66$, $p = .11$. Analyses performed for the speeded task and non-speeded task separately showed that in the speeded task, the interaction between frequency and imageability was significant, $F_1(1,231) = 19.29$, *partial* $\eta^2 = .08$; In HF-words, significantly more errors were made in HI-words, whereas in LF-words, no significant difference was found. In the non-speeded task, the interaction was not significant, $F_1 < 1$.

The three-way interaction of instruction by imageability by grade was significant, $F_1(1,231) = 22.66$, *partial* $\eta^2 = .09$; $F_2(1,112) = 5.96$, *partial* $\eta^2 = .05$. Analyses performed for the lower grades and higher grades separately showed that in the lower grades, the interaction of instruction by imageability was significant in the subject analysis, $F_1(1,88) = 17.37$, *partial* $\eta^2 = .17$, and marginally significant in the item analysis, $F_2(1,112) = 3.97$, *partial* $\eta^2 = .03$, $p = .05$; In the speeded task, significantly more errors were made in HI-words than in LI-words, whereas in the non-speeded task, no significant differences were revealed. In the higher grades, the instruction by imageability interaction was not significant, $F_1 < 1$, $F_2 < 1$.

The interaction of imageability by grade was significant in the subject analysis, $F_1(1,231) = 4.19$, *partial* $\eta^2 = .02$, but not in the item analysis, $F_2(1,112) = 1.19$, $p = .28$. Subsequent analyses of the effect of imageability performed for each grade level separately suggested that in the lower grades, more errors were made in HI-words than in LI-words, whereas in the higher grades, more errors were made in LI-words. However, the statistical analyses indicated no significance.

The interaction between instruction and imageability was also significant in the subject analysis, $F_1(1,231) = 16.03$, *partial* $\eta^2 = .07$, but not in the item analysis, $F_2(1,112) = 2.40$, $p = .12$. Subsequent analyses of the effect of imageability performed for the speeded task and non-speeded task separately showed that in the speeded task, the imageability effect was significant, $F_1(1,231) = 14.55$, *partial* $\eta^2 = .06$; in HI-words, more errors were made than in LI-words. In the non-speeded task, there was no significant imageability effect.

Next, the interaction between frequency and imageability was significant in the subject analysis, $F_1(1,231) = 10.13$, *partial* $\eta^2 = .04$, but not in the item analysis, $F_2(1,112) = 1.44$, $p = .23$. Subsequent analyses of the effect of imageability performed for the HF-words and LF-words separately showed that the effect was only significant in HF-words, $F_1(1,231) = 15.87$, *partial* $\eta^2 = .06$; in HI-words, significantly more errors were made than in LI-words.

Among the main effects, there were significant main effects in the subject analysis for instruction, $F_1(1,231) = 15.83$, *partial* $\eta^2 = .06$; frequency, $F_1(1,231) = 12.06$, *partial* $\eta^2 = .05$; grade, $F_1\ (1,231) = 42.82$, *partial* $\eta^2 = .16$. The main effect of grade was also significant in the item analysis, $F_2(1,112) = 92.76$, *partial* $\eta^2 = .45$. The main effect of imageability was not significant, $F_1(1,231) = 2.83$, $p = 0.09$; $F_2 < 1$. In the speeded task, more errors were made than in the non-speeded task; in HF-words, fewer errors were made than in LF-words; in the higher grades less errors were made than in the lower grades.

To summarize, similar to the error analyses in the lexical-decision task, no significant main effect of imageability was found in the error analyses in the naming task. A significant 4-way interaction effect between instruction, frequency, imageability, and grade, however, revealed an unexpected pattern; In the lower grades, there was a significant interaction between instruction, frequency, and imageability. It turned out that in these grades, only in the speeded task, a significant frequency by imageability effect was found; In high-frequency word, more errors were made in HI-words than in LI-words.

General discussion

The present study showed that reading performance in children can be influenced by imageability under certain circumstances, which implies that semantics does affect printed word perception in young readers. However, the effects are small and dependent on word characteristics (e.g., low-frequency words vs. high-frequency words), instruction (speeded vs. non-speeded), the age of the participants (lower grades vs. higher grades), and the type of reading task (naming vs. lexical decision). Several results provide evidence for these suggestions.

First, in lexical decision, the imageability effect depended on word frequency: The interaction between imageability and word frequency was only significant in the speeded task and only in HF-words: HI-words were responded to slower than LI-words. In LF-words, the imageability effect was not significant. This result contradicts previous studies with adult participants which demonstrated an imageability advantage in LF-words (e.g., de Groot, 1989; Kroll & Merves, 1986; Raman & Baluch, 2001; Shibahara et al., 2003; Strain et al., 1995).

In addition, several researchers (e.g., Cortese et al., 1997; Shibahara et al., 2003; Strain et al., 1995; Strain & Herdman, 1999) showed a significant interaction between imageability and regularity in naming tasks, with imageability effects being larger for exception words. It has been concluded from these studies (e.g., Strain et al., 1995) that the effect of semantics increases when phonological processing is slow. In our study, regularity and consistency effects could not be examined, because of the rather consistent grapheme-to-phoneme mappings in Dutch.

Second, imageability effects were also dependent on task instruction: The present study demonstrated that speed manipulation was responsible for the emergence of imageability effects. Overall reaction times in lexical decision revealed a limited effect of imageability in the speeded task, whereas in the non-speeded task, HI-words were responded to significantly faster than LI-words. The overall error rates of children in the higher grades showed an interaction between instruction and imageability in the opposite direction: In the speeded task, fewer errors were made in HI-words than in LI-words, whereas in the non-speeded task, the advantage disappeared. Thus, in a lexical-decision task without time pressure, semantic effects were only observed in reaction times, whereas in a speeded task, semantic effects were only apparent in accuracy scores. Similarly, McFalls, Schwanenflugel, and Stahl (1996) demonstrated an effect of imageability in speeded lexical decision only in accuracy scores, not in latencies. However, the difference in accuracy for concrete and abstract words, 3.1% for words not appearing in children's reading curriculum and 15.7% for words taught in the reading program was much larger than in our study (about 1% overall difference in errors), albeit our results suggest that semantics plays a role in printed-word perception, even in speeded reading.

What are the theoretical consequences of this finding on processing speed? Some authors (e.g., Strain et al., 1995) assume that semantic activation builds up gradually, which reduces semantic effects in speeded word recognition. This assumption was supported by results of Strain et al.'s Experiment 3. Strain et al. showed that in speeded naming, more errors were made in (low frequency) high-imageability exception words than in the non-speeded naming condition. However, in (low-frequency) low-imageability exception words, no significant differences between errors in both conditions were found. The authors suggested that in the speeded task, the contribution of semantics was reduced and as a consequence, the advantage of high-imageability words disappeared, resulting in more errors in high-imageability words in speeded naming than in non-speeded naming. Because reading low-imageability words is not facilitated by semantics, speeding did not affect reading performance of these words. In our study, however, speeding negatively affected accuracy and reaction times of both HI-words and LI-words, with LI-words even more affected. Thus, the results of the present study do not support a graded semantic activation account. In the speeded task, reaction times were not affected by imageability, but accuracy was if participants were forced to speed their response.

Third, the effect of imageability depended on the amount of reading experience of participants. In the present study, in less experienced readers (Grades 2 and 3), no effect of imageability was found in accuracy in the lexical-decision task. In contrasts, in the higher grades (Grades 4, 5, and 6), imageability did affect lexical-decision performance. Thus, imageability only affected word recognition time in relatively experienced readers. Schwanenflugel and Akin (1994) and McFalls et al. (1996) however, who ran a lexical-decision task, found reliable imageability effects in second and third graders. Moreover, Schwanenflugel and Akin showed that the imageability effect was much more pronounced for children (221 ms) than it was for adults (32 ms), a result that is not to be expected from our results. At this point, we are uncertain about the factors that are responsible for the different results. In general, semantic effects may be more apparent in an irregular orthography like English, with a lot of irregular and exception words, than in a relatively regular orthography like Dutch. In an irregular orthography, semantic information might support opaque conversions from graphemes to phonemes early in childhood. The absence of imageability effects in the lower grades in Dutch is supported by literature on naming in other relatively transparent languages like Turkish (Raman, 2000) and Greek (Porpodas et al.,1990). Both researchers investigated the role of imageability in beginning readers and found no imageability effect.

Finally, the effect of imageability depended on the type of reading task that was used. Results of Experiment 2 revealed a different picture of the effect of imageability if reading is assessed by a naming task instead of a lexical-decision task. Accuracy scores showed that imageability only affected reading performance in the speeded task and only in high-frequency words. Moreover, this pattern was only significant in the lower grades and in the opposite direction of the lexical-decision task: Separate analyses showed that more errors were made in high frequency HI-words than in high frequency LI-words. This unexpected finding was also apparent in reaction times in the lexical-decision task. To explore possible explanations, we looked at the qualitative data. It turned out that proportionally many children made errors on the naming of ‘deken’ (blanket), a HF-HI word. Most children named the word as if it comprised a closed syllable, like ‘dekken’ (to cover) or ‘denken’ (to think). Words with open syllables often pose a problem four young children, because it deviates from the regular orthography to phonology mapping. Thus, we suggest that the unexpected finding of more errors in high frequency HI-words than in high frequency LI-words in the lower grades might be due to the Dutch spelling rules with respect to open and closed syllables. Although all stimuli were matched on a number of word characteristics, we did not control for this variable. Future research on imageability effects in reading performance of (Dutch) children needs to consider this issue. We want to stress, however, the difficulties in investigating semantic effects in beginning readers. Because of the very limited reading experience in these children, it is hard to find a representative amount of words that is matched on all relevant variables.

In short, imageability can affect word perception in Dutch children, but the effects are small and depend on word characteristics (HF-words versus LF-words), type of reading task (lexical decision versus naming), instruction and outcome measure (in a non-speeded lexical-decision task, semantic effects were only observed in reaction times, whereas in a speeded task, semantic effects were only apparent in accuracy scores), and participant characteristics (more prevalent in the higher grades than lower grades). Other researchers also demonstrated that experimental variables do affect the emergence of imageability effects (e.g., Kroll & Merves, 1986; Zevin & Balota, 2000). Kroll and Merves found effects of concreteness in lexical decision, but the effect was dependent on the order of presentation. Participants were first presented pure lists of one type of words (i.e., concrete words or abstract words) in a lexical decision task. Afterwards, they were presented pure list of the other type of words. It turned out that there was no impact of concreteness if abstract words preceded concrete words. Zevin and Balota (Experiment 4) investigated imageability effects in a naming task with adult skilled readers as participants. The task consisted of naming primes (LF-exception words and nonwords) and targets, mainly drawn from the lists reported in Strain et al. (1995) and Cortese et al. (1997). Results of target response latencies demonstrated that imageability effects were apparent for the LF-exception word-prime condition, but not for the nonword-prime condition. In short, imageability effects depended on the context. Zevin en Balota argued that readers can “bias their processing style at a very basic level to meet task demands within the context of an experiment” (p. 121), referred to as the ‘Attentional Control Hypothesis’.

The final and major question is: How can imageability effects be explained in current theories of visual word perception? One way to explain reading is by means of a symbolic, localist model with the Dual Route Model (e.g., Coltheart, 1978; Coltheart et al., 2001; Jackson & Coltheart, 2001) as prototype. This model has several slightly different representational forms, but the key assumption is the distinction of two different routes to the identification of a letter string: A lexical route and a non-lexical route. The lexical route assumes letter units activating an orthographic lexicon, which in turn, will send input to the phonological lexicon. The non-lexical route implies grapheme-phoneme rules to convert letter units into speech. The model has been implemented into a computational model, which is referred to as the Dual Route Cascaded Model (Coltheart, 2001), from now on DRC model. Several adaptations have been made tot the traditional model, to fit the model to experimental findings. In the DRC-model, the lexical route distinguishes a lexical-semantic route, in which activation flows from the orthographic lexicon via a semantic system to the phonological lexicon, and a lexical non-semantic route, without accessing a semantic system. Based on the results of Strain et al. (1995), it is suggested that the lexical semantic system can contribute to correct word identification, but only when the other two routes operate too slowly. This is the case in reading low-frequency irregular words, which causes competition between the lexical route and the non-lexical route. Semantic contribution is made possible, because activation flows in both directions. The DRC model can also account for the different results for beginning and advanced readers, which are found in our study. It is suggested that beginning readers mainly rely upon the non-lexical route, whereas with increasing reading experience, the lexical route will become dominant. Thus, only for skilled readers, the model predicts semantic contribution in visual word perception, which is supported by our results of the lexical-decision task.

Another way of explaining semantic effects is by means of distributed accounts of reading. Probably the most influential model of this type is the distributed, developmental model of word recognition by Seidenberg and McClelland (1989), from now on SM89 model. This model assumes that reading words involves the computation of three types of codes, i.e., orthographic, phonological, and semantic codes. Each of these codes is a distributed representation. In this view, a word is not a local unit in memory, like in the DRC model, but a pattern of distributed activation across nodes. In the SM89 model, the semantic nodes have not been implemented, so it is difficult to make assumptions about semantic effects, but the key assumption is that all three nodes are computed simultaneously. In regular and high-frequency words, the mappings from orthography to phonology are suggested to be too efficient to show semantic effects. However, in low-frequency words and exception words, the computation from orthography to phonology is inefficient and as a result, semantics may affect reading performance. The semantic pathway might improve with reading experience, explaining the absence of an imageability effect in the lower grades in our study. Several shortcomings of the implemented SM89 model (the model could not account well for reading nonwords) led to an adaptation of the connectionist model, developed by Plaut et al. (1996). In general, a framework of interactivity between orthography, semantics, and phonology has been put forward by several other researchers. The Phonological Coherence Model of Van Orden, Pennington, and Stone (1990), which focuses on the process of reading, is an example of a recurrent network that has been fully implemented (Farrar & Van Orden, 2001). Because of this recurrence, semantic codes can provide support in orthography to phonology mappings, especially when processing is relatively slowly, like in low-frequency words.

Thus, semantic effects can be explained in both localist and distributed accounts of visual word perception. We favor the distributed model, because in our view, this model provides a better explanation of development of reading skills. More research is needed to get better insight in a developmental shift in semantic effects. In addition to imageability, it would be useful to study other semantic manipulations in the visual word perception of young readers.

REFERENCES

Baddeley, A. D., Ellis, N. C., Miles, T. R., & Lewis, V. J. (1982). Developmental and acquired dyslexia: A comparison. *Cognition, 11*, 185-199.

Balota, D. A., Ferraro, F. R., & Connor, L. T. (1991). On the early influence of meaning in word recognition : A review of the literature. In Schwanenflugel, P. J., *The psychology of word meanings* (pp. 187-222). Hillsdale, NJ: Lawrence Erlbaum Associates.

Baluch, B., & Besner, D. (2001). Basic processes in reading: semantics affects speeded naming of high-frequency words in an alphabetic script. *Canadian Journal of Experimental psychology, 55,* 63-69.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers, 34,* 424-434.

Borowsky,R. & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 63-85.

Bosman, A. M. T., & de Groot, A. M. B. (1996). Phonologic mediation is fundamental to reading: Evidence from beginning readers. *The Quarterly Journal of Experimental Psychology, 49A, 715-744*

Bosman, A. M. T., & van Hell, J. G (2002). Orthography, phonology, and semantics: Concerted action in word perception. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 165-187). Amsterdam: John Benjamins.

Brysbaert, M., Lange, M., & Van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology, 12,* 65-85.

Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory and Cognition, 12,* 590-606.

Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology, 79,* 1-12.

Coltheart, V., Laxon, V. J., Keating, G. C., & Pool, M. M. (1986). Direct access and phonological encoding processes in children's reading: Effects of word characteristics. *British Journal of Educational Psychology, 56,* 255-270.

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review, 4,* 226-231.

de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(5),* 824-845.

Farrar, W. T., & Van Orden, G. C. (2001). Errors as multistable response options. *Nonlinear Dynamics, Psychology, and Life Sciences, 5*, 223-265.

Hino, Y. Lupker, S. J., & Pexman, P. M.(2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 686-713.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120*, 358-394.

Kroll, J. F., & Merves, J. S. (1986). Lexical access for concrete and abstract words. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 12*, 92-107.

Locker, JR L., Simpson, G. B., & Yates, M. (2003). Semantic neighborhood effects on the recognition of ambiguous words. *Memory & Cognition, 31*, 505-515.

Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with spelling-sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 183-206.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology, 50A*, 528-559.

Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology, 91*, 167-180.

Pecher, D. (2001). Perception is a two-way junction: Feedback semantics in word recognition. *Psychonomic Bulletin & Review, 8*, 545-551.

Pexman, P. M., & Lupker, S. J. (1999). Ambiguity and visual word recognition: Can feedback explain both homophone and polysemy effects? *Canadian Journal of experimental psychology, 53*, 323-334.

Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: Number-of-features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review, 9*, 542-549.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115.

Porpodas, C. D., Pantelis, S. N., & Hantziou, E. (1990). Phonological and lexical encoding processes in beginning readers: Effects of age and word characteristics. *Reading and Writing, 2*, 197-208.

Raman, I. (2000). Word imageability effects on naming: A pilot investigation of beginning readers of Turkish. *Perceptual and Motor Skills, 472-474.*

Raman, I., & Baluch, B. (2001). Semantic effects as a function of reading skill in word naming of a transparant orthography. *Reading and Writing, 14,* 599-614.

Rolf, P. C., & Van Rijnsoever, R. J. (1984). *Positionele letterfrequenties van het Nederlands* [Positional letter frequencies of Dutch]. Lisse: Swets & Zeitlinger.

Schaerlaekens, A., Kohnstamm, G. A., & Lejaegere, M. (1999). *Streeflijst woordenschat voor zesjarigen* [Target list vocabulary for 6-year-old children]. Lisse, The Netherlands: Swets & Zeitlinger.

Schneider, W., Eschman, A. & Zuccolotto, A. (2002). *E-Prime*. Pittsburgh, PA: Psychology software tools.

Schrooten, W., & Vermeer, A. (1994). *Woorden in het basisonderwijs: 15.000 woorden aangeboden aan leerlingen* [Words in primary school: 15.000 words presented to students]. Tilburg: University Press.

Schwanenflugel, P. J. (1991). Why are abstract concepts hard to understand? In Schwanenflugel, P. J., *The psychology of word meanings* (pp. 223-250). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schwanenflugel, P. J., & Akin, C. E. (1994). Developmental trends in lexical decisions for abstract and concrete words. *Reading Research Quarterly, 29*, 251-264.

Shibahara, N., Zorzi, M., Hill, M. P., Wydell, T., & Butterworth, B. (2003). Semantic effects in word naming: Evidence from English and Japanese Kanji. *The Quarterly Journal of Experimental Psychology, 56A*, 263-286.

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: an individual differences analysis. *Canadian Journal of Experimental Psychology, 53*, 347-359.

Strain, E, Patterson, K, & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of experimental psychology: Learning, Memory, and Cognition, 21,* 1140-1154.

Taft, M., & van Graan, F. (1998). Lack of phonological mediation in a semantic categorization task. *Journal of Memory and Language, 38,* 203-224.

Van Orden, G. C., Bosman, A. M. T., Goldinger, S. D., & Farrar, W. T. (1997). A recurrent network account of reading, spelling, and dyslexia. In J. W. Donahoe & V. Packard Dorsel (Eds.), *Neural network models of cognition: A biobehavioral foundation* (pp. 522-538). Amsterdam, the Netherlands: Elsevier Science Publishers.

Van Orden, G. C., Johnston, J. C. & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 371-386.

Van Loon-Vervoorn, W. A. (1985). *Voorstelbaarheidswaarden van Nederlandse Woorden: 4600 substantieven, 1000 verba en 500* adjectieven [Imageability ratings of Dutch words: 4600 substantives, 1000 verbs and 500 adjectives]. Lisse: Swets & Zeitlinger.

Zevin, J. D., & Balota, D. A. (2000). Priming and attentional control of lexical and sublexical pathways during naming. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 26*, 121-135.

Appendix A

Characteristics of the Words in the Lexical-Decision Task and Naming task.

Stimulus	Translation	Speed	Length	Imag	AoA	Freq(log)	Trigramfreg (mean)	Word type
bedrag	amount	0	6	4,50		1,40	97,50	1
bedtijd	bed time	0	7	3,70	97	,60	56,20	1
boel	lots	0	4	3,75	78	1,11	98,50	1
geest	sense	0	5	3,95		1,73	418,00	1
haat	hate	0	4	4,13		,78	1513,50	1
hitte	warmth	0	5	4,37	78	1,34	121,00	1
hoofdpijn	headache	0	9	4,03	96	1,49	55,00	1
moed	guts	0	4	3,90		1,57	528,50	1
onzin	nonsense	0	5	2,63	80	1,79	51,00	1
raadsel	riddle	0	7	2,90	84	1,15	42,60	1
scheet	fart	0	6	4,03	84	,60	286,00	1
stank	stink	0	5	3,73	84	1,26	399,67	1
tempo	speed	0	5	3,33		1,63	23,00	1
voorkant	front	0	8	3,87	93	1,57	207,33	1
warmte	warmth	0	6	3,00	85	1,64	62,50	1
dorst	thirst	1	5	4,40	93	1,72	141,67	1
haast	hurry	1	5	4,13	75	1,71	199,67	1
herrie	noise	1	6	4,27	85	1,70	35,50	1
kabaal	noise	1	6	3,97	75	1,41		1
keus	choice	1	4	3,83		1,26	66,50	1
kunstje	trick	1	7	3,83	87	1,65	69,80	1
lol	fun	1	3	4,00	89	1,59	5,00	1
mop	joke	1	3	4,33	92	1,34	2,00	1
pech	trouble	1	4	3,97	91	1,74	4,00	1
schap	shelf	1	5	3,20		,30		1
spreuk	slogan	1	6	3,60		,90	36,00	1
stukje	piece	1	6	4,03		,48	103,25	1
verdriet	sorrow	1	8	4,47	96	1,69	366,00	1
voornaam	first name	1	8	3,20	96	1,18	208,67	1
zijkant	flank	1	7	3,70	91	1,62	31,80	1
beetje	bit	0	6	3,67	97	3,28	193,25	2
beurt	turn	0	5	2,87	89	2,12	96,67	2
eind	end	0	4	3,73	84	2,44	302,00	2
geheim	secret	0	6	3,27	92	2,02	200,25	2
gek	mad	0	3	3,93	97	2,64	45,00	2
geluk	happiness	0	5	3,80	85	2,12	172,67	2
grap	joke	0	4	3,70	92	2,11	63,50	2
idee	idea	0	4	2,53	79	2,38	80,00	2
leven	life	0	5	3,50	90	2,70	876,00	2
ochtend	morning	0	7	4,10	90	2,21	108,00	2
plezier	pleasure	0	7	4,21	99	2,11	86,80	2
probleem	problem	0	8	3,17	76	2,23	135,50	2
trots	proud	0	5	3,80	84	2,29	296,67	2
vandaag	today	0	7	3,07	99	2,85	183,40	2
voorbeeld	example	0	9	2,10	82	2,42	264,14	2
afscheid	farewell	1	8	4,17	75	1,87	179,00	2
afspraak	appointment	1	8	3,56	78	1,92	70,50	2
begin	start	1	5	4,11	88	2,34	223,00	2
dag	day	1	3	4,03	97	3,27	359,00	2
geluid	sound	1	6	4,18	95	2,57	152,00	2
gevaar	danger	1	6	4,33	85	1,89	267,00	2
honger	hunger	1	6	4,33	98	2,27	291,50	2
iemand	somebody	1	6	3,77	87	2,99	224,25	2
manier	method	1	6	2,33	77	2,64	242,50	2

pauze	break	1	5	3,70	93	1,99	37,67	2
poos	while	1	4	3,70	82	2,08	31,50	2
recht	right	1	5	2,63	93	2,37	460,00	2
spul	stuff	1	4	2,73		1,86	9,50	2
verhaal	story	1	7	3,93	98	2,98	348,40	2
vraag	question	1	5	4,33	86	2,88	457,00	2
asbak	ash-tray	0	5	6,77	87	1,00	9,00	3
drank	drinks	0	5	6,17	80	1,04	113,33	3
dwerg	dwarf	0	5	6,30	82	1,74		3
gitaar	guitar	0	6	6,53	92	1,53	181,75	3
patat	chips	0	5	6,77	93	1,57		3
piano	piano	0	5	6,90	93	1,75	16,67	3
pleister	plaster	0	8	6,50	99	1,53	188,00	3
ridder	knight	0	6	6,10	78	1,78	649,75	3
rits	zipper	0	4	6,60	99	1,00	135,50	3
rolstoel	wheel chair	0	8	6,83	88	,85	19,00	3
sandaal	sandal	0	7	6,27	77	,85	192,20	3
speeltuin	playground	0	9	6,17	98	1,81	21,57	3
tang	pincers	0	4	6,40	78	,85	275,50	3
vla	pudding	0	3	6,37	87	,85		3
zandbak	sandbox	0	7	6,57	98	1,46	52,80	3
aardbei	strawberry	1	7	6,67	95	1,59		3
bips	bottom	1	4	6,53	77	1,04		3
gieter	watering can	1	6	6,60	91	1,32	503,75	3
ladder	stepladder	1	6	6,73	97	1,64	695,00	3
leesboek	reading book	1	8	6,27	95	1,70	40,67	3
parkiet	parakeet	1	7	6,61	77	1,15	27,80	3
perzik	peach	1	6	6,47	79	,90	26,75	3
pil	pill	1	3	6,63	91	1,36	8,00	3
pruik	wig	1	5	6,23	79	1,49	45,67	3
rozijn	raisin	1	6	6,37	79	1,15	13,75	3
sigaar	cigar	1	6	6,70	81	1,41	197,75	3
stier	bull	1	5	6,57	76	1,08	29,33	3
tegel	tile	1	5	6,27	95	1,68	696,33	3
wijn	wine	1	4	6,57	78	1,65	4016,00	3
zoen	kiss	1	4	6,23	98	1,80		3
afval	rubbish	0	5	6,37	84	1,89	94,33	4
foto	photo	0	4	6,70	89	2,62	30,00	4
groep	group	0	5	6,27	88	2,94	765,00	4
jager	hunter	0	5	6,27	78	2,26	202,00	4
lichaam	body	0	7	6,77	86	1,97	155,80	4
monster	monster	0	7	6,07	88	2,10	186,00	4
park	public garden	0	4	6,07	86	2,20	28,50	4
rots	rock	0	4	6,77	76	1,91	3,00	4
schouder	shoulder	0	8	6,37	95	2,46	220,83	4
straat	street	0	6	6,50	95	2,83	229,75	4
tijger	tiger	0	6	6,80	95	2,30	72,75	4
uil	owl	0	3	6,63	86	2,08	12,00	4
vijver	pool	0	6	6,07	93	2,16	57,50	4
vleugel	wing	0	7	6,30	93	2,09	34,80	4
walvis	whale	0	6	6,63	87	2,34	15,75	4
deken	blanket	1	5	6,17	91	2,06	510,33	4
getal	number	1	5	6,17	90	2,44	21,67	4
huid	skin	1	4	6,10	81	2,00	217,50	4
letter	letter	1	6	6,53	92	2,50	453,00	4
markt	market	1	5	6,35	88	2,29	107,33	4
mol	mole	1	3	6,23	87	2,29	7,00	4
prinses	princess	1	7	6,37	93	2,39	97,60	4
schaar	scissors	1	6	6,71	95	1,99	351,50	4
snavel	pecker	1	6	6,76	95	1,96	16,50	4

tekening	picture	1	8	6,47	97	2,76	275,00	4
toren	tower	1	5	6,76	90	2,29	656,33	4
vierkant	quadrangle	1	8	6,13	91	2,15	69,00	4
viool	violin	1	5	6,77	78	1,89	5,00	4
voorhoofd	forehead	1	9	6,43	89	1,90	160,43	4
wolf	wolf	1	4	6,23	88	2,48	16,00	4

Note. type 1 = LI-LF words, type 2 = LI-HF words, type 3 = HI-LF words, type 4 = HI-HF words.

Appendix B

Characteristics of the Nonwords in the Lexical-Decision Task and Naming task.

LI-LF-nonwords

Non-speeded	Speeded
andin	borkloet
beplag	daast
burtaad	donkt
duisgent	herkee
geenk	kabaar
hoonstijn	klonsje
hoste	koes
jol	mon
kaat	pach
moep	schak
piedsel	scheuk
soel	staag
spreet	sturda
tekka	vijklaam
wirste	zuikala

LI-HF nonwords

Non-speeded	Speeded
adee	abstried
burtje	begor
fand	binger
gezeum	braag
grank	decht
grop	geloes
lunen	gevoek
nak	marder
onklend	monglaak
pledink	oevand
stigreem	pag
teurt	pauma
trong	puis
vagtaag	stul
voonbeels	vesliek

HI-LF nonwords

Non-speeded	Speeded
aprak	aarkwee
drans	bers
golaar	kijn
jang	kodder
klorg	leemsoek
pates	mouter
piase	paggiet
renker	pig
ronsteum	pogel
sorgaal	proek
speektuim	rozeup
treisjer	sigoep
vits	stiep
vle	tarzik
zantrak	zeun

HI-HF nonwords

Non-speeded	Speeded
apral	dekof
eel	getes
groel	hied
jagon	lenker
kallis	madet
lisbaam	mel
loeger	nagening
monstig	pristes
neuver	schoer
pars	soren
pato	trakel
preugel	viaak
schieper	viefkald
strief	voofheels
vots	walf

Chapter 6: General discussion

The main purpose of the present thesis was to examine in which way semantic processes contribute to word identification of young children. Semantic contribution was investigated both in experimental settings and in a longitudinal-training study with poor beginning readers. In addition, because the focus was on readers who struggle in the initial process of learning to read, a prediction study was carried out, to investigate whether children who are at risk for reading disabilities could already be identified in kindergarten. In the next section, we will discuss the prediction of early reading difficulties, early reading intervention, and the role of semantic skills in early reading. Each topic starts with a brief summary of the major findings, followed by theoretical implications, practical implications, and limitations of the study. Finally, suggestions for future research will be presented.

Prediction of early reading difficulties

The study on predictive factors for subsequent reading skills has been described in Chapter 2. The predictive value of risk factors, cognitive factors, and teachers' judgments for subsequent reading skills was studied in a sample of 462 kindergartners. Results suggested that group membership (i.e., reading disabilities or not) in Dutch students at the beginning of Grade 1 could be predicted with moderate accuracy in kindergarten. Although letter knowledge was the strongest correlate of word decoding, it could not on its own correctly identify students who developed reading difficulties. A combination of productive letter knowledge, kindergarten teachers' predictions, and rapid naming of colors increased the accuracy of identification. The study demonstrated that although specific kindergarten variables (e.g., letter knowledge and rapid naming) show high correlations with reading abilities in Grade 1, reliable discrimination between poor readers and good readers on the basis of these variables was not achieved. Thus, although the results are fairly positive, accurate prediction at the individual level in kindergarten is still a challenge, both for researchers and for practitioners.

Theoretical implications

Results of the present study suggest that predicting reading difficulties in kindergarten is a rather complicated issue. Even the best predictor of reading will not discriminate sufficiently between poor readers and good readers (e.g., Scarborough, 1998). The vast amount of factors that is responsible for future reading skills certainly complicates fast and accurate prediction. These factors include reader characteristics (e.g., genetic factors, and cognitive factors), and environmental factors (e.g., classroom instruction, and home literacy). There are probably even more variables that affect reading performance. However, I choose to report on these factors because they cover a broad range of variables that are related to reading performance. The discussion is limited to a description of variables, because theoretical frameworks concerning the prevention and identification of reading difficulties are mostly restricted to verbal theories and correlational research.

First, reader characteristics have been demonstrated to contribute significantly to future reading skills. These characteristics include genetic-risk factors and cognitive factors. Although the present study failed to demonstrate a significant genetic predictor for reading difficulties, other studies have shown the predictive value of dyslexia or speech and language problems in first-grade relatives (e.g., Lewis, Freebairn, & Taylor, 2000; Snowling, Gallagher, & Frith, 2003). Cognitive skills of the student (e.g., letter knowledge, phonological skills) also predict future reading skills. An overview of cognitive processes entailed in learning to read is provided by Vellutino, Fletcher, Snowling, and Scanlon (2004). They outline a variety of processes and types of knowledge that are related to reading. Their descriptive model illustrates the broad range of skills that are included in decoding and which might be responsible for reading difficulties. Letter knowledge and rapid-naming skills, both significantly correlated with reading in the present study, constitute only a small part of the model.

Second, environmental factors have been found to affect children's reading skills. Inadequate instruction or limited home literacy experience might result in reading difficulties that are not caused by cognitive deficits (e.g., Vellutino et al., 2004). Neither the predictive value of literacy activities at home nor the impact of instruction has been examined in the present study. However, it is important to realize that these factors may cause significant variations in reading outcome, and therefore, make accurate predictions harder.

In sum, reading is a complex skill, and children's reading level is affected by a number of different skills. This certainly complicates accurate identification of those children who will develop reading difficulties and those who will not. Importantly, despite this high number of variables that are correlated with reading performance, teachers could relatively accurately predict which children would develop reading difficulties: Teachers' predictions yielded the highest number of valid positives, that is, the number of students that were predicted to have reading difficulties and who turned out to have reading difficulties.

Practical implications

Although a number of researchers have studied the prediction of reading difficulties, the best screening measure has not been found yet. Evidently, it should be inexpensive and quickly to administer. Moreover, it should accurately predict those children who will develop reading difficulties and accurately identify those children who will develop adequate reading skills. Unfortunately, the number of misidentifications is still unsatisfying and until now, no screening procedure can perfectly estimate future reading performance. Although results from this thesis showed that at-risk status for reading disabilities can be predicted relatively accurately in kindergarten, there is no risk factor that can accurately predict reading difficulties on its own. Therefore, it is recommended to include a combination of several predictive factors. Tests of letter knowledge and phonological skills (e.g., rapid naming) might provide a good indication of children's initial reading skills. Second, in the literature, it is suggested that graduated scoring and corrective feedback in phonological tasks might increase the predictive validity of tests (O'Connor & Jenkins, 1999). In graduated scoring (in contrast to an all-or none score), items are not scored as right or wrong, rather are scored in a gradual way. For example, in the nonword-repetition task, items could have been scored on multiple criteria like the correct number of syllables repeated or correct initial phonemes repeated. Providing corrective feedback enables the experimenter to examine the learning capacities of the child. Third, kindergarten teachers' perceptions and predictions contribute to correct identification of readers who are at risk for reading disabilities.

Because there is no screening procedure that perfectly predicts reading difficulties, I recommend to monitor kindergartners performance on reading-related skills carefully and to gather information about a child's language profile. Cooperation with a speech- and language therapist would be useful on this subject. Kindergarten teachers should be instructed profoundly on beginning literacy during their education, so that they are able to notice early predictors adequately.

In the Netherlands, the implementation of the Protocol Leesproblemen en Dyslexie [Protocol Reading problems and Dyslexia], formulated by Wentink and Verhoeven (2001) contributes in early prediction and prevention of reading difficulties. Children at risk for reading disabilities should be stimulated in kindergarten and should be monitored on reading performance early in Grade 1. After all, word-identification skills can be accurately assessed after only two months of formal reading instruction. Screening at the beginning of first grade makes it possible to provide early intervention.

Limitations

In the present study, only the relationship between kindergarten variables and beginning reading skills (that is, reading CVC-words) was measured. Although several researchers have demonstrated relatively stable levels of reading skills through primary grades (Bast & Reitsma, 1998; Verhoeven & van Leeuwe, 2003), it would be interesting to know the relationship with reading skills at the end of first grade or second grade. The strength of the kindergarten predictors might have been different if predictor variables were related to reading pseudowords or connected text after one or two years of reading experience. Moreover, we only included a rapid-naming test as a phonological measure. The inclusion of a phonological-awareness test might have caused additional explained variance, although the additional value would probably have been small. After all, letter knowledge is the best (phonological) predictor.

Early reading intervention

The crucial question in Chapter 3 was whether a semantically-oriented training was effective for improving word-decoding skills of poor readers. Results of the first experiment revealed that after four months of training, children from the semantic group surpassed those from the phonological group in letter knowledge and reading isolated words. However, the superiority of the semantic training disappeared at the end of the training, when reading performance of the semantic group and the phonological group was statistically equal. Importantly, both experimental groups performed at least as good as the control group, in which the majority of children received additional help from a remedial teacher. This finding demonstrates the ecological validity of the training programs. A second experiment provided additional support for the ecological validity. Gains in word-identification skills were equal for children in the control group and children in the experimental groups. Moreover, there was no difference in performance between the semantic group and phonological group. Thus, in the end, the type of instructions in a training program seems of minor importance: A semantically-oriented training resulted in similar gains in reading performance as a phonologically-oriented training. This result challenges the generally accepted view that a reading-intervention program should focus primarily on phonological skills.

Theoretical implications

Theories of word identification do not provide direct evidence for instructions or remediation: A hypothesized cognitive process in reading does not imply that training this process will lead to successful outcomes in reading performance (van den Broeck, 1993). However, a theory of reading might provide preliminary indications for the construction of an intervention program and results of intervention studies could be used to change or adapt theoretical models of visual word perception. In the present study, the Phonological Coherence Model (Van Orden et al., 1990) was chosen as a source of inspiration for the construction of a reading-intervention program. This model has been described in Chapter 1. It was hypothesized that the activation of semantics in the word-identification process might support the correspondence between grapheme nodes and phoneme nodes. This should be helpful when the translations from graphemes to phonemes are slow or error-prone, as in poor readers. However, the findings in this thesis showed that explicit focusing on semantic characteristics of words in an intervention program does not lead to better word-decoding performance than focusing on phonological characteristics in a training program. Because phonological, orthographic, and semantic information are activated simultaneously, and because activation flows bi-directionally, it is impossible to disentangle the unique contribution of semantics. Thus, the present results cannot be used to evaluate the key assumptions of the Phonological Coherence Model. It should be stressed that the model only served as a source of inspiration for the design of a reading intervention. The results do however demonstrate that repeated focus on semantic characteristics of words *during intervention* does not result in better word-decoding skills in an isolated word-reading task or in reading text than repeated focus on phonological characteristics does. This is an important finding, which contributes to our knowledge about reading in general and reading intervention in particular.

However, many questions remain unanswered. For example, it is unclear whether or not the semantic-phonological connections of those children who received a semantic training were improved. If explicit focus on semantic characteristics improved the connections between phonological and semantic nodes, one might expect that these strengthened connections would support word decoding.

To conclude, the results of the present study suggest that focusing on semantic characteristics in a training program leads to similar gains in reading than focusing on phonological characteristics of words. A comprehensive theoretical model of reading intervention should be helpful to construct and evaluate effective reading intervention in the future.

Practical implications

The current findings on intervention, together with clinical observations and suggestions from previous research, provide several implications for practitioners. First, the present research has demonstrated that intervention can be implemented early in Grade 1. It is advisable to start reading intervention as early as possible. In the Dutch curriculum, children have been taught a considerable amount of letters already in the first two months of Grade 1. These graphemes and several combinations of graphemes can be implemented immediately in a reading-intervention program.

Second, the intervention should be intensive (preferable on a daily basis) and last for a sufficient amount of time. A meta-analysis carried out by Therrien (2004) showed that it is advisable to use a performance criterion (e.g. a fixed number of words read correctly or in a limited amount of time), rather than a fixed number of reading sessions. As a consequence, poor readers need a lot of rehearsals to reach a required performance level.

Third, the exact type of instructions in an intervention program that is supplemental to (phonics) instruction in the classroom seems of minor importance. The present thesis showed that a phonologically-oriented program and a semantically-oriented program resulted in similar gains in word-decoding skills. I recommend the inclusion of exercises that focus on both semantics and phonology of printed words. This will increase the relationship between orthography, semantics, and phonology. In addition, it is recommended to include orthography from the start. Several studies (e.g., Bus & van IJzendoorn, 1999; Ehri et al., 2001) have demonstrated that exercises that appeal to children's phonological skills without the inclusion of graphemes are not as effective as instructions that include graphemes.

Fourth, children, particularly those who have poor reading skills, need sufficient encouragement to practice their reading skills. A motivating attitude of the teacher may contribute here. I experienced that children enjoyed the reading exercises to a greater extent if teachers disseminated a positive attitude towards the intervention. Another factor that contributes to strong engagement is the inclusion of visual and auditory feedback in a computer program. In our training program, different variants of positive auditory feedback were included (e.g., "Well done"," Excellent", and applause) and children were noticeable encouraged by these types of feedback. Furthermore, it is recommendable to provide feedback on the speed of their responses to encourage the automatization of word identification.

Limitations

In the present design and implementation of the intervention study, a certain trade-off between theoretical objectives and practical attainability was experienced. Although I tried to minimize the limitations that Lyon and Moats (1997) reported as much as possible, they could not be completely avoided. First, the intensity of instruction related to the intensive classroom instruction and practice was small. Second, it will be valuable to evaluate the effect of the experimental-training programs in higher grades, when classroom instruction and practice on word decoding has been decreased. Third, a proper distinction between the phonological and the semantic program could not be attained at each difficulty level. For example, at the grapheme level, it was impossible to implement pure semantic activities. In addition, reading strategies of children sometimes led to other strategies than the ones that were intended in the instruction. For example, in the word-hunt exercise, the semantic training provided a comprehensive context whereas in the phonological training, sentences were incoherent. However, instead of reading the consistent (semantic) or inconsistent (phonological) set of sentences as a whole, some children just scanned the text for the target words without focusing on the context. As a consequence, the eventual profit of a meaningful context disappeared.

The role of semantic skills in early reading

To respond accurately in a semantic-oriented training program as designed in the present thesis, a minimum level of semantic skills is required. To examine whether poor readers possess adequate semantic skills, the relationship between semantic skills and word-decoding skills was studied across primary grades. The outcome of this study has been described in Chapter 4. In the first experiment, children from Grade 1 performed a word-association task and a semantic-categorization task and task performance was related to reading performance. In Experiment 2, we included readers from Grades 1 to 6 and assessed a different type of semantic-categorization tasks. Although poor readers and good readers demonstrated the same level of word-association skills, results of the semantic-categorization tasks were somewhat conflicting. Experiment 1 revealed that poor readers made more errors than good readers did, whereas in Experiment 2, no differences in accuracy were found for poor readers, average reader, and good readers. However, in Experiment 2 ,we did find a difference in speed of semantic categorization that was not caused by differences in decoding speed in the test. Poor readers showed longer reaction times in semantic categorization than average readers or poor readers did. Outcome differences across both experiments may be attributed to differences in age and task instruction in the experiments.

Thus, the tentative conclusion in Chapter 4 was that poor readers have slower semantic access in a semantic-categorization task than average or good readers. However, semantic skills as measured in offline-processing tasks (i.e., a semantic-categorization task or a word-association task) do not reflect indirect semantic-processing skills (e.g., semantic priming), which may be more linked to word decoding. To examine whether young children benefit from semantic facilitation in isolated-word reading, we performed a study on imageability effects, which has been described in Chapter 5. Two experiments were carried out to examine the effect of imageability in lexical decision and in naming. Results demonstrated that the imageability advantage was only present in lexical decision. Reaction times were shorter for high-imageability words than for low-imageability words. In addition, in the higher grades, and only in the speeded test, more errors were made in low-imageability words than in high-imageability words. Results of Experiment 2 indicated that in naming, no imageability advantage was demonstrated for accuracy scores. Thus, semantic attributes affected word decoding only in a restricted way.

Theoretical implications

The major question is: How can the present results be interpreted within a theoretical model of reading? As outlined before, the Phonological Coherence Model, as formulated by Van Orden and colleagues (Farrar & Van Orden, 2001; Van Orden, Pennington, & Stone, 1990; Van Orden, Bosman, Goldinger, & Farrar, 1997) was chosen as a theoretical framework for the present thesis. It is important to note that this reading model refers to the identification of words in adult, experienced readers. The results of the present study offer some tentative conclusions that can be used to evaluate the model for younger, less-experienced readers.

First, because of the assumed orthographic, phonological, and semantic activation during reading, it was hypothesized that children’s skills in all three domains might contribute to correct word decoding. Results on semantic-categorization skills across all grades (Chapter 4, Experiment 2) suggested that poor readers have slower semantic access than average or good readers and it was concluded that semantic skills and reading skills are related to each other. Previous research has already pointed out that poor readers are challenged in both phonological skills (e.g., Elbro, Borstrøm, & Petersen, 1998; O’Connor & Jenkins, 1999) and semantic skills (e.g., Gillon & Dodd, 1994). These results can be explained in a recurrent network of reading in which orthographic, phonological, as well as semantic skills can contribute to accurate and fast word decoding.

Second, it was investigated whether semantic contribution affected reading performance implicitly. In naming accuracy, we failed to find an imageability effect in decoding isolated words. High-imageability words were read as accurately as low-imageability words, suggesting that semantic attributes of isolated words do not facilitate decoding those words for Dutch children. This does not imply that semantics are not activated during word decoding in young readers. Rather, the activated semantic information probably does not contribute to more efficient decoding. The absence of semantic contribution might be caused by the type of semantic manipulation (imageability) , the relatively transparent orthography of Dutch, or the type of reading task. In oral reading, an overt, oral response is required, and the response can be provided without any knowledge of word meanings. Thus, semantic information might be activated in the task (a result that is hypothesized in connectionist models of reading), but its contribution is of minor importance. In contrast, in a lexical-decision task, the participant has to determine whether the letter string constitutes a word or not. Indeed, results of our study demonstrated an imageability effect in a lexical-decision task. Results on latencies and future research on other semantic manipulations are needed to test this assumption. In summary, implicit semantic knowledge (e.g., imageability) only marginally affects the reading of words in young children.

Practical implications

The study on semantic factors in early reading does not provide direct clinical implications. Both the experiments on semantic skills and imageability were designed to evaluate theoretical models of visual-word perception. Results do not answer the issue of causality and thus do not provide suggestions for the educational practice.

Limitations

In the study on semantic skills (Chapter 4), the mean age of the poor readers and good readers in Experiment 2 was different. In future research, it is recommended to include a poor readers group, age-level matched group and reading-level matched group to prevent such shortcomings. Furthermore, although we only selected those concepts that were highly likely to be known by six-years-old Dutch children (Schaerlaekens, Kohnstamm, Lejaegere, & de Vries, 1999; familiarity rating $\geq .80$ on a scale from 0 to 1), we did not test whether all pictures could be correctly identified by all children. It would have been proper to perform a pilot-study in which this issue was addressed albeit there appeared to be no problems in this respect. In addition, we did not have adult ratings of semantic, phonological, and perceptual similarity in the Exemplar-stimuli tests. Finally, in the study on imageability effects (Chapter 5), it would have been better to have a speed measure in the naming task, because accuracy scores were relatively high and thus decreased the variability.

Future issues

To conclude, the present study was an attempt to increase our knowledge of semantic processes in word-identification of young children. Future research should extend our understanding of semantic processes and semantic skills in young readers, to investigate their role in reading and reading instruction. Future intervention studies should be longitudinal, to investigate the effects over a significant period of time and studies should address both short-term and long-term gains in reading. Moreover, intervention studies including large samples of participants should be supplemented with single case studies. This will probably lead to better insight in developmental changes. Relatively large heterogeneity of groups and great differences in response to intervention suggest that the use of large-sample statistics is not sufficient for answering the complicated questions regarding reading skills and reading intervention. Another solution for the heterogeneity of groups, which will be a useful method for the future, has been found in the study of intra-individual development. The intra-individual development can be studied by means of individual growth curves once participants are being assessed at several points in time. The latter can be realized by studying the reading progress that is made in the intervention: Results of each training session can be used to study intra-individual changes.

With respect to semantic facilitation in word decoding, I choose to investigate the effect of imageability. Further research should investigate the effects of other semantic manipulations (e.g., ambiguity, synonymy) to obtain a more profound insight in semantic effects in isolated word reading in children. In this thesis, I already emphasized that small manipulations or decisions in the design of a study have a great impact on the results and conclusions. Therefore, in the future, it is important to study semantic effects in a number of different conditions, which may result in general conclusions. Cross-linguistic studies can shed a light on the influence of a language's orthography on semantic effects in word decoding in young children.

Finally, researchers and practitioners should sufficiently communicate and provide input to each other. An open-minded view of researchers and collaboration with other experts in the field of reading is essential for future success.

REFERENCES

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology, 34,* 1373-1399.

Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91,* 403-414.

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

Elbro, C., Borstrøm, I., & Petersen, D. K. (1998). Predicting dyslexia from kindergarten. The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly,* 33, 36-60.

Farrar, W. T., & Van Orden, G. C. (2001). Errors as multistable response options. *Nonlinear Dynamics, Psychology, and Life Sciences, 5*, 223-265.

Gillon, G., & Dodd, B. J. (1994). A prospective study of the relationship between phonological, semantic and syntactic skills and specific reading disability. *Reading and Writing, 6,* 321-345.

Lewis, B. A., Freebairn, L. A., & Taylor, H. G. (2000). Academic outcomes in students with histories of speech sound disorders. *Journal of Communication Disorders,* 11-30.

Lyon, G. R., & Moats, L. C. (1997). Critical conceptual and methodological considerations in reading intervention. research. *Journal ofLearning Disabilities, 30, 6,* 578-588.

O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading, 3,* 159-197.

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

Schaerlaekens, A., Kohnstamm, G. A., & Lejaegere, M. (1999). *Streeflijst woordenschat voor zesjarigen [Target list vocabulary for 6-year-old children].* Lisse, The Netherlands: Swets & Zeitlinger.

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

Therrien, W. J. (2004). Fluency and Comprehension Gains as a Result of Repeated Reading: A Meta-Analysis. *Remedial and Special Education, 25,* 252-261.

Van den Broeck, W.(1993). Theorieën van woordherkenning en praktische implicaties [Theories of word recognition and practical implications]. *Tijdschrift voor Orthopedagogiek, 11*, 474-492.

Van Orden, G.C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97,4*, 488-522.

Van Orden, G. C., Bosman, A. M. T., Goldinger, S. D., & Farrar, W. T. (1997). A recurrent network account of reading, spelling, and dyslexia. In J. W. Donahoe & V. Packard Dorsel (Eds.), *Neural network models of cognition: A biobehavioral foundation* (pp. 522-538). Amsterdam, the Netherlands: Elsevier Science Publishers.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M.(2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45, 1,* 2-40.

Verhoeven, L., & van Leeuwe, J. (2003). Ontwikkeling van decodeervaardigheid in het basisonderwijs [Development of decoding skills in primary school]. *Pedagogische Studiën, 80,* 257-271.

Wentink, H., & Verhoeven, L. (2001). *Protocol voor leesproblemen en Dyslexie.* Nijmegen: Expertisecentrum Nederlands.

Summary

The aim of the present dissertation was to obtain insight in semantic effects in word identification of Dutch children. In a larger framework, the goal of the study was to increase our knowledge about the prediction and prevention of severe reading disabilities. Chapter 2 described the results of a longitudinal prediction study. This study focused on the predictive value of risk factors, cognitive factors, and teachers' judgments in a sample of 462 kindergartners for their early reading skills and reading failure at the beginning of Grade 1. With respect to risk factors, enrollment in speech-language therapy, history of dyslexia or speech-language problems in the family, and the role of gender were considered. None of these risk factors were significantly related to reading performance in Grade 1. Cognitive factors in this study included letter knowledge, rapid-naming ability, and nonword-repetition skills. Of these skills, letter knowledge showed the highest correlation with reading. Kindergarten teachers' judgments, including a task-assignment scale and teachers' predictions, demonstrated a significant relationship with reading. Finally, to judge whether these predictors could identify reading disabilities, the discriminatory power of all predictors was assessed and appeared to be insufficient. Results of a discriminant function analysis, however, demonstrated that a combination of productive letter knowledge, rapid naming of colors, and teachers' predictions increased the accuracy of prediction to an overall accuracy rate of 71%. In conclusion, the results of this study suggested that group membership (reading disabilities or not) in Dutch students at the beginning of Grade 1 can be moderately predicted in kindergarten.

If children are diagnosed to develop reading difficulties, their early reading development has to be monitored carefully and intervention should be started as soon as possible. Chapter 3 describes the results of two intervention studies in Grade 1, in which effects of a semantically-oriented training program and a phonologically-oriented training program are discussed. In the first study, 121 poor beginning readers (mean age 6.5 years) from 22 regular primary schools were assigned to a semantically-oriented training, a phonologically-oriented training or a control group. Results showed an advantage for the semantic training over the phonological training after four months of training. At post-test, however, children in both experimental training programs showed similar gains in word-identification skills. In the second study, the experimental-training programs were modified and extended. About 83 poor beginning readers participated in one of both experimental

groups or were assigned to a control group. Letter knowledge, word-identification skills, text-reading skill and receptive vocabulary were assessed during the training and reading skills were assessed at follow-up, mid-Grade 2. Result showed that all groups performed statistically equally across all measures. Thus, a word-identification training program for poor beginning readers focusing on the semantics of words is as effective as a training program focusing on the phonology of words. Provided with this result, additional information about semantic effects in word reading in young children is useful. To investigate the relationship between semantic skills and reading performance, two experiments were conducted. Results of these experiments have been described in Chapter 4.

In Chapter 4, the relationship between semantic skills and word-decoding skills was examined. In Experiment 1, 99 first graders participated in a semantic-categorization task, a word-association task and a word-decoding test. Results revealed no differences between poor decoders and good decoders in word-association skills, whereas poor readers were more error prone in a semantic-categorization test. In Experiment 2, children from Grades 1 to 6 participated in two types of semantic-categorization tasks and a word-decoding test. The categorization tasks were performed in different modalities: Concepts were presented by means of printed words, spoken words, or pictures. Response options were all pictures. It turned out that poor readers showed longer reaction times on both types of categorization tasks than average readers and good readers. This difference did neither vary across grades nor across different modalities of the stimuli. The results suggested that semantic skills are related to reading performance if semantic skills are measured with offline procedures, that is, tasks that require conscious processing like semantic categorization. To investigate semantic skills in beginning readers in online procedures, which address automatic components of semantic organization, two additional experiments have been carried out and have been described in Chapter 5.

Chapter 5 focused on imageability effects in isolated-word reading in Dutch children from Grades 2 to 6. Word-reading skills were assessed by lexical decision (Experiment 1) and naming (Experiment 2). In both experiments, a speeded task as well as a non-speeded task were administered. Results of Experiment 1 revealed an imageability advantage in the non-speeded task in latencies and an imageability advantage in low-frequency words in accuracy; in low-frequency words, more errors were made in low-imageability (LI-) words than in high-imageability (HI-) words. Moreover, accuracy analyses revealed a significant grade by instruction by imageability interaction; in the higher grades, only in the speeded test more errors were made in LI-words than in HI-words. Results of Experiment 2 indicated that in naming, no imageability advantage was demonstrated for accuracy scores. Unexpectedly, children in the lower grades made more errors in HI-words than in LI-words, but only in high-frequency words and only in the speeded task. In short, imageability can affect word perception in Dutch children, but the effects are small and depend on word characteristics (HF-words versus LF-words), type of reading task (lexical decision versus naming), instruction and outcome measures (in a non-speeded lexical-decision task, semantic effects were only observed in reaction times, whereas in a speeded task, semantic effects were only apparent in accuracy scores), and participant characteristics (more prevalent in the higher grades than lower grades).

In general, the conclusion of the thesis is that a training program in Grade 1 focusing on semantic characteristics of words leads to similar gains in reading than focusing on phonological characteristics of words. Second, although performance in a semantic-categorization task was related to reading performance, semantic attributes (imageability ratings) affected word decoding in beginning readers only in a restricted way. Thus, the present research has revealed a relationship between semantics and word reading in beginning readers. However, the impact of semantics on visual word identification in the early grades appears to be modest.

Samenvatting

In deze dissertatie wordt verslag gedaan van een onderzoek naar semantische effecten in het aanvankelijk leesproces. De studie beoogt een bijdrage te leveren aan de huidige kennis over het voorspellen en voorkomen van ernstige leesproblemen.

Hoofdstuk 2 beschrijft de resultaten van een longitudinale predictie studie. In deze studie is de voorspellende waarde van risicofactoren, cognitieve factoren en leerkrachtoordelen onderzocht bij 462 kinderen in groep 2 voor hun leesprestaties in het eerste semester van groep 3. Risicofactoren die zijn onderzocht betreffen het krijgen (of in het verleden gekregen hebben) van logopedie, het voorkomen van dyslexie en/of spraak- en taalproblemen in de familie en het geslacht van de leerling. Geen van deze risicofactoren bleek significant gerelateerd te zijn aan de leesvaardigheid in groep 3. Cognitieve factoren die zijn onderzocht, zijn letterkennis, snelheid van benoemen van kleuren en plaatjes en nonwoord repetitie. Letterkennis in groep 2 bleek het hoogst te correleren met de toekomstige leesprestaties in groep 3. Leerkrachtoordelen betroffen de resultaten van een werkhoudingslijst die is ingevuld door leerkrachten van groep 2 en de verwachtingen van de leerkrachten omtrent de ontwikkeling van leesproblemen bij de kinderen. Deze leerkrachtoordelen bleken significant samen te hangen met de leesvaardigheid in groep 3. De afzonderlijke voorspellers (risicofactoren, cognitieve factoren en leerkrachtoordelen) waren echter onvoldoende in staat om het voorkomen van leesproblemen in groep 3 correct te voorspellen. Wanneer een combinatie van actieve letterkennis, het snel benoemen van kleuren en leerkrachtoordelen werd meegenomen, was de voorspelling accurater en kon 71% van de kinderen goed geclassificeerd worden, dat wil zeggen, een correcte voorspelling wel/geen leesproblemen.

Wanneer leesproblemen verwacht worden, is het van groot belang om het beginnend leesproces nauwlettend in de gaten te houden en zo vroeg mogelijk een adequate leesinterventie op te zetten. Hoofdstuk 3 beschrijft de resultaten van twee interventiestudies waarbij de effecten van een semantisch georiënteerd trainingsprogramma werden onderzocht. In het eerste experiment werden 121 zwakke lezers (gemiddeld 6.5 jaar) van 22 reguliere basisscholen toegewezen aan een semantisch georiënteerde training, fonologisch georiënteerde training, of een controlegroep. Na 4 maanden presteerden de leerlingen van de semantisch georiënteerde training significant beter dan de leerlingen van de fonologisch georiënteerde training. Aan het einde van de training, bij de posttest, waren de effecten van beide trainingsprogramma's echter even groot. In het tweede experiment werden de trainingsprogramma's aangepast en uitgebreid. Ruim 80 zwakke lezers in groep 3 werden opnieuw toegewezen aan de semantisch georiënteerde training, fonologisch georiënteerde training, of een controlegroep. De kinderen werden getest op letterkennis, leesvaardigheid van woorden en korte teksten en woordenschat. De resultaten lieten zien dat de effecten van training van alle groepen (zowel de experimentele groepen als controlegroep) statistisch gelijk waren. Dus, een trainingsprogramma voor technisch lezen voor zwakke lezers in groep 3 dat de nadruk legt op de semantiek van woorden is even effectief als een training die zich met name richt op de fonologie van woorden. Aanvullende informatie over semantische effecten in het technisch lezen bij jonge leerlingen is wenselijk. Om de relatie tussen semantische vaardigheden en leesvaardigheid te onderzoeken, zijn twee experimenten uitgevoerd en beschreven in hoofdstuk 4.

In het eerste experiment van hoofdstuk 4 hebben 99 leerlingen uit groep 3 een semantische categorisatietaak en een woord-associatietaak uitgevoerd. De prestaties op deze taken zijn vergeleken met de prestaties op een leestest voor het lezen van losse woorden. Er bleek geen verschil te zijn in associatievaardigheid tussen zwakke lezers en goede lezers. De zwakke lezers maakten echter significant meer fouten dan goede lezers in de semantische categorisatietaak. In het tweede experiment hebben kinderen van groep 3 tot en met 8 twee typen semantische categorisatietaken uitgevoerd en opnieuw werden de prestaties vergeleken met de prestaties op een leestest voor het lezen van losse woorden. In de categorisatietaken werden de stimuli aangeboden als plaatje, gesproken woord of geschreven woord. De antwoordopties waren telkens plaatjes. Resultaten lieten zien dat zwakke lezers langere reactietijden hadden op de semantische categorisatietaken dan gemiddelde lezers en goede lezers. Dit verschil had geen relatie met de groep waarin de kinderen zaten of de modaliteit waarin de stimuli werden aangeboden. De resultaten suggereren dat semantische vaardigheden gerelateerd zijn aan leesvaardigheid indien taken worden gebruikt die bewuste verwerking vragen, zoals de semantische categorisatietaak. Om semantische effecten te onderzoeken in taken die automatische componenten van semantiek bevatten, zijn twee aanvullende experimenten uitgevoerd en beschreven in hoofdstuk 5.

In hoofdstuk 5 wordt verslag gedaan van een onderzoek naar het effect van voorstelbaarheid op het lezen van losse woorden bij kinderen van groep 4 tot en met 8. De leesvaardigheid van losse woorden werd zowel met een lexicale-decisie taak (experiment 1) als met een hardop leestaak (experiment 2) gemeten. Beide taken werden met tijdsdruk en zonder tijdsdruk afgenomen. Resultaten van de lexicale-decisietaak lieten kortere reactietijden zien voor hoogverstelbare woorden dan voor laagvoorstelbare woorden in de taak zonder tijdsdruk. Wat betreft accuratesse was er een effect van voorstelbaarheid bij de laagfrequente woorden: bij laagfrequente woorden werden meer fouten gemaakt in laagvoorstelbare woorden dan in hoogvoorstelbare woorden. De accuratesse scores lieten verder een significante interactie zien van groep x instructie x voorstelbaarheid; in de hogere groepen werd alleen in de taak met tijdsdruk meer fouten gemaakt in laagvoorstelbare woorden dan in hoogvoorstelbare woorden. Resultaten van de hardop leestaak konden in de accuratesse scores geen voordeel aantonen voor hoogvoorstelbare woorden. Kinderen in de lagere groepen maakten zelfs meer fouten in hoogvoorstelbare woorden dan in laagvoorstelbare woorden, maar dit bleek alleen het geval in hoogfrequente woorden in de taak met tijdsdruk. Dus, de mate van voorstelbaarheid van woorden kan het technisch lezen van kinderen wel beïnvloeden, maar de effecten zijn klein en zijn afhankelijk van woordkarakteristieken (hoogfrequent versus laagfrequent), type taak (hardop lezen versus lexicale decisie), instructie (met of zonder tijdsdruk), maat van prestatie (snelheid of accuratesse) en groep (onderbouw versus midden- en bovenbouw).

De algemene conclusie van dit proefschrift is dat een leestraining die gericht is op de semantiek van woorden in groep 3 even effectief is als een training die gericht is op de fonologie van woorden. Hoewel prestaties op een semantische categorisatietaak gerelateerd waren aan de leesvaardigheid van beginnende en gevorderde lezers, bleek verder dat semantische eigenschappen van woorden (de mate van voorstelbaarheid) de woordidentificatie bij beginnende lezers slechts beperkt beïnvloedt. Het onderzoek heeft dus een relatie aangetoond tussen semantiek en woordidentificatie bij beginnende lezers. De invloed van semantische effecten en semantische vaardigheden lijkt echter zeer bescheiden bij beginnende lezers.

Curriculum Vitae

Martine Gijsel is geboren op 11 maart 1976 te Terneuzen. Na het behalen van haar VWO-diploma op scholengemeenschap Petrus Hondius te Terneuzen in 1994, startte zij in datzelfde jaar de opleiding Logopedie en Audiologie aan de Hogeschool Vesalius te Gent. Na het succesvol afronden van deze studie in 1997, startte zij de studie Spraak- en Taalpathologie aan de Katholieke Universiteit Nijmegen (KUN). Tijdens de eindfase van haar studie is ze tevens werkzaam geweest als logopediste. De studie heeft ze voltooid in 2000, waarna ze korte tijd als onderzoeksassistent aan de KUN, sectie Orthopedagogiek Leren en Ontwikkeling heeft gewerkt. In 2001 startte ze bij deze vakgroep haar promotieonderzoek. Naast het onderzoek, dat in dit proefschrift is beschreven, is ze werkzaam geweest als docent logopedie aan de Hogeschool van Arnhem en Nijmegen. Sinds mei 2006 is ze werkzaam bij het Expertisecentrum Nederlands.