J. Child Lang. **38** (2011), 356–379.
doi:10.1017/S0305000909990420

Semantic categorization and reading skill across Dutch primary grades: development yes, relationship no*

MARTINE A. R. GIJSEL

Radboud University Nijmegen, Behavioral Science Institute

ELLEN A. ORMEL

Radboud University Nijmegen, Behavioral Science Institute and Pontem, St. Michielsgestel

DAAN HERMANS

Pontem, St. Michielsgestel

L. VERHOEVEN

Radboud University Nijmegen, Behavioral Science Institute

AND

ANNA M. T. BOSMAN

Radboud University Nijmegen, Behavioral Science Institute

(*Received* 4 *January* 2008 – *Revised* 2 *January* 2009 – *Accepted* 9 *November* 2009 – *First published online* 28 *May* 2010)

ABSTRACT

In the present study, the development of semantic categorization and its relationship with reading was investigated across Dutch primary grade students. Three Exemplar-level tasks (Experiment 1) and two Superordinate-level tasks (Experiment 2) with different types of distracters (phonological, semantic and perceptual) were administered to assess semantic categorization skills. Reading was measured with a standardized word-reading test. Results of both experiments demonstrated that children in the higher grades had shorter reaction

[*] We thank Corina Michielsen, Annelies Leechburch-Auwers and Jikke Planting who collected most of the data and assisted in creating all test materials. We are also grateful to all the children who participated in the study and to the teachers who gave their permission to run the experiment. Address for correspondence: Martine A. R. Gijsel, Expertisecentrum Nederlands, Postbus 6610, 6503 GC Nijmegen, The Netherlands. tel: +31-24-36 15624; fax +31-24-36 15644; e-mail: m.gijsel@taalonderwijs.nl

times and fewer errors than children in the lower grades. Reading skill, however, was not related to semantic categorization performance. Moreover, neither grade level nor reading skill was related to the effect of distracter type on error percentages. Based on the results of this study, we suggest a substantial development of semantic categorization skills over time, and reject the notion that Dutch poor readers have less advanced semantic categorization skills than typical readers.

INTRODUCTION

Decades of reading research led to an almost universal agreement about three fundamental aspects that govern reading: orthography (or spelling), phonology (or sound) and semantics. All three appear to contribute interactively to the reading process. Empirical evidence for the role of orthography and phonology in experienced and fluent reading is abundant (e.g. Barker, Torgesen & Wagner, 1992; Berent & Perfetti, 1995; Bosman & van Hell, 2002; Juel, Griffith & Gough, 1986; Stanovich & West, 1989; Van Orden, Pennington & Stone, 1990; Seidenberg & McClelland, 1989).

Although the role of semantics in reading (e.g. Azuma & Van Orden, 1997; Meyer & Schvaneveldt, 1971; Strain, Patterson & Seidenberg, 1995) and learning to read (e.g. Assink, Van Bergen, Van Teeseling & Knuijt, 2004; Scarborough, 1991) has been recognized, it has not received as much attention as the orthographic and phonological aspects have. The main goal of this study, therefore, is to add to the understanding of the role of semantics to reading in students attending primary grades. We will address the issue of semantic development and the relationship between semantic skills and word reading.

Semantic(s) is a rather abstract notion, revealed by the many different ways in which it has been operationalized in the domain of experimental psycholinguistics. Examples are recognizing and defining words (receptive and expressive vocabulary tests; e.g. Vellutino, Scanlon & Spearing, 1995), generating words in an association test (semantic fluency; e.g. Nation & Snowling, 2004), making synonym judgments (e.g. Nation & Snowling, 2004), detecting common features in semantic concepts (e.g. Vellutino *et al.*, 1995), generating exemplars of a semantic category (e.g. Ben-Dror, Bentin & Frost, 1995) and verifying whether a word belongs to a certain semantic category (e.g. Ben-Dror *et al.*, 1995). In addition, 'semantic skills' has also been used to refer to the notion of higher levels of processing, that is, the ability to predict the plot of a story, extract the meaning of a story and discuss a book after reading.

The first empirical demonstration of the role of semantics in word identification is the seminal study by Meyer & Schvaneveldt (1971) on semantic priming. In their lexical decision task, participants responded faster and more accurately to the second word, *nurse*, in a semantically related word pair (*doctor* followed by *nurse*) than in a semantically unrelated word pair (e.g. *cloud* followed by *nurse*). Semantic priming studies have been carried out with both adult participants (see Neely, 1991, for a review) and children (e.g. Assink *et al.*, 2004). In general, these studies have demonstrated that target words preceded by a semantically related prime word are recognized faster and with fewer errors than words preceded by an unrelated prime word, and suggest that semantic activation appears to facilitate word identification.

A second line of evidence for the role of semantics in word identification has been provided by studies on single word reading. Words associated with multiple meanings (Azuma & Van Orden, 1997) and ambiguous words (Borowsky & Masson, 1996) are processed more quickly in a lexical decision task than words with single meanings or unambiguous words. Words without clear synonyms (Pecher, 2001), highly imageable words (Strain *et al.*, 1995) and words with a large number of features (Pexman, Lupker & Hino, 2002) are also read relatively more quickly. The general conclusion is that semantically rich words (i.e. ambiguous words, highly imageable words and words with a large number of features) are recognized faster and with fewer errors than words with fewer and less enriched semantic features.

The notion that semantics contributes to the identification of single words raises the issue of the relationship between semantic skills and reading development. Scarborough (1991), for instance, showed that a low level of semantic skills, such as vocabulary, language comprehension and narrative skills, play a role in the development of reading difficulties (see also Keenan & Betjemann, 2008). Berends & Reitsma (2006) found evidence for the superiority of semantic training. Focusing on the semantic characteristics of words improved reading fluency more than orthographic training in Grade 2 students.

A large part of experimental studies on the role of semantics in word reading has focused on semantic categorization. Semantic categorization is a means to assess taxonomic knowledge and is often used to obtain insight into the organization of the semantic memory system. Taxonomic knowledge refers to the organization of concepts in a hierarchical structure. It pertains, for example, to the knowledge that a poodle belongs to the category 'dogs' or that cats and dogs share meaning similarities because they both belong to the semantic category 'animals' or 'pets'.

It is assumed that semantic knowledge in adults is organized hierarchically, and that even young children are able to use taxonomic relationships between concepts (e.g. Nguyen & Murphy, 2003). With respect to taxonomic knowledge, a distinction can be made between horizontal and vertical relationships. Horizontal relationships refer to items at the same hierarchical level (e.g. 'cat – dog') and vertical relationships refer to items at different hierarchical levels, including the use of superordinate terms (e.g. 'animal – dog'). Lucariello, Kyratzis & Nelson (1992) showed that conventional, horizontal and vertical, taxonomic knowledge develops with age and emerges around the age of seven.

Several studies revealed a positive relationship between semantic categorization skills and reading in primary grades (e.g. Ben-Dror *et al.*, 1995; Howell & Manis, 1986; Vellutino *et al.*, 1995). Howell & 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 differed as a function of age and reading skill. Poor readers were significantly slower in making decisions than controls and this difference was larger in younger readers than in older readers. Slower semantic performance of poor readers was apparent in the pictures as well as the printed words condition, which suggests that semantic skills are not specific to written words.

Ben-Dror *et al.* (1995) addressed the relationships among semantic categorization, morphological skills, phonological skills and reading in Hebrew. In their task, children were presented with a semantic category (e.g. 'clothing'), followed by positive exemplars (e.g. 'dress') and negative exemplars (e.g. 'boat'). Participants were both typical readers and children with reading difficulties from Grade 5 and a control group of typical readers from Grade 3. Results showed that poor readers performed worse (more errors and longer reaction times) than typical readers at the same age. In short, these studies reveal that poor semantic categorization skills are associated with poor word reading skills.

There are, however, researchers who failed to demonstrate a relationship between semantic categorization and word reading. Vellutino, Scanlon & Tanzman (1990) presented poor readers and typical readers from Grades 1, 2, 4 and 6 with triads of words. They were asked to 'put together words that go together'. They used words with a horizontal semantic relationship (e.g. *roof* and *wall*), a vertical semantic relationship (e.g. *food* and *cake*) or a phonological/orthographic relationship (e.g. *rest* and *roof*). It appeared that poor and typical decoders did not differ with respect to their semantic categorization skill.

Silva-Pereyra *et al.* (2003) also challenged the claim that poor readers exhibit semantic deficits and attributed poor readers' lower performance in semantic tests to deficiencies in the processing of words, rather than to poor semantic skills per se. In their experiment, children were presented with pictures and words representing animals or non-animals (e.g. household items). They were asked whether the stimulus was an animal or not. Poor readers differed from control readers on a word categorization task, but not on a picture categorization task.

The reason for these contradicting results concerning the relationship between semantic skills and word reading might be that it changes with age. After all, this phenomenon has also been demonstrated in the relationship between word identification and phonological skills, and word identification and orthographic skills. Wagner *et al.* (1997) showed that the effect of phonological awareness on naming faded with development, whereas the relationship between orthographic skills and word reading increased with age (Juel *et al.*, 1986). Juel *et al.* found that orthographic skills explained more of the variance in word reading in second graders than in first graders.

Vellutino *et al.* (1995, Experiment 1A) studied the issue of semantic development in primary grade students. In a category verification test, a word was presented verbally and the child had to decide whether it belonged to a specific semantic category. Both typical readers and poor readers of Grades 2 and 6 performed the tests. It turned out that typical readers were generally faster than poor readers and reader group differences were smaller in second grade than in sixth grade.

To summarize, evidence for a positive relationship between semantic categorization and reading is not fully convincing. Moreover, few studies have been carried out with beginning readers and none were performed in the Dutch language. We, therefore, decided to study semantic categorization and its relationship with reading (i.e. word identification) in Dutch-speaking children from Grades 1 to 6, covering all grades of primary school.

Dutch is a relatively transparent language; that is, it has relatively unambiguous spelling-to-sound relationships. Formal reading instruction in the Netherlands starts in Grade 1 and as soon as children master the decoding rules, relatively few errors are made and variations among readers are largely marked by speed differences (for a concise description of the Dutch orthographic system and its relationship to learning to read and spell, please consult Bosman, de Graaff & Gijsel, 2006).

In short, the present study investigated the role of semantic skills in a relatively transparent language across all primary grades. Because studies of conceptual knowledge of children have yielded mixed results concerning the order of development of different kinds of relationships (e.g. horizontal and vertical relationships), we decided to use two types of semantic categorization tasks, namely, Exemplar-level tasks (Experiment 1) and Superordinate-level tasks (Experiment 2). In an Exemplar-level task, knowledge of horizontal relations is assessed, and in Superordinate-level tasks, knowledge of vertical relationships is investigated.

To ensure the participation of young readers with limited word recognition skills, stimuli were not just written words, but also spoken words and pictures; for comparison purposes among condition, response options were always pictures (e.g. Vellutino, Scanlon, DeSetto & Pruzek, 1981; Vellutino *et al.*, 1990). The different presentation conditions enabled us to investigate whether semantic skills are independent of task specifics. Additionally, we manipulated distracter type to examine the effect of phonological, semantic and perceptual characteristics. Vellutino *et al.* (1990, Experiment 1) created word pairs with related semantic/syntactic attributes (S/S attributes, such as *roof–wall* and *old–man*) and with orthographic/phonological attributes (O/P attributes, such as *room–goose*, *rest–roof*). They found that both poor and typical readers predominantly matched words based on S/S attributes rather than O/P attributes, and the percentage of S/S categorizations increased significantly from Grade 1 to Grade 4. In their second experiment, the S/S categorization principle was omitted and words could only be categorized based on either an orthographic or a phonological similarity. It turned out that typical readers used phonological similarity more often than poor readers did. Poor readers used the phonological categorization principle infrequently, suggesting that phonological attributes of words are more weakly represented in poor readers than orthographic attributes.

In both tasks of the present study, children were instructed to categorize on the basis of semantics. Apart from semantic distracters, we also included phonological distracters, because of its major role in reading, and perceptual distracters, because perceptual similarity appears to be involved in learning word meanings (e.g. Gentner & Namy, 1999). Moreover, response options were always pictures.

In sum, this study aims at investigating whether: (a) semantic categorization skills increase over time; (b) poor readers have lower semantic categorization skills than typical or good readers; and (c) poor readers and younger readers use different categorization aspects (semantic, phonological or perceptual) than typical or good readers and older readers.

EXPERIMENT 1: EXEMPLAR LEVEL

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. Both schools were randomly selected and all children were included in the experiment. The group of participants consisted of 47 children in the lower grades (Grades 1 and 2), 47 children in the intermediate grades (Grades 3 and 4), and 47 children the higher grades (Grades 5 and 6). The mean age ranged from 7.6 ($SD=0{\cdot}7$) in the lower grades to 9.5 ($SD=0.7$) in the intermediate grades, and 11.6 ($SD=0.7$) in the higher grades (Grades 5 and 6). The majority of the children (93%) were native Dutch speakers.

Materials

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

Forced-choice semantic tasks

Three Exemplar-level tasks (including a picture, spoken word or written word as target stimulus) were included. The response options for each target stimulus in each task always contained four pictures. Children were instructed to choose the picture that best matched the target stimulus. All pictures originated from 'Leesladder' [Reading Ladder] from Irausquin & Mommers (2001), a computer program for children with reading disabilities. The pictures were coloured line drawings and represented nouns which were well known by six-year-old Dutch children (Schaerlaekens, Kohnstamm & Lejaegere, 1999; familiarity rating $\geqslant 0{\cdot}80$ on a scale from 0 to 1). Moreover, only high-imageability nouns were selected (van Loon-Vervoorn, 1985; imageability rating $> 5{\cdot}5$ on a seven-point scale). All stimuli were presented on a laptop. All tasks consisted of twenty experimental trials, preceded by three practice trials, and both accuracies and reaction times were registered. A detailed description of each task is presented below.

Picture condition. 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 was a different semantic category, except for the last 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 (ten out of twenty trials), was a semantic distracter: a concept (e.g. 'egg') that belonged to the superordinate category 'food', but was not an exemplar of the subcategory 'fruit', represented by the target stimulus ('orange'). A second distracter picture was a phonological distracter (eleven out of twenty trials, e.g. *sneeuw* 'snow' for the stimulus *leeuw* 'lion') or a perceptual distracter (nine out of twenty trials, e.g. 'ball' with the stimulus 'orange'). The criterion for phonological similarity was sharing the same end-rime with the stimulus. Perceptual similarity was created by similar contours, similar perceptual features or a similar colour of the distracter to the stimulus. A third distracter picture, included in all twenty trials, was an unrelated picture (e.g. 'chair' with the stimulus 'orange'). Trials without a semantic distracter included two unrelated response options. The criteria for the unrelated concept were the absence of semantic (taxonomic or associative), perceptual and phonological similarity.

Spoken word condition. The target stimuli were members from the same categories as those in the picture task. However, the target stimuli in this task were presented verbally. Again, response options were presented by means of pictures. The identity of the target stimuli and response options differed from those in the pictures task but were of comparable difficulty, as indicated by the familiarity ratings of Schaerlaekens *et al*. (1999). The main body of the target stimuli constituted of one-syllable CVC-, CCVC- or CVCC-words (C stands for consonant, V stands for vowel); six stimuli consisted of two syllables. None of the word stimuli were semantically ambiguous. The types and distribution of the distracters were identical to the picture task, except for the distribution of phonological and perceptual distracters; this task involved exactly ten trials for each distracter type.

Written word condition. The target stimuli were derived from the same categories that were used in the picture condition and spoken word condition. However, target stimuli in this task were presented as written letter strings. Again, the identity of the target stimuli and response options differed from the other Exemplar-level tasks, but were of comparable difficulty. The main body of the target stimuli contained one-syllable CVC-, CCVC-, CVCC or CCCVC-words; two stimuli consisted of two syllables. The types and distribution of the distracters were identical to the picture condition.

Reading test

Drie-Minuten Toets [Three-Minute Test] by Verhoeven (1995). This standardized word reading test consists of three cards with three different types of words. Card 1 consists of 150 one-syllable CVC-, CV- and VC-words. Card 2 consists of 150 one-syllable CCVC-, CCVCC-, CCCVC- and CVCCCC-words. Card 3 consists of 120 polysyllabic words. Children are asked to read the words as quickly and as accurately as possible within one minute. The score is the number of items read correctly. In Grades 1 and 2, Card 1 was administered. In Grades 3 to 6, Card 3 was administered.

Apparatus and procedure

All forced-choice tasks were designed in E-prime (Schneider, Eschman & Zuccolotto, 2002). Words were recorded in 'Spraak' [Speech] by Boersma & Weenink (2004). The experiments were performed on a laptop. First, a fixation stimulus (a plus sign, 50-point Times New Roman font) was presented for 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

https://doi.org/10.1017/S0305000909990420 Published online by Cambridge University Press

best matched 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 '.' on 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 with 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 all three forced-choice tasks in groups of eight students. Each experimental forced-choice task started with three practice items. The practice items were used to explain the relationship between exemplars (e.g. 'bicycle' and 'car') and their superordinate label (e.g. 'vehicles'). There were six different orders of the subtasks in the Exemplar-level experiment (pictures, spoken words and written words). Children who were tested simultaneously received the subtasks in the same order. Trials within a subtask were randomized. Children from Grades 1 and 2 were tested in two different sessions. Children from Grades 3 to 6 performed all tasks in one single session. The reading test was administered in March (Grades 2 to 6) and May (Grade 1). To examine the effect of grade, three groups were distinguished: lower grades (Grades 1 and 2), intermediate grades (Grades 3 and 4) and higher grades (Grades 5 and 6).

RESULTS

For each participant, mean reaction times and accuracy percentages were calculated for each semantic task. For each task, incorrect trials and reaction times more than two standard deviations below or above a participant's mean reaction time were excluded from the latencies analyses.[1] The percentages of outliers in each task are listed in Table 1.

Grade level and semantic categorization

Table 2 shows the results of the Exemplar-level tasks for all children and for the lower grades, intermediate grades and higher grades, separately. A 3 (task condition: written word vs. spoken word vs. picture) $\times$ 3 (grade: lower vs. intermediate vs. higher) GLM Repeated Measures analysis was performed on reaction times and error rates with grade as between-subjects factor and

[1] 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 1. *Percentages of outliers in the Exemplar-level task removed from the analyses*

		Reaction times			Accuracy		
	n	<2 *SD*	>2 *SD*	total	practice items	errors	total
Pictures	141	0·4	7·3	7·6	13·0	19·7	32·7
Spoken	141	0·4	8·2	8·6	13·0	28·0	41
Written	141	0·4	9·3	9·8	13·0	22·3	35·3

TABLE 2. *Mean reaction times in ms., standard deviations and error percentages in the Exemplar-level task for each grade level*

	Grade level							
	Lower (*n*=47)		Intermediate (*n*=47)		Higher (*n*=47)		Total (*n*=141)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Reaction time								
written	3269	1015	2788	723	2445	666	2834	878
spoken	3759	1438	3240	886	2888	1006	3295	1183
pictures	3204	976	3088	1331	2423	751	2905	1095
Error percentage								
written	33·0	14·7	23·2	11·1	20·7	11·6	25·6	13·6
spoken	38·6	15·2	30·1	16·4	27·7	12·1	32·1	15·3
pictures	27·8	13·6	21·9	14·0	18·1	13·3	22·6	14·1

task condition as within-subjects factors. Significant effects are based on an alpha level of 0·05. The interaction effect between grade level and task condition did not reach significant levels, and neither did reaction times ($F(4, 274) = 1{\cdot}13$, $p > 0{\cdot}05$), nor errors ($F < 1$).

The main effect of task condition was significant, both for reaction times ($F(2, 137) = 17{\cdot}72$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}21$) and errors ($F(2, 137) = 29{\cdot}31$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}30$). Contrast analyses showed that error percentages were significantly higher in the spoken word condition (32·1%) than in the written word condition (25·6%) ($F(1, 138) = 29{\cdot}76$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}18$) and in the picture condition (22·6%) ($F(1, 138) = 55{\cdot}02$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}29$). Error percentages in the picture condition and written word condition also differed significantly from each other ($F(1, 138) = 6{\cdot}06$, $p = 0{\cdot}02$, *partial* $\eta^2 = 0{\cdot}04$). Reaction times were significantly higher in the spoken word condition (3295 ms) than in the written word condition (2834 ms) ($F(1, 138) = 34{\cdot}35$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}20$) and in the picture condition (2905 ms) ($F(1, 138) = 17{\cdot}48$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}11$).

https://doi.org/10.1017/S0305000909990420 Published online by Cambridge University Press

The main effect of grade level was significant, both for reaction times ($F(2, 138) = 11{\cdot}91$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}15$) and for errors ($F(2, 138) = 13{\cdot}22$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}16$). A post hoc Bonferroni test revealed that reaction times in the higher grades (2585 ms) were significantly shorter than those in the intermediate grades (3039 ms) and in the lower grades (3411 ms). Reaction times in the lower grades and intermediate grades did not differ significantly from each other ($p = 0{\cdot}09$). A post hoc Bonferroni test on error percentages revealed that error percentages in the lower grades (33·1%) were significantly higher than those in the intermediate grades (25·1%) and in the higher grades (22·2%). Error percentages in the intermediate grades and higher grades did not differ significantly from each other.

Reading ability and semantic categorization

To examine the relationship between word reading ability and semantic categorization, we selected a group of poor readers (PR) from the higher grades (Grades 5 and 6). This group of poor readers had reading scores in the lowest quartile of the distribution (scores on Three-Minute Test ranging from 62 through 82). Second, we defined a reading-level (RL) control group. This group included younger normal readers of the intermediate grades (Grades 3 and 4), who were reading at the same developmental level as the poor readers group from the higher grades. Finally, we defined a chronological age (CA) control group. This group consisted of children with the same chronological age as the poor readers in the higher grades, but with normal reading skills. The lower grades were excluded from the analyses, because in these grades a different card of the Three-Minute Test was used (Card 1 instead of Card 3). Table 3 summarizes the results of all groups.

Next, a 3 (task condition: written word vs. spoken word vs. picture) × 3 (reading group: PR-group vs. RL-control vs. CA-control) GLM Repeated Measures analysis was performed on reaction times and error percentages with reading group as between-subjects factor and task condition as within-subjects factor. Significant effects are based on an alpha level of 0·05. The interaction effect between reading group and task condition did not reach significant levels, neither for reaction times nor for errors (both $Fs < 1$).

The main effect of task condition was significant, both for reaction times ($F(2, 32) = 4{\cdot}75$, $p = 0{\cdot}02$, *partial* $\eta^2 = 0{\cdot}23$) and for errors ($F(2, 32) = 8{\cdot}89$, $p = 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}36$). Contrast analyses showed that error percentages were significantly higher in the spoken word condition (27·5%) than in the written word condition (22·1%) ($F(1, 33) = 6{\cdot}95$, $p = 0{\cdot}01$, *partial* $\eta^2 = 0{\cdot}17$) and in the picture condition (18·3%) ($F(1, 33) = 16{\cdot}67$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}34$). Reaction times were significantly higher in the spoken word condition (2860 ms) than in the written word condition (2466 ms)

TABLE 3. *Mean age, reading scores, reaction times in ms. (standard deviations in parentheses), and error percentages in the Exemplar-level task for each reading level*

	Reading group		
	Poor readers	CA-control group	RL-control group
N	12	12	12
Age	11·7 (0·9)	11·7 (0·7)	9·7 (0·7)
Reading score	76·7 (5·9)	89·2 (9·0)	75·4 (7·5)
Reaction Time			
Pictures	2384	2151	2635
Spoken	3010	2793	2776
Written	2683	2220	2495
Error percentage			
Pictures	19·6 (8·7)	16·3 (14·0)	19·2 (10·6)
Spoken	30·0 (13·3)	24·2 (10·0)	28·3 (12·5)
Written	23·3 (14·2)	20·4 (13·0)	22·5 (10·3)

($F(1, 33) = 7{\cdot}15$, $p = 0{\cdot}01$, *partial* $\eta^2 = 0{\cdot}18$) and in the picture condition (2390 ms) ($F(1, 33) = 8{\cdot}62$, $p = 0{\cdot}006$, *partial* $\eta^2 = 0{\cdot}21$). The main effect of reading group was not significant, neither for reaction times, nor for errors (both $Fs < 1$).

Distracter options

Third, we investigated the effect of distracter type. To that end, we computed for each of the distracter types (i.e. phonological, perceptual and semantic) the percentages of distracter choices. This refers to the number of actual distracter choices made by the children divided by the number of potential distracter options of this type, multiplied by 100. Table 4 shows for each distracter type the percentage of distracter choices.

Subsequently, we examined whether the effect of distracter type varied among readers of different grade levels (lower vs. intermediate and higher grades) and different reading groups (PR-group vs. RL-control vs. CA-control).

Distracter type and grade level. A GLM Repeated Measures analysis with task condition (spoken word vs. picture vs. written word) and distracter type (semantic, phonological, perceptual) as within-subjects variables, and grade level (lower vs. intermediate vs. higher) as a between-subjects variable was performed on the percentages of distracter choices ($N = 141$).

The interaction between task condition and distracter type was significant ($F(4, 135) = 21{\cdot}56$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}39$). Separate analyses on the spoken word, written word, and picture condition showed significant main effects of distracter type in all three task conditions: spoken word

https://doi.org/10.1017/S0305000909990420 Published online by Cambridge University Press

TABLE 4. *Percentages of distracter choices for each distracter type*

	Task condition					
	Spoken words		Pictures		Written words	
	M	*SD*	*M*	*SD*	*M*	*SD*
Distracter						
Semantic						
sem, phon, unrel	25·4	15·7	7·8	10·3	26·8	16·2
Sem, perc, unrel	22·6	18·3	18·0	18·4	7·6	14·0
Total	24·0	12·5	11·9	9·5	19·2	10·8
Phonological						
Phon, sem, unrel	8·2	16·4	3·9	10·5	15·1	19·0
Phon, unrel, unrel	8·5	16·6	5·5	14·2	4·1	12·8
Total	8·4	14·9	4·6	10·3	10·1	14·0
Perceptual						
Perc, sem, unrel	19·9	22·3	13·3	21·0	6·6	14·2
Perc, unrel, unrel	26·0	19·7	28·2	20·3	23·1	17·1
Total	22·9	17·4	21·6	16·5	15·8	12·5

NOTE: Percentages of distracter choices have been provided for each combination of response options. 'Sem' is semantic, 'phon' is phonological, 'unrel' is unrelated.

($F(2, 137) = 55{\cdot}54$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}49$); picture ($F(2, 137) = 79{\cdot}43$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}54$); written word ($F(2, 137) = 15{\cdot}80$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}19$). In the picture and in the written word condition, all percentages of distracter choices differed significantly from each other. However, in the spoken word condition, the percentages of perceptual and semantic distracter choices did not differ significantly from each other.

The main effect of task condition was significant ($F(2, 137) = 29{\cdot}60$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}30$). Percentage of distracter choices in the spoken word condition (18·4%) were significantly larger than those in the written word condition (15%) and in the picture condition (12·7%), which in turn also differed significantly from each other.

The main effect of distracter type was also significant ($F(2, 137) = 71{\cdot}23$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}51$). Percentages of phonological distracter choices (7·7%) were significantly lower than those of semantic (18·4%) and perceptual distracter choices (20·1%), which in turn did not differ significantly from each other.

Finally, the main effect of grade was also significant ($F(2, 138) = 12{\cdot}67$, $p < 0{\cdot}001$, *partial* $\eta^2 = 0{\cdot}16$). Percentages of distracter choices in the lower grades (18·8%) were significantly higher than those in the intermediate grades (14·5%) and those in the higher grades (12·9%). Percentages of distracter choices of the intermediate grades and higher grades did not differ significantly from each other.

Distracter type and reading skill. A GLM Repeated Measures analysis with task condition (spoken word vs. picture vs. written word) and distracter type (phonological, perceptual, semantic) as within-subjects variables and reading group (PR-group vs. RL-control vs. CA-control) as between-subjects variable was performed on the percentages of distracter choices ($N=36$).

The interaction effect between task condition and distracter type was significant ($F(4, 30)=7{\cdot}55$, $p<0{\cdot}001$, *partial* $\eta^2=0{\cdot}50$). Separate analyses on the spoken word, picture, and written word condition showed significant main effects of distracter type in all three task conditions: spoken word ($F(2, 32)=21{\cdot}45$, $p<0{\cdot}001$, *partial* $\eta^2=0{\cdot}57$); picture ($F(2, 32)=21{\cdot}11$, $p<0{\cdot}001$, *partial* $\eta^2=0{\cdot}57$); written word ($F(2, 32)=9{\cdot}22$, $p=0{\cdot}001$, *partial* $\eta^2=0{\cdot}37$). In the picture condition, all percentages of distracter choices differed significantly from each other. However, in the written word condition and spoken word condition, percentages of perceptual and semantic distracter choices did not differ significantly from each other.

The main effect of task condition was significant ($F(2, 32)=6{\cdot}61$, $p=0{\cdot}004$, *partial* $\eta^2=0{\cdot}29$). Percentages of distracter choices in the spoken word condition (16·1%) were significantly larger than those in the written word condition (13·4%) and picture condition (10·6%), which in turn did not differ significantly from each other.

The main effect of distracter type was also significant ($F(2, 32)=34{\cdot}7$, $p<0{\cdot}001$, *partial* $\eta^2=0{\cdot}68$). Percentages of phonological distracter choices (5·2%) were significantly lower than those of semantic (16·6%) and perceptual distracter choices (18·2%), which in turn did not differ significantly from each other.

CONCLUSION

To sum up, in Experiment 1, readers from Grades 1 to 6 performed three Exemplar-level tasks. Results demonstrated longer reaction times and more errors in the spoken word condition than in the written word and the picture condition. The lower grades showed longer reaction times and more errors than the higher grades. No significant differences in semantic categorization skills were found between poor readers, chronological age matched readers and reading level matched readers.

The effect of distracter type (phonological, perceptual, semantic) on percentages of distracter choices of children of different grade levels and different reading skills showed that the percentage of phonological distracter choices was significantly lower than the number of semantic distracter choices and perceptual distracter choices. In the picture condition, the difference between percentages of semantic distracter choices and perceptual distracter choices also reached significance. In the written word condition, this difference reached significance in the analyses on grade. Neither grade level nor reading skills influenced the effect of distracter type. In other words, children from different grade levels and different reading skills were not differently affected by the phonological, perceptual or semantic distracters.

EXPERIMENT 2: SUPERORDINATE LEVEL

In Experiment 2, we address the same issues as in Experiment 1. However, instead of using an Exemplar-level task to assess semantic categorization skills, we used a Superordinate-level task, which measures knowledge of vertical relationships. We investigated whether semantic categorization skills increase over time, and whether poor readers have lower semantic categorization skills than typical or good readers. Because the Superordinate-level tasks included only semantic distracters and unrelated distracters, we decided not to investigate the effect of distracter type.

METHOD

Participants

The same children as in Experiment 1 participated in Experiment 2, except for the children in the lower grades. The Superordinate-level tasks included multisyllable words like *muziekinstrumenten* 'musical instruments' and *vervoersmiddelen* 'means of transportation'. These words require a minimum level of reading skills not yet mastered by children in Grades 1 and 2. Therefore, children from the lower grades were excluded. The number of participants in Experiment 2 consisted of 94 children: 47 children in the intermediate grades and 47 children in the higher grades.

Materials, apparatus and procedure

The materials comprised two forced-choice semantic tasks to assess taxonomic superordinate knowledge, and the word-reading test that was described in Experiment 1. Both experimental tasks consisted of twenty-five experimental trials and were preceded by two practice items. Accuracy as well as reaction times was measured. The same apparatus and procedure that was used in Experiment 1 was used in this experiment.

Forced-choice semantic tasks

Spoken word condition. The 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

https://doi.org/10.1017/S0305000909990420 Published online by Cambridge University Press

TABLE 5. *Percentages of outliers in the Superordinate-level task removed from the analyses*

		Reaction times			Accuracy		
	n	<2 *SD*	>2 *SD*	total	practice items	errors	total
Spoken	141	0·1	8·4	8·5	7·4	16·0	23·4
Written	94	0·0	8·4	8·4	7·4	10·1	17·5

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, fifteen stimuli included two semantic distracters and one unrelated picture in the response options. Ten stimuli included three unrelated pictures in the response options. Criteria for these distracters have been described earlier.

Written word condition. This test consisted of twenty-five experimental trials, preceded by two practice items. Categories that were included were identical to the spoken word task. However, the stimuli in this subtask were presented by written words. Response options were all pictures. The identity of the stimuli and response options differed from the spoken word condition, but were of comparable difficulty (see 'Results' section). The types and distribution of distracters was the same as those included in the spoken word condition. Both accuracy and reaction time were measured.

RESULTS

For each participant, mean reaction times and accuracy percentages were calculated for each semantic task. For each task, incorrect trials and reaction times more than two standard deviations below or above a participant's mean reaction time were excluded from the latencies analyses.[1] The percentages of outliers in each task are listed in Table 5.

Grade level and semantic categorization

Table 6 shows the results of the Superordinate-level tasks for all children and for the intermediate grades (Grades 3, 4) and higher grades (Grades 5, 6) separately. A 2 (task condition: written words vs. spoken words) × 2 (Grade: intermediate vs. higher) ANOVA was performed on reaction times and error rates with grade as between-subjects factor and task condition as within-subjects factor. Significant effects are based on an alpha level of 0·05. The interaction effect between grade level and task condition did not reach

TABLE 6. *Mean reaction times in ms., standard deviations and error percentages in the Superordinate-level task for each grade level*

	Grade level					
	Intermediate ($n=47$)		Higher ($n=47$)		Total ($n=141$)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Reaction time						
written	2153	535	1683	345	1918	506
spoken	2088	390	1745	301	1917	387
Error percentage						
written	13·2	8·5	8·5	6·7	10·9	8·0
spoken	15·0	10·2	10·6	6·9	12·8	8·9

significant levels, neither for reaction times ($F(1, 92)=3{\cdot}07$, $p>0{\cdot}05$) nor for errors ($F<1$).

The main effect of task condition was significant only for errors ($F(1, 92)=4{\cdot}23$, $p=0{\cdot}04$, *partial* $\eta^2=0{\cdot}04$). Error percentages were significantly higher in the spoken word condition (12·8%) than in the written word condition (10·9%).

The main effect of grade was also significant, both for reaction times ($F(1, 92)=29{\cdot}58$, $p<0{\cdot}001$, *partial* $\eta^2=0{\cdot}24$) and for errors ($F(1, 92)=10{\cdot}42$, $p=0{\cdot}002$, *partial* $\eta^2=0{\cdot}10$). Reaction times in the higher grades (1714 ms) were significantly shorter than those in the intermediate grades (2120 ms). Error percentages in the higher grades (9·6%) were significantly lower than those in the intermediate grades (14·1%).

Reading ability and semantic categorization

To examine the relationship between reading and semantic categorization skills, the selection criteria used in Experiment 1 were applied here to compose a poor reader group (PR), a reading level (RL) control group, and a chronological age (CA) control group. Table 7 summarizes the results of all groups.

Next, a 2 (task condition: written words vs. spoken words) × 3 (reading group: PR-group vs. RL-control vs. CA-control) GLM Repeated Measures analysis was performed on reaction times and error rates with reading group as between-subjects factor and task condition as within-subjects factor. Significant effects are based on an alpha level of 0·05. The interaction effect between reading group and task condition did not reach significant levels, neither for reaction times nor for errors (both $Fs<1$).

The main effect of task condition was also not significant, neither for reaction times ($F(1, 33)=1{\cdot}71$, $p>0{\cdot}05$) nor for errors ($F(1, 33)=3{\cdot}81$,

https://doi.org/10.1017/S0305000909990420 Published online by Cambridge University Press

TABLE 7. *Mean age, reading scores, reaction times in ms (standard deviations in parentheses) and error percentages in the Superordinate-level tasks for each reading level*

	Reading group		
	Poor readers	CA-control group	RL-control group
N	12	12	12
Age	11·7 (0·9)	11·7 (0·7)	9·7 (0·7)
Reading score	76·7 (5·9)	89·2 (9·0)	75·4 (7·5)
Reaction Time			
Spoken	1672	1686	2011
Written	1655	1562	1972
Error Percentage			
Spoken	10·7 (6·0)	10·0 (7·7)	16·3 (15·9)
Written	7·7 (5·8)	7·7 (5·8)	10·7 (5·7)

$p > 0{\cdot}05$). The main effect of reading group was only significant for reaction times ($F(2, 33) = 6{\cdot}52$, $p = 0{\cdot}004$, *partial* $\eta^2 = 0{\cdot}28$). Reaction times of the RL-control group (1991 ms) were significantly larger than those of the PR-group (1663 ms) and CA-control group (1624 ms).

CONCLUSION

To sum up, in Experiment 2, readers from Grades 3 to 6 performed two Superordinate-level tasks. More errors were made in the spoken word condition than in the written word condition. Moreover, results demonstrated longer reaction times and more errors in the intermediate grades than in the higher grades. The effect of reading skill on performance of the Superordinate-level tasks was visible only in the reaction time analyses. The children in the (younger) reading level matched group were slower than those in the older poor reader group and in the chronological age matched group.

GENERAL DISCUSSION

The focus of the present study was the development of semantic categorization across primary grades and its relationship with reading. Three Exemplar-level tasks (Experiment 1) and two Superordinate-level tasks (Experiment 2) with different types of distracters (phonological, semantic, perceptual, unrelated) were administered to assess semantic categorization skills.

Before discussing the main question of this study, we address how children's semantic categorization performance changes with task requirements.

https://doi.org/10.1017/S0305000909990420 Published online by Cambridge University Press

Children took longer to respond and committed more errors in the spoken word condition than in the picture and in the written word condition in both the Exemplar-level and the Superordinate-level task. The lower performance in the spoken word condition is most likely due to the nature of the stimulus category. In the spoken word condition, the presentation of stimulus category is relatively transient compared to the written word and the picture condition. In these last two conditions, the stimulus remained visible until the child decided to ask for the answer options, whereas in the spoken word condition the answer options appeared after the stimulus had been presented. Thus, in the spoken word condition, the stimulus naturally disappeared, whereas in the written word and in the picture condition, the children had control over the disappearance of the stimulus. This implies that the spoken word condition required careful attention and a good short-term memory compared to the other two conditions in which the child could influence the time span of the stimulus presentation.

Not only did task condition affect semantic categorization, but also distracter types influenced children's performance. In the Exemplar-level tasks, children chose the phonological distracter significantly less frequently than the perceptual distracter or semantic distracter. It appears that in the semantic categorization tasks developed here, children are less inclined to categorize words on the basis of their phonological information when pictures are included as response options.

The major research question of the present study was whether semantic categorization is affected by grade and/or reading level. First, in the Exemplar-Level tasks as well as in the Superordinate-level tasks, higher grades were faster and more accurate in all task conditions. This strongly suggests that the test is sensitive to developmental changes. Increasing accuracy in semantic categorization with age is in line with previous studies (e.g. Nguyen & Murphy, 2003). Moreover, in the Superordinate-level tasks, the reaction times of younger typical readers were larger than those of older poor readers and chronological age matched readers. Recall that, in the Superordinate-level tasks, knowledge of vertical taxonomic relations between concepts was assessed. For example, the stimulus *clothes* was presented and the child had to indicate which picture belonged to the semantic category *clothes*, with *trousers* as the correct response and *bag, cloth and box* as distracters. It appears that knowledge of vertical taxonomic relationships increases with age, and this development is independent of reading skills. During primary grades, children's conceptual knowledge increases over time and semantic structures become more specified.

Second, there was no relationship between reading level and semantic categorization performance. In all task conditions, poor readers showed similar semantic categorization performance as typical readers. This result is in line with the results of Silva-Pereyra *et al.* (2003). They also failed to find a relationship between figure categorization skills and reading. Silva-Pereyra *et al*. demonstrated both with behavioural data and with ERP recordings that poor readers performed worse than normal readers on a word categorization task, but not on a picture categorization task. These and the present finding suggest that there is no reason to assume that poor readers have incomplete semantic networks (expressed in error percentages) or have slower access to semantic memory (expressed in reaction times) than typical readers. Note that in the present study, semantic categorization times were measured separately from the time it took to encode the (written) stimulus; this experimental choice disentangles semantic access from reading performance to a large degree.

Some authors, however, have suggested that semantic processing is actually the preferred strategy in poor readers, if semantic information is available. Waterman & Lewandowski (1993), for example, orally presented participants with a set of words. After listening to the list of words, they were presented with a new set of words and had to decide whether each word they heard had been in the first list. The test words 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. Thus, both poor and good readers more often incorrectly reported that they had heard a semantic target (e.g. *sack*) than a semantic control (e.g. *mess*). However, results also showed that the 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. From these results, the authors concluded that the semantic processing system of poor readers is intact and might even be the preferred coding strategy in poor readers. Because our poor readers' taxonomic knowledge was similar to the taxonomic knowledge of good readers, our results also support the conclusion that the semantic processing of poor readers is highly similar to that of good readers.

To conclude, our study does not provide evidence for a positive relationship between semantic categorization skills and word reading skills. Of course, we do not believe that these findings can be generalized to text reading, because it is generally assumed that semantic skills such as vocabulary do play a role in text reading and text comprehension.

Our study did not provide evidence for the superiority of semantic skills in poor decoders, as suggested by Waterman & Lewandowski (1993). Note, however, that the procedure in their study was different from ours. Their task addressed implicit semantic knowledge (i.e. on-line procedures), whereas the semantic categorization tasks in our study required purposive semantic processing (i.e. off-line procedures). Moreover, the authors were unable to replicate and extend their findings in a subsequent study when stimuli were presented visually, rather than verbally (Waterman & Lewandowski, 1994). In the latter study, poor readers and good readers were equally affected by semantic confusability. Several researchers (e.g. Assink *et al.*, 2004) have argued that off-line tasks are less suitable for examining implicit semantic knowledge, because they are more directly linked to word reading than conscious processes are, suggesting that semantic priming studies might be a better alternative. Assink *et al.* carried out a semantic priming experiment with poor readers and controls and found no sensitivity differences to semantic cues between poor readers and typical readers. Although semantic processes can be assessed either in a direct way (purposive, off-line processing) or in a rather indirect way (e.g. semantic priming), the two methods appear to address different processes. We believe that both on-line procedures and off-line procedures are important in the study of the semantic processing of poor readers, because deficiencies in purposive semantic processing do not automatically imply deficits in implicit semantic processing and vice versa.

Finally, we examined whether poor readers and younger readers were more distracted by other characteristics of concepts (phonology vs. semantics vs. shape) than good readers or older readers were. Neither grade level nor reading skills affected the outcomes. Thus, poor readers were no differently affected by the phonological, perceptual or semantic distracters than good readers. This result is in accordance with the results of Waterman & Lewandowski (1994). They manipulated phonological, orthographic and semantic confusability. The task for the poor and good readers was to indicate whether stimuli had been visually presented a second time. Poor readers and good readers were equally affected by orthographic and semantic confusability. Phonological confusability was not observed for either of the groups. These results, together with the results of the semantic priming studies (e.g. Assink *et al.*, 2004), suggest that semantic processing in poor readers is qualitatively similar to that of good readers. This mirrors the results on reading tasks in relatively transparent languages: children with dyslexia in the Netherlands show identical patterns of reading errors to normal readers and they are equally affected by spelling (in)consistencies as normal readers (e.g. Bosman, Vonk and van Zwam, 2006).

To conclude, the semantic categorization skills of poor and good readers are equally advanced, suggesting that semantic processing in poor readers is not deficient. This finding needs to be incorporated in theories of word reading. As noted in the 'Introduction', almost all models of reading acknowledge the fundamental and interactive role of phonology, orthography and semantics. If semantic processing in poor readers is not dissimilar to that of good readers, poor reading needs to be explained from the other aspects that reading involves. All models that assume recurrent interaction between all three aspects that govern reading are able to explain this phenomenon (e.g. Bosman & Van Orden, 1997; Harm & Seidenberg, 1999; Van Orden *et al.*, 1990). Interactive models cause the presentation of a printed word to activate letter nodes, which in turn activate phoneme and semantic nodes. Following initial activation, recurrent feedback dynamics begin among all these node families. It is assumed that, in poor readers, the connections between the phonological nodes and orthographic nodes are inefficient or less stable. However, because of the interconnectivity between all nodes, activation of semantic information may support the word identification process. Farrar & Van Orden (2001) investigated this assumption in a simulation study on English words, and indeed found that intact semantics supports the reading process, in case connections between phonology and orthography are weak or even absent. To substantiate the claim of semantics supporting the reading process, cross-linguistic research is needed to answer the question of whether semantics also plays a role in more transparent languages such as Dutch. After all, the relationships between semantics, orthography and phonology might differ between deep and shallow orthographies, resulting in different findings concerning the role of semantic skills in word identification.

In sum, the results of this study demonstrated a strong development (in both speed and accuracy) in semantic skills through primary grades. The development was apparent in all Exemplar-level and Superordinate-level tasks. No differences emerged in semantic performance between poor readers and typical readers, and the effect of distracter type was the same for readers of different grade levels and different reading skills. We, therefore, reject the notion of a qualitative difference in taxonomic knowledge between reader groups with respect to word reading. Identical error patterns of all reader groups and similar preferences for different types of distracters support this view.

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. *Journal of Genetic Psychology* **165**, 67–79.

Azuma, T. & Van Orden, G. C. (1997). Why SAFE is better than FAST: The relatedness of a word's meanings affects lexical decision times. *Journal of Memory and Language* **36**, 484–504.

Barker, T. A., Torgesen, J. K. & Wagner, R. K. (1992). The role of orthographic processing on five different reading tasks. *Reading Research Quarterly* **27**(4), 334–45.

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**, 876–93.

Berends, I. E. & Reitsma, P. (2006). Addressing semantics promotes the development of reading fluency. *Applied Psycholinguistics* **27**, 247–65.

Berent, I. & Perfetti, C. A. (1995). A Rose is a REEZ: The two-cycles model of phonology assembly in reading English. *Psychological Review* **102**, 146–84.

Boersma, P. & Weenink, D. (2004). Praat: Doing phonetics by computer. University of Amsterdam: Institute of Phonetic Sciences.
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 Graaff, S. & Gijsel, M. A. R. (2006). Double Dutch: the Dutch spelling system and learning to spell in Dutch. In R. M. Joshi & P. G. Aron (eds), *Handbook of orthography and literacy*, 135–50. Mahwah, NJ: Lawrence Erlbaum Associates.
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*, 165–87. 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 & M. Fayol (eds), *Learning to spell: Research, theory, and practice across languages*, 173–94. Hillsdale, NJ: Lawrence Erlbaum Associates.
Bosman, A. M. T., Vonk, W. & van Zwam, M. (2006). Spelling consistency affects reading in students with and without dyslexia. *Annals of Dyslexia* **56**, 271–300.
Farrar, W. T. & Van Orden, G. C. (2001). Errors as multistable response options. *Nonlinear Dynamics, Psychology, and Life Sciences* **5**, 223–65.
Gentner, D. & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development* **14**, 487–513.
Harm, M. W. & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights form connectionist models. *Psychological Review* **106**, 491–528.
Howell, M. J. & Manis, F. R. (1986). Developmental and reader ability differences in semantic processing efficiency. *Journal of Educational Psychology* **78**, 124–29.
Irausquin, R. & Mommers, C. (2001). Leesladder. Een programma voor kinderen met leesmoeilijkheden. [Reading ladder. A program for children with reading difficulties]. Tilburg: Zwijsen.
Juel, C., Griffith, P. L. & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology* **78**, 243–55.
Keenan, J. M. & Betjemann, R. S. (2008). Comprehension of single words: The role of semantics in word identification and reading disability. In E. L. Grigorenko & A. J. Naples (eds), *Single-word reading: Behavioral and biological perspectives*, 191–209. Mahwah, NJ: Erlbaum.
Lucariello, J., Kyratzis, A. & Nelson, K. (1992). Taxonomic knowledge: What kind and when? *Child Development* **63**, 978–98.
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**, 227–34.
Nation, K. & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading* **27**, 342–56.
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*, 264–336. Hillsdale, NJ: Erlbaum.
Nguyen, S. P. & Murphy, G. L. (2003). An apple is more than just a fruit: Cross-classification in children's concepts. *Child Development* **74**, 1783–806.
Pecher, D. (2001). Perception is a two-way junction: Feedback semantics in word recognition. *Psychonomic Bulletin & Review* **8**, 545–51.
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**, 542–49.
Scarborough, H. S. (1991). Antecedents to reading disability: Preschool language development and literacy experiences of children from dyslexic families. *Reading and Writing* **3**, 219–33.

Schaerlaekens, A., Kohnstamm, G. A. & Lejaegere, M. (1999). *Streeflijst woordenschat voor zesjarigen: Derde herziene versie gebaseerd op nieuw onderzoek in Nederland en België*. [Target list vocabulary for 6-year-old children]. Lisse: Swets & Zeitlinger.

Schneider, W., Eschman, A. & Zuccolotto, A. (2002). E-Prime. Pittsburgh, PA: Psychology Software Tools.

Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review* **96**, 523–68.

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**, 187–99.

Stanovich, K. E. & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly* **24**, 402–433.

Strain, E., Patterson, K. & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **21**, 1140–54.

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.

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., DeSetto, L. & Pruzek, R. M. (1981). Developmental trends in the salience of meaning versus structural attributes of written words. *Psychological Research* **43**, 131–53.

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. and 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.

Verhoeven, L. (1995). *Drie-Minuten-Toets*. [Three-Minute Test]. Arnhem: CITO.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J. & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology* **33**, 468–79.

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.

Waterman, B. & Lewandowski, L. (1994). Orthographic, phonologic, and semantic processing in reading disabled and nondisabled subjects. *Perceptual and Motor Skills* **79**, 35–45.