Cognitive Component Skills of Reading Comprehension in Developing Readers

Dissertation
Zur Erlangung des akademischen Grades
Doktor der Philosophie (Dr. phil.)

Eingereicht am Fachbereich 01 Humanwissenschaften
der Universität Kassel

Vorgelegt von
Julia Knoepke, M.A.
Juni 2016
Erstgutachter: Prof. Dr. Tobias Richter
Zweitgutachterin: Prof. Dr. Cordula Artelt
Drittgutachterin: Prof. Dr. Mirjam Ebersbach

For my parents
# Table of Contents

**Summary** .......................................................................................................................... 5  
**Chapter I: Introduction and Aims** .................................................................................. 7  
**Chapter II: Study 1** .......................................................................................................... 43  
Phonological recoding, orthographic decoding, and comprehension skills during reading acquisition (Knoepke, Richter, Isberner, Naumann, & Neeb, 2014)  
**Chapter III: Study 2** ......................................................................................................... 71  
Leseverstehen = Hörverstehen X Dekodieren? Ein stringenter Test der Simple View of Reading bei deutschsprachigen Grundschulkindern [Reading comprehension = listening comprehension X decoding? A stringent test of the simple view of reading in German primary school children] (Knoepke, Richter, Isberner, Neeb, & Naumann, 2013)  
**Chapter IV: Study 3** .......................................................................................................... 93  
Processing of positive-causal and negative-causal coherence relations in primary school children and adults: A test of the cumulative cognitive complexity approach in German. (Knoepke, Richter, Isberner, Naumann, Neeb, & Weinert, 2016)  
**Chapter V: Study 4** .......................................................................................................... 126  
Construct validity of a process-oriented test assessing skills of syntactic integration in German primary school children (Knoepke, Richter, Isberner, Naumann, & Neeb, submitted)  
**Chapter VI: Sources of Poor Reading Comprehension** ................................................. 154  
Reading comprehension: Individual differences, disorders, and underlying cognitive processes (Knoepke, & Richter, in press)  
**Chapter VII: General Discussion** .................................................................................... 174  
**References** ......................................................................................................................... 201  
**Erklärung zum Eigenanteil** ............................................................................................. 227  
**Eidesstattliche Versicherung und Erklärung** ................................................................. 230  
**Acknowledgements** ........................................................................................................ 231
Summary

The ability to comprehend written text, that is, the ability to establish a coherent mental representation of text content, is one of the necessary preconditions for successful educational and career development in a literate society. Poor readers not only face severe difficulties in everyday life when they come into touch with e-mails, text messages, official letters, and forms. Their career prospects are greatly reduced by the inability to adequately comprehend written language. Thus, one of the most important and pressing goals of the educational system is to identify poor readers as early as possible to develop individualized and target-oriented intervention programs to foster and enhance their reading abilities (Artelt & Dörfler, 2010). This task requires profound knowledge of the complex concept of reading competence, its acquisition, and development. Although it is commonly assumed that reading competence comprises several cognitive component skills at the word, sentence, and text level (e.g., Lenhard & Artelt, 2009; Müller & Richter, 2014; Perfetti, Landi, & Oakhill, 2005; Richter & Christmann, 2009), and that individual differences in these skills differentiate skilled and poor readers (e.g., Cain & Oakhill, 2004), several questions concerning the cognitive component skills of reading comprehension, their interrelatedness, and their development in beginning readers remain unanswered. The aim of this dissertation was to advance a more comprehensive understanding of the complex concept of reading comprehension. It contributes to existing literature and empirical research by exploring a subset of open research questions on the cognitive component skills of reading comprehension, their interrelatedness, and their development in beginning German readers.

Study 1 examined the extent that phonological recoding and orthographical decoding skills contribute to reading comprehension at the sentence and text level and how both skills develop in German 2nd to 4th graders. Results suggest that both visual word recognition skills made significant and unique contributions to reading comprehension and that their relative weight did not change across grade levels. Orthographical decoding skills explained substantially more variance in reading comprehension than phonological recoding skills at all grade levels, suggesting that, by the end of Grade 2, German readers are already able to recognize the majority of written words in age-appropriate texts directly via orthographical decoding. Nevertheless, the results indicate that German readers continuously use phonological information at all grade levels to optimize visual word recognition.

Study 2 amplifies empirical research on one of the most prominent models of reading comprehension—the simple view of reading (Gough & Tunmer, 1986; Hoover & Gough,
Summary

The study tested the simple view of reading \((R = D \times C)\) by using optimized and methodologically stringent measures of the model constituents and probed its generalizability to German developing readers in Grades 3 and 4. It was shown that the simple view of reading in its most simplistic form could not withstand a methodologically stringent test and was not readily applicable to German 3rd and 4th graders. The results provided only weak evidence in favor of a multiplicative combination of decoding skills \((D)\) and listening comprehension skills \((C)\) as proposed by the simple view. The fact that a considerable amount of variance in reading comprehension \((R)\) could not be explained by \(D\) and \(C\) indicates that the model might not be complete and needs to be extended by further components.

Study 3 investigated the establishment of positive-causal and negative-causal coherence relations in German developing readers and adults in written and spoken language comprehension. Consistent with the cumulative cognitive complexity approach by Evers-Vermeul and Sanders (2009) and Spooren and Sanders (2008), accuracy data in a semantic verification task with sentence-pairs revealed that the processing of negative-causal coherence relations was cognitively more demanding than the processing of positive-causal coherence relations in German 1st to 4th graders and adults. Moreover, developmental trends in the processing of negative-causal coherence relations seem to lag behind developmental trends in the processing of positive-causal coherence relations and the comprehension of negative-causal coherence relations seems to be still developing by the end of Grade 4.

Finally, Study 4 aimed to demonstrate the usefulness of process-oriented reading tests that assess individual differences in cognitive component skills of reading comprehension selectively. Explanatory item-response models provided empirical evidence in favor of the construct validity of the ProDi-L and ProDi-H grammaticality judgment tasks, which assess syntactic integration skills in German primary school children. Results indicate that both tasks are valid diagnostic instruments to assess individual differences in syntactic integration skills in spoken and written language processing in German 1st to 4th graders and to detect children with exceptionally poor syntactic integration skills.

The findings reported in this dissertation constitute a step on the way to a more comprehensive understanding of the cognitive component skills of reading comprehension in developing readers. Such a comprehensive understanding is indispensable for an optimal conception of reading lessons, the construction of learning materials, reading instructions, and textbooks. Most importantly, it is the basis for a meaningful diagnosis of individual patterns of reading deficits and for the construction of adaptive and target-oriented reading interventions to improve reading comprehension in poor readers.
Chapter I

Introduction and Aims
Introduction and Aims

There is general agreement that, at least up to a basic level, being able to read is essential for life, not just for the individual but also for the well-being of the whole society (UNESCO, 2005). In order to maximize individual life chances, every child should be given the possibility to learn to read and to be able and motivated to use this skill effectively and on a high level. For providing such learning opportunities, researchers as well as educators need to understand how individuals acquire the ability to read and why some learn and practice it so successfully whereas others struggle or fail.

- Pfost, Artelt, & Weinert (2013, p. 7)

Written language is practically unavoidable in everyday life. Whether as advertised taglines, street signs, price tags, ingredients on groceries, text messages, e-mails, instruction manuals, newspapers, novels, or educational textbooks, written information is often processed automatically and with little effort. The ability to comprehend written language, that is, the ability to establish a coherent mental representation of text content, is one of the preconditions for successful educational and career development in a literate society. Given the crucial role of reading comprehension in daily life, the fact that more than 15% of the primary school children in Germany are not able to draw simple inferences from texts, to link distributed information, and to integrate text information with prior knowledge into a coherent mental representation of the text by the end of Grade 4 is alarming (IGLU, Bos, Tarelli, Bremerich-Vos, & Schwippert, 2012). Consequently, poor readers are impeded in several educational achievement outcomes such as gaining new information from textbooks, second language acquisition, or solving word problems. This lack is particularly worrying, because the achievement gap between good and poor readers tends to become even larger during further reading development (Matthew effects, Stanovich, 1986; see also Pfost, Dörfler, & Artelt, 2012; see Pfost, Hattie, Dörfler, & Artelt, 2014 for a critical review and discussion of Matthew effects in reading development). Good readers tend to get better throughout reading development, whereas poor readers tend to develop very slowly or stagnate. Under these circumstances, the persistence of poor reading skills into adulthood is not surprising. Almost 15% of 15-year-olds in Germany are not capable of interpreting written text, drawing inferences, establishing logical relations, linking distant information, generating text-related opinions, and comprehending conflicting information (PISA, Hohn, Schiepe-Tiska, Sälzer, & Artelt, 2012). More than 14% of the working age population in Germany are functional illiterates who suffer from deficient text comprehension (i.e., they are not able to extract
explicit information even from short texts) despite adequate word and sentence level comprehension (Leo. - Level-One Studie, Grotlüschen & Riekmann, 2011). Apart from severe difficulties in everyday life caused by illiteracy, career prospects are greatly reduced by the inability to adequately comprehend written language. Thus, one of the most important and pressing goals of the educational system is to identify poor readers as early as possible during reading acquisition and to develop individualized and target-oriented intervention programs to foster and enhance their reading abilities (Artelt & Dörfler, 2010).

This task requires a profound knowledge of the complex concept of reading competence, its acquisition, and development. In recent decades, numerous studies have been published in various disciplines, including cognitive, developmental, and educational psychology and psycholinguistics. Researchers commonly assume that reading competence comprises several cognitive component skills at the word, sentence, and text level (e.g., Lenhard & Artelt, 2009; Müller & Richter, 2014; Perfetti, 1999, 2001; Perfetti et al., 2005; Richter & Christmann 2009) and that individual differences in these skills differentiate skilled and poor readers. At the word level, skilled readers recognize written word forms via two different routes (see Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). To recognize infrequent or unknown word forms, readers use grapheme-to-phoneme conversion to translate words grapheme-by-grapheme into a phonological code (*phonological recoding*), which is then used to access entries in the mental lexicon. In contrast, orthographic forms of frequent and well-known words are recognized directly (*orthographical decoding*) and can be directly mapped onto entries in the mental lexicon without the necessity of grapheme-to-phoneme conversion. In addition to the visual recognition of a word form, its lexical entry in the mental lexicon has to be accessed to retrieve its meaning (*lexical access*). At the sentence level, readers integrate word meanings semantically into a coherent mental representation of the sentence (*semantic integration*) according to the sentence’s syntactic structure (*syntactic integration*). Finally, at the text and discourse level, propositions must be integrated with each other and with prior knowledge into a coherent mental representation of the text by establishing mental relations of local and global coherence, by drawing knowledge-based inferences, and by monitoring the comprehension process.

Each of these cognitive component skills is assumed to make a unique and essential contribution to reading comprehension. That is, reading comprehension can only be successful to the extent that each of its component skills is successfully accomplished. Conversely, this also implies that reading comprehension is negatively affected when one or more of its component skills are deficient (e.g., Cain & Oakhill, 2004). Patterns of reading
deficits can thus be very heterogeneous and multi-facetted (Cain & Oakhill). In contrast to the common assumption that reading difficulties are primarily due to a phonological processing deficit (e.g., phonological-core variable-difference model by Stanovich, 1988), poor reading performance can be caused by several separate or combined deficits in cognitive component skills of reading comprehension at the word, sentence, and text level, which can vary in severity. The construction of diagnostic tests that identify individual sources of reading difficulties and of target-oriented training and intervention programs tailored to the individual needs of poor readers requires a comprehensive and profound knowledge of the cognitive component skill that constitute reading comprehension.

Although reading comprehension is assumed to depend on cognitive component skills at the word, sentence, and text level, several questions concerning the extent that these skills are interdependent, their relationship to corresponding skills in listening comprehension, and their development during reading acquisition remain unanswered. The aim of this dissertation is to advance a more comprehensive picture of reading comprehension by empirically exploring a selective subset of open research questions. The first empirical study explores the extent that phonological recoding and orthographical decoding skills contribute to reading comprehension at the sentence and text level and how both skills of visual word recognition develop in children learning to read in German. The second empirical study aims to test the simple view of reading (SVR; Gough & Tunmer, 1986; Hoover & Gough, 1990), one of the most prominent and influential theories in the field of reading research, by using optimized and methodologically stringent measures of the model components in German 3rd and 4th graders. The third empirical study explores how German primary school children and adults process positive-causal and negative-causal coherence relations in written and spoken language comprehension and how the processing of both coherence relations develops throughout primary school years. Finally, the fourth empirical study discusses and demonstrates the usefulness of diagnostic instruments assessing individual differences in cognitive component skills of reading and listening comprehension separately in a process-oriented manner.

The following sections in Chapter I provide a theoretical overview of the cognitive component skills of reading comprehension at the word, sentence, and text level. Previous research demonstrating the separability of the cognitive components skills and their unique contributions to reading comprehension is presented and discussed. Against this background, the aims and scope of this dissertation will be elaborated. At the end of the chapter, I present ProDi-L, a process-oriented reading assessment test for German 1st to 4th graders, which was
used to explore the research questions of this dissertation. Chapters II to V report the four empirical studies, which are the core of this dissertation. Chapter VI provides a short review of existing literature on reading disability (dyslexia) and discusses how poor readers can be characterized in terms of individual patterns of deficits in cognitive component skills of reading comprehension. Findings and practical implications of the four empirical studies reported in this dissertation are summarized and discussed in Chapter VII.
Component Skills of Reading Comprehension at the Word Level

Most theories in reading research assign visual word recognition a key role in reading comprehension. This seems fairly plausible, given that the decoding of written word forms is a necessary prerequisite to be able to process written text. One theory emphasizing the crucial role of visual word recognition in reading comprehension is Perfetti’s verbal efficiency hypothesis (1985), which states that efficient (i.e., reliable and rapid) word recognition is at the core of successful reading comprehension. Efficient word recognition saves cognitive resources at the word level so that more are available for higher cognitive comprehension processes at the sentence and text level. Conversely, if visual word recognition is inefficient, cognitive resources are bound to the word level and are consequently not available for comprehension processes at the sentence and text level, resulting in impeded reading comprehension. In a revised version of the verbal efficiency hypothesis, the lexical quality hypothesis, Perfetti and Hart (2001; 2002; see also Perfetti, 2010) assume that a high quality of word form representations (including phonological, orthographical, meaning and morphosyntactic information) and their effortless and efficient retrieval from memory are the basis for successful reading comprehension. Thus, reading comprehension crucially depends on the amount of high-quality lexical representations in the reader’s mental lexicon (Perfetti & Hart, 2001; Richter, Isberner, Naumann, & Neeb, 2013).

Another prominent theory focusing on the central role of visual word recognition in reading comprehension is the Simple View of Reading (SVR) advanced by Gough and Tunmer (1986) and Hoover and Gough (1990). The SVR states that reading comprehension (R) is the result of the product of two cognitive abilities, namely, decoding (D) and a more general linguistic comprehension (C; usually operationalized as listening comprehension):

\[ R = D \times C \]

According to the SVR formula, the only process that differentiates between reading and listening comprehension is the decoding of written word forms (see also Perfetti, 1999). After written word forms have been decoded, cognitive processes of comprehension are basically the same for reading and listening comprehension (but see also the discussion on differences between listening and reading comprehension in Kirby & Savage, 2008). The multiplicative combination of decoding and listening comprehension skills accounts for the observation that reading comprehension is impossible when one of the two components has a value of zero. Despite perfect decoding skills, reading comprehension is prevented when
listening comprehension is lacking. Conversely, when no decoding skills are evident, reading comprehension is impossible even when listening comprehension is perfect.

In sum, reading comprehension crucially depends on the successful accomplishment of visual word recognition processes (Perfetti, 1985; Shankweiler, 1989; Shankweiler et al., 1999). The following sections provide an overview of two cognitive component skills of visual word recognition—(1) phonological recoding and (2) orthographical decoding—and their substantial and unique contributions to reading comprehension.

**Visual Word Recognition – Phonological Recoding and Orthographical Decoding**

According to the *dual route cascaded model (DRC)* advanced by Coltheart et al. (2001; Coltheart, 2005), recognition of written word forms can be accomplished via a *non-lexical* route (phonological recoding) and a *lexical* route (orthographical decoding). Unknown or infrequent word forms are recognized via phonological recoding. The reader translates single graphemes or clusters of graphemes into their respective phonological representations by means of grapheme-to-phoneme correspondence rules of the particular language. These phonological representations of written word forms are processed in the same way as spoken word forms, that is, they are used to search the mental lexicon and to access matching lexical entries. Phonological recoding plays a central role for beginning readers in particular because this type of recoding is the only way to recognize written word forms that are unfamiliar to the reader (which applies to most written word forms in beginning readers). However, because of the laborious letter-by-letter translation, this route is cognitively demanding and time-consuming. Thus, more advanced readers make use of less demanding orthographical decoding. Advanced readers have already developed a *sight vocabulary* (Ehri, 2005b) of sufficient size that includes orthographical forms of familiar words with direct links to the respective entries in the mental lexicon. When encountering familiar written word forms, readers recognize them directly without the necessity of cognitively demanding grapheme-to-phoneme conversion. The DRC model assumes that during visual word recognition both routes start a “race” in parallel (Paap & Noel, 1991, p. 13). The more efficient route that recognizes the written word form faster and more reliably wins the race and accesses the entry in the mental lexicon.

Although models of visual word recognition have been advanced without the assumption of two functionally distinct routes, such as the *parallel-distributed-processing (PDP)* model by Seidenberg and McClelland (1989) or the *connectionist triangle model* by Plaut, McClelland, Seidenberg, and Patterson (1996), the dual-route models of visual word
recognition offer an advantage, because they account for several phenomena of visual word recognition in good and poor readers, and they explain developmental changes in visual word recognition throughout reading acquisition in a fairly simple and intuitive way. Several simulation studies and experimental investigations with skilled and poor readers have provided evidence in favor of dual route models of visual word recognition. Some of these studies are reviewed in the following section. One exception for the applicability of the dual route models of visual word recognition might be that they are not suitable to account for reading phenomena and acquisition in non-alphabetic scripts such as Chinese and Japanese (e.g., Frost, 2012). In contrast, they appear to be highly suitable for alphabetical scripts. Thus, the work presented in this dissertation is based upon the assumption of two functionally distinct routes of visual word recognition.

Findings Supporting the Assumption of two Functionally Distinct Routes of Visual Word Recognition

Coltheart, Curtis, Atkins, and Haller (1993) reviewed several studies on visual word recognition phenomena in skilled and poor readers. They found that dual-route models account for exception and non-word reading performance, for lexical decision task outcomes, and for several forms of acquired and developmental dyslexia. In contrast, the PDP model simulated exception word reading sufficiently but failed to simulate lexical decision outcomes and patterns of dyslexia. Similar findings were obtained by Ziegler, Perry, and Coltheart (2000) who demonstrated that the DRC model sufficiently simulated reading of regular words and irregularly spelled loan words in German.

In a word naming experiment, Paap and Noel (1991) tested the assumption of the DRC model that both routes start in parallel and compete during visual word recognition. They replicated the so-called exception word effect, finding that naming latencies for low-frequency exception words (irregular spelling) were significantly longer than naming latencies for low-frequency words with regular spelling. This effect can be explained by competition between both routes when exception words are named. Given that orthographical decoding is inefficient in low-frequency words and phonological recoding is inefficient in words with irregular spelling, both routes are equally inefficient and competitive in exception words. In contrast, phonological recoding is more efficient than orthographical decoding in low-frequency words with regular spelling. For these words, the routes compete less. To further strengthen the assumption of competition between both routes, they manipulated the competition by increasing working memory load with a secondary digit-memory task. The
cognitively demanding route of phonological recoding was expected to slow down with an increase in memory load. However, the authors had not expected the efficiency of orthographical decoding to be affected by increased memory load, because this route is highly automatized (see also Ehri, 2005b) and requires little cognitive resources. Contrary to intuition but consistent with the predictions based on the DRC model, naming speed for low-frequency exception words increased because of reduced competition between both routes—orthographical decoding was superior to impeded phonological recoding. Furthermore, naming speed decreased for low-frequency regular words because of the phonological recoding route deceleration, indicating an increase in competition between both routes. These findings of Paap and Noel support the assumption of two functionally distinct and competing routes of visual word recognition.

Further support comes from studies with poor readers. Castles and Coltheart (1993) obtained a pattern of double dissociation for 34% of the poor readers in their study. Poor readers who exhibited difficulties reading unfamiliar and non-words had difficulties with phonological recoding despite intact orthographical decoding (so-called *deep dyslexics*), whereas the so-called *surface dyslexics* suffered from deficient orthographical decoding but not phonological recoding. The latter had difficulties reading words in an automatized manner and overgeneralized irregularly spelled word forms (see also Manis, Seidenberg, Doi, McBride-Chang, & Petersen, 1996). The fact that reading of nonwords and irregular words can be strongly impaired independent of each other suggests two functionally distinct routes of visual word recognition.

**Unique Contributions of Both Routes to Reading Comprehension**

Extant research on poor readers in particular revealed a strong relationship between skills of visual word recognition and reading comprehension (e.g., Shankweiler et al., 1999). Poor reading skills are commonly assumed to stem from a general phonological deficit that impedes sufficient acquisition of phonological recoding skills as proposed by the phonological-core variable-difference model (Stanovich, 1988; see also Shankweiler, 1989). General phonological deficits can be caused by deficient phonological representations (e.g., Boada & Pennington, 2006) or deficient phonological retrieval processes (e.g., Dickie, Ota, & Clark, 2013; Griffith & Snowling, 2001; Ramus, Marshall, Rosen, & van der Lely, 2013). As a result, these readers have difficulties acquiring grapheme-to-phoneme conversion rules (Snowling, 1980) or their phonological recoding is very slow (Wimmer, 1996; Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Körne, 2003; see also Mayringer & Wimmer, 2000).
Longitudinal and training studies strengthen a causal link between phonological skills and reading ability (e.g., Rack, Snowling, & Olson, 1992; Vellutino, Fletcher, Snowling, & Scanlon, 2004; for a detailed and critical discussion on a causal relationship between phonological skills and reading comprehension, see Castles & Coltheart, 2004 and Hulme, Snowling, Caravolas, & Carroll, 2005). For example, reading ability in primary school could be predicted by phonological processing skills in kindergarten (e.g., Bradley & Bryant, 1983; Scanlon & Vellutino, 1996; see also Wilson & Lesaux, 2001 and Ransby & Swanson, 2003 for the persistence of poor phonological processing skills into adulthood), and enhancing phonological skills in preschoolers and beginning readers was found to improve reading comprehension skills in higher grade levels (e.g., Bradley & Bryant, 1983; Scanlon & Vellutino, 1996).

Another likely source of poor reading comprehension is poor orthographical decoding skills. Castles and Coltheart (1993) revealed not only the unique contribution of phonological recoding but also of orthographical decoding to reading comprehension. In their study of 8- to 14-year-old poor readers, 45 of 53 readers showed a pattern of dissociation between the two routes. Twenty-nine of the 45 readers exhibited more difficulties with phonological recoding than would have been expected from their orthographical decoding skills, and the remaining 16 readers exhibited more difficulties with orthographical decoding than would have been expected from their phonological recoding skills. Overall, 18 readers even showed a complete dissociation, that is, they exhibited a deficit in one of the two routes whereas the other route was completely unaffected. The findings by Castles and Coltheart suggest unique roles of both visual word recognition skills for reading comprehension. When visual word recognition is poor, either as a result of deficient phonological recoding or deficient orthographical decoding, reading comprehension is negatively affected (see also Manis et al., 1996).

Finally, a cross-sectional study by Richter et al. (2013) demonstrated that phonological recoding and orthographical decoding skills explained unique variance in the reading comprehension of 247 German primary school children in Grades 1 to 4. Moreover, an indirect effect of both visual word recognition skills was found on reading comprehension, mediated by lexical access to word meanings (for further discussion see the section **Access to word meanings**).

In sum, these studies strongly suggest that successful reading comprehension crucially depends on the successful accomplishment of phonological recoding and orthographical decoding as proposed by dual route models of visual word recognition. Unknown or infrequent regular words are recognized via phonological recoding, and familiar and irregular
words, which are already stored in sight vocabulary, are recognized via orthographical decoding. Alternative models such as the strong phonological model by Frost (1998) even assume that phonological information is always activated during visual word recognition independent of the reader’s familiarity with the word. However, the idea that advanced readers make use of phonological recoding at least to some extent (e.g., when encountering unfamiliar word forms) is in contrast to the three-stage developmental model by Frith (1985, 1986), which states that readers acquire and use phonological recoding and orthographical decoding in a strictly serial order instead of using both routes simultaneously. The following section discusses the roles of phonological recoding and orthographical decoding in developing readers with respect to the DRC model and the three-stage developmental model.

**Different Roles of Phonological Recoding and Orthographical Decoding Throughout Reading Development**

The DCR model implies that the relative contributions of phonological recoding and orthographical decoding in reading comprehension change throughout the course of reading development. Given that most written word forms are unfamiliar to beginning readers, beginning readers should recognize them primarily via phonological recoding. In contrast, more advanced readers, who have stored most written word forms in sight vocabulary (Ehri, 2005b), should recognize written word forms primarily via the cognitively less demanding and more efficient route of orthographical decoding (Hoover & Gough, 1990; Tunmer & Chapman, 2012). Thus, the relative contribution of phonological recoding to reading comprehension should decrease with increasing reading experience, whereas the relative contribution of orthographical decoding to reading comprehension should increase.

The idea of a shift from phonological recoding to orthographical decoding during reading acquisition is a central assumption in hierarchical stage or phase models of literacy development, such as the foundation literacy framework (Seymor, 1997; Seymour, Aro, & Erskine, 2003) or Frith’s three-stage developmental model (1985, 1986). Whereas the phases in the foundation literacy framework are not necessarily acquired consecutively, the three-stage developmental model assumes that developing readers progress through the following three critical developmental stages of visual word recognition sequentially: logographic, alphabetic, and orthographic. Moreover, the use of the three stages is not under the strategic control of the reader. Prior to actual reading, illiterate subjects can recognize some familiar written words in an iconic fashion. At this initial logographic stage, they are not yet aware
that words consist of separate graphemes, which can be mapped onto phonemes. Instead, they store familiar words like pictures. With beginning reading instruction, readers proceed to the alphabetic stage. At this stage, readers learn to recognize single graphemes and to map them onto phonemes by using the grapheme-to-phoneme-conversion rules of their language. Word recognition at this stage corresponds to phonological recoding in the DRC model. Finally, when readers have fully mastered the alphabetic stage, they proceed to the orthographic stage. At this stage, word forms or orthographic units such as morphemes are recognized directly without the necessity of grapheme-to-phoneme conversion. Word recognition at the orthographic stage corresponds to orthographical decoding as proposed by the DRC model.

Although the DRC model and the three-stage developmental model share the idea of a shift from phonological recoding (alphabetic stage) to orthographical decoding (orthographic stage) during reading acquisition, some fundamental differences exist between the two models. The DRC model states that even advanced readers use phonological recoding to some extent, whereas the three-stage developmental model states that the readers progress through critical developmental stages sequentially. That is, after the alphabetic stage has been mastered and the reader has proceeded to the orthographic stage, written words are no longer recognized via grapheme-to-phoneme conversion. According to Frith, there is a short period of overlap when both stages are “merged” (Frith, 1985, p. 309). That is, grapheme-to-phoneme-conversion is still retained to help the reader grasp the functioning of the new stage. But eventually, readers fully shift to the orthographic stage in which written words or orthographic units are recognized directly.

The three-stage developmental model states further that a shift to the next stage cannot occur before the preceding stage has been completely mastered. For example, failing to master the alphabetic stage precludes a shift to the orthographic stage. Consequently, visual word recognition cannot develop further without specific training, and reading comprehension is hampered (Frith, 1985, 1986). Findings by Zoccolotti et al. (2005) supported the assumption that poor reading skills are caused by a failure to shift from the alphabetic to the orthographic stage. In their study, poor and skilled beginning readers named written words of differing length. The authors expected word length to have an effect on naming latencies only at the alphabetic stage, because at this stage written word forms are decoded letter-by-letter. In contrast, word length was not expected to affect naming latencies at the orthographic stage, because at this stage word forms are recognized directly. Zoccolotti and colleagues found that skilled readers’ naming latencies were less affected by word length at Grade 2 than at Grade 1, indicating a shift from the alphabetic to the orthographic stage. However, poor readers
naming latencies were affected by word length in both grades, indicating that poor readers had become stuck at the alphabetic stage.

In contrast to Frith’s (1985, 1986) assumption that readers cannot proceed to the orthographic stage before they have fully mastered the alphabetic stage, the DRC model assumes that acquiring orthographical decoding skills is possible even when phonological recoding skills are poor (Castles & Coltheart, 1993; Coltheart et al., 2001). Thus, the finding that some poor readers appear to have good orthographical decoding skills despite poor phonological recoding skills (e.g., Castles & Coltheart, 1993; Manis et al., 1996) can be explained by the DRC model.

In sum, the empirical literature reviewed strongly suggests that phonological recoding and orthographical decoding contribute substantially and uniquely to reading comprehension in beginning and advanced readers. However, the relative contributions of both visual word recognition skills on reading comprehension can be assumed to change to some extent as reading skills increase—phonological recoding can be considered to play a more prominent role in beginning readers, whereas advanced readers rely increasingly on orthographical decoding with increasing sight vocabulary.

Access to Word Meanings

In addition to visual word recognition, comprehending written language at the word level involves the successful retrieval of word meanings from the mental lexicon. Retrieval of meaning information is a crucial component of Perfetti and Hart’s lexical quality hypothesis (2001, 2002; Perfetti, 2007). According to the lexical quality hypothesis, reading comprehension crucially depends on readers’ word knowledge, that is, on the quality of their lexical representations. Lexical representation quality depends on the quality of the following four representational word features (or constituents) and the strength of their interconnectedness: phonological, orthographical, meaning, and morpho-syntactic constituents (Perfetti & Hart, 2001, 2002 subsume the meaning and morpho-syntactic constituents under a single ‘semantic’ constituent). If one or more of these constituents are low in quality, reading comprehension can be considerably impeded.

In a semantic verification judgment task, Perfetti and Hart (2001) found that poor reading skills in adults were associated with lexical representations of low quality. Readers were presented with written word pairs that could be semantically related such as king – royalty, not related such as evening – royalty, or they contained a homophone such as night – royalty, with the non-intended homophone meaning (knight) being semantically related to the
second word. Readers were asked to judge whether the words of a word pair were semantically related or not. Perfetti and Hart (2001) reported two central findings. First, skilled readers responded faster than less-skilled readers to all word pairs. The authors inferred that skilled readers have high quality lexical representations and therefore are able to access word meanings more efficiently than less-skilled readers. Second, skilled readers exhibited an early homophone interference effect (i.e., longer response latencies for word pairs containing homophones) at an SOA of 150ms, whereas the same effect was observed at a longer SOA of 450ms for less-skilled readers. Furthermore, this effect differed for high- and low-frequency homophones. Skilled readers exhibited an early interference effect (SOA: 150ms) for word pairs with low-frequency homophones but no interference effect for high-frequency homophones. In contrast, less-skilled readers exhibited a late interference effect (SOA: 450ms) for word pairs containing high-frequency homophones but no interference effect for low-frequency homophones. Again, the authors argued that this pattern of findings can be explained by differences in lexical representation quality between skilled and less-skilled readers. They reasoned that the lexical representation quality of high-frequency homophones in skilled readers is high, meaning that the word constituents are tightly bound and are retrieved as a unit. Consequently, processing high-frequency homophones is not prone to interference with the non-intended meaning in skilled readers. In contrast, retrieval of low-frequency homophone constituents can be affected by the non-intended meaning of the homophone, because they are less tightly bound. In less-skilled readers, high-frequency homophones are of lower lexical quality than in skilled readers (i.e., the word constituents are not that tightly bound). As a result, the processing is prone to interference with the non-intended meaning. Low-frequency homophones, however, are of very low quality in less-skilled readers. General processing difficulties for these words might conceal homophone interference effects in less-skilled readers (Perfetti & Hart, 2001; see also Perfetti, 2007 for a detailed interpretation and discussion of these findings). These findings indicate that individual differences in reading ability are related to individual differences in the quality of lexical representations.

The lexical quality hypothesis was also tested by Richter et al. (2013). In addition to the significant contributions of phonological recoding and orthographical decoding to reading comprehension, the authors found that the effects of both visual word recognition skills were mediated by access to word meanings. These findings support Perfetti and Hart’s (2001, 2002) assumption that individual differences in reading comprehension can be attributed to individual differences in the quality of lexical representations, that is, in the quality of
phonological, orthographical, meaning, and morpho-syntactic representations and the ease of their access.

Component Skills of Reading Comprehension at the Sentence Level

The majority of models and studies on reading comprehension and development have focused on word level skills to explain individual differences in reading ability. However, several studies show that poor reading ability can emerge despite adequate word-level skills (e.g., Cain, Oakhill, Barnes, & Bryant, 2001; Nation, Cocksey, Taylor, & Bishop, 2010; Nation & Snowling, 1997, Study 2, 1998, Study 2, 1999; Stothard & Hulme, 1992). These studies suggest that further cognitive component skills contribute to reading comprehension beyond word level skills (Cain, 2009; Cain & Oakhill, 2004). This assumption is consistent with the SVR, which states that reading comprehension is the product of decoding skills and more general linguistic (or listening) comprehension skills. The simple view implies that reading difficulties must be attributed to poor listening comprehension skills if decoding skills are highly developed (Gough & Tunmer, 1986; Hoover & Gough, 1990; for empirical support of this assumption see Catts, Hogan, & Fey, 2003; Johnston & Kirby 2006; Joshi & Aaron, 2000; Kendeou, Savage, & van den Broek, 2009; Nation & Snowling, 1997, Study 2; Ransby & Swanson, 2003).

General linguistic (or listening) comprehension comprises several cognitive component skills at the sentence and text level (e.g., Hogan, Sittner Bridges, Justice, & Cain, 2011; Cain & Oakhill, 2004; Nation, 2005). To understand a sentence’s intended meaning it is not sufficient to simply derive its word meanings and to combine them. Moreover, the reader must consider the syntactic structure of the sentence and, if available, context information to integrate the word meanings into a coherent mental representation of the sentence (e.g., Lenhard & Artelt, 2009; Müller & Richter, 2014; Richter & Christman, 2009). For example, in the sentence Katie hits John, understanding who hit whom is impossible based on word meanings alone, because both referents are possible hitters and possible victims. Only word order (syntactic information) provides the reader with the information that Katie is the hitter and John is the one who was hit, because the initial noun phrase is always the subject and agent of the sentence in English transitive main clauses with active voice, and the final noun phrase is the direct object and patient. Interpreting figurative meaning is another example of the necessity to derive sentence meanings not only on the basis of word-meaning combinations alone. The idiom to hit the roof implies that someone is really angry. In this
• Introduction and Aims •

In case, the combination of literal word meanings would result in the miscomprehension that someone actually hits the ceiling of the room or the roof of the house. The reader must process the whole word string in a specific order to recognize the idiom. The reader then relies on semantic context information to interpret the sentence either literally or figuratively. If the reader, for example, reads a story describing someone falling onto a roof, the figurative meaning clearly would make no sense. However, if the reader reads a story about a couple fighting, the figurative meaning makes more sense than the literal meaning. The semantic context can also be useful to solve ambiguities. A bug means something entirely different when the sentence reads the bug was killed by the man in the gray suit (bug: insect) or the bug was carefully placed by the man in the gray suit (bug: device for spies). Thus, to ensure successful sentence comprehension, readers must also process syntactic and semantic context information in addition to information derived at the word level.

Skills of Syntactic Integration

Several studies have demonstrated a strong relationship between individual differences in auditory and visual skills of syntactic integration, that is, a reader’s “ability to reflect upon and to manipulate aspects of the internal grammatical structure of sentences” (Tunmer, Nesdale, & Wright, 1987, p. 25) and individual differences in reading comprehension. Stothard and Hulme (1992) found that the performance of 7- to 8-year-old poor readers on Bishop’s test for the reception of grammar (TROG, 1983) was inferior to the performance of skilled readers matched for chronological age, reading accuracy, and vocabulary age. Grammaticality judgments and corrections were shown to be positively related to reading skills at the end of first grade, even when phonological awareness, naming speed, and auditory memory were statistically controlled (Plaza & Cohen, 2003). Poor readers exhibited difficulties correcting ungrammatical spoken sentences, supplying missing words in an oral cloze task (Tunmer et al., 1987), or unscrambling words back into their correct order in a sentence (Nation & Snowling, 2000). Using a picture-choice task, Byrne (1981) demonstrated that poor readers were less likely than skilled readers to identify a picture matching a syntactically complex spoken sentence when the picture could be correctly identified only with the aid of syntactic information. Nation et al. (2010) investigated the impact of syntactic integration skills on reading comprehension in a longitudinal study. They found that the ability to recall sentences of varying length and syntactic complexity and to apply syntactic rules in a picture-choice task in 6- to 7-year-olds was predictive of their comprehension skills at 8 years. Furthermore, poor comprehenders performed worse than good comprehenders on
Introduction and Aims

the Test for Reception of Grammar-2 (TROG-2, Bishop, 2003) at 5 years, but their performance on phonological measures, such as non-word repetition and phoneme elision, did not differ from good comprehenders during the three-year span. These findings can be interpreted in favor of a causal link between individual differences in syntactic integration skills and individual differences in reading comprehension.

Skills of Semantic Integration

In addition to syntactic integration skills, semantic integration skills make a unique contribution to reading comprehension. Readers must integrate the word meanings contained in a sentence in a meaningful way to establish a coherent mental representation of the sentence content. This mental representation can be further enriched by contextual and world knowledge to solve local or global ambiguities or to recognize figurative meanings of word strings such as metaphors or idioms. If semantic integration skills are poorly developed, reading comprehension can be negatively affected. Hannon and Daneman (2004) demonstrated that poor readers were less likely than skilled readers to detect semantic anomalies within a sentence such as Amanda was bouncing all over because of too many sedatives (p. 197), indicating shallower semantic processing in poor readers compared to skilled readers. The authors assumed that poor readers put less cognitive effort in the establishment of coherent sentence meanings compared to skilled readers. Instead, poor readers appear to invest their cognitive resources in the establishment of global text coherence. However, semantic integration of sentence contents in this study required the establishment of a local coherence relation between adjacent intrasentential propositions (e.g., between Amanda was bouncing all over and she took too many sedatives). Consequently, processes of semantic integration and establishing coherence relations are not possible to disentangle in this study (for a discussion of Hannon & Daneman’s findings in terms of establishing coherence relations, see the section Establishing Coherence in this Chapter).

To the same extent that poor semantic integration skills can be associated with reading difficulties, good semantic integration skills can be associated with good reading comprehension. Well-developed semantic integration skills can even help readers to compensate at least partially for poorly developed cognitive component skills at the word level (e.g., see the interactive-compensatory model by Stanovich, 1980). To a certain degree, poor decoders can rely on the semantic context to anticipate and infer words that are difficult to recognize. Although using contextual information to infer words is error prone and certainly cannot ‘repair’ deficient word recognition, the strategy can, at least, support visual
Introducing and Aims

Several studies investigating the use of contextual information during visual word recognition in readers with good and poor decoding skills suggest that contextual information supports visual word recognition for poor decoders in particular. West and Stanovich (1978) found that a sentence context that was congruent with a target word facilitated the recognition of the target word in 4th graders, 6th graders, and college students compared to an incongruent context or to context-free word recognition. Moreover, naming latencies for words were reduced more strongly by a facilitating context for poor readers than for skilled readers. West and Stanovich concluded that readers with poor word recognition skills make use of semantic context information to compensate for a lack of automaticity in visual word recognition skills. Similar results were obtained by Nation and Snowling (1998, Exp. 2) who found greater relative contextual facilitation in a visual word recognition task for 7- to 10-year-old readers with poor visual word recognition skills compared to normal readers and readers with poor comprehension skills.

However, processing semantic context information can also have a hindering effect on visual word recognition in poor readers when inappropriate meanings are required to be suppressed. Gernsbacher, Varner, and Faust (1990, Exp. 4) presented poor and skilled readers with written sentences ending in an ambiguous or unambiguous word. After 100ms (immediate condition) or 850ms (delayed condition) participants were presented with a target word. Their task was to judge whether or not the target word matched the sentence meaning. In the immediate condition, both skilled and poor readers exhibited an interference effect when the target word did not match the sentence meaning but the non-intended meaning of the ambiguous word. That is, both groups needed time to suppress the inappropriate meaning of the ambiguous word in the immediate condition. However, in the delayed condition, the interference effect was found only for poor readers, which suggests that they are not able to suppress the inappropriate meanings of ambiguous words even after a delay of 850ms. Similar findings were obtained by Gernsbacher and Faust (1991, Exp. 1) for sentences ending in homophones. Using a similar paradigm, they found that poor readers had more difficulties than skilled readers in suppressing the non-intended meanings of homophones. The findings by Gernsbacher et al. and Gernsbacher and Faust suggest that individual differences in the inability to suppress information, which is inappropriate within the context of a specific sentence, can be related to individual differences in text comprehension (but see also Perfetti & Hart, 2001 who found no homophone interference effect for poor readers when homophones were presented in the context of a sentence).
In this dissertation, ample evidence has been presented showing that individual differences in reading comprehension are related to individual differences in cognitive component skills at the word and sentence level. However, to fully comprehend text, readers must accomplish various text level processes that link the contents of its sentences and paragraphs in a meaningful way, as in the text below.

(1.1) *Paul opened his apartment door and entered the living room. Instantly his eye was caught by the tattered curtains. He swung around furiously looking for Hercules. After a few minutes he noticed the red furred tail behind the couch.*

Readers are required to identify *Paul* as the referent of the personal pronouns *he* and *his* and establish a causal relationship between the *tattered curtains* (cause) and Paul’s anger (consequence). Readers must also infer that Paul is convinced that *Hercules* is responsible for the mess in his living room. Based on the information provided by the text (*red furred tail, behind the couch*) and based on readers’ world knowledge (cats are able to squeeze behind a couch and they love to tatter curtains, whereas the mythical figure *Hercules* probably would not) they must infer that *Hercules* does not refer to the mythical figure but to a cat called *Hercules* etc.

Text comprehension can be described as the construction of a coherent *mental model* (Johnson-Laird, 1981) or a *situation model* (Van Dijk & Kintsch, 1983) of the text (both terms are used interchangeably in this dissertation). In addition to mentally representing information that is directly conveyed by the text (*propositional textbase*), a mental model includes prior text-based and more general world knowledge of the reader that relates to the situation described in the text. In this vein, mental models are richer and more elaborated than a mental representation of the textbase alone. Thus, to establish a coherent mental model of the text, readers must integrate the propositional textbase with their prior knowledge.

The establishment of a mental model involves several cognitive activities such as mentally linking the contents of adjacent propositions (establishing local coherence relations) and distant propositions (establishing global coherence relations; Van Dijk & Kintsch, 1983), drawing knowledge-based inferences (e.g., Graesser, Singer, & Trabasso, 1994; Oakhill, Cain, & Bryant, 2003), monitoring the plausibility of the text (Isberner & Richter, 2013), and monitoring the comprehension process itself (Nation 2005; for a short overview of text level
comprehension processes see Cain, 2009 and Hogan et al., 2011). Previous research provides evidence that individual differences in drawing inferences, monitoring comprehension processes, or establishing coherence relations are associated with individual differences in reading comprehension.

**Drawing Inferences and Comprehension Monitoring**

In cross-sectional studies, Cain and Oakhill (1999) and Cain et al. (2001) found that 7- to 8-year-old poor readers exhibited more difficulties than their skilled peers drawing inferences that connect information provided by the text (coherence/text-connecting inferences) and drawing more elaborative inferences by integrating text-based information and world knowledge to fill informational gaps in the text or to elaborate the mental model (elaborative/gap-filling inferences; note that, strictly speaking, coherence/text-connecting inferences are coherence relations, because they mentally link propositions of a text—see the section Establishing Coherence in this Chapter). Children were presented with short texts and answered comprehension questions that required the drawing of text-connecting and gap-filling inferences. Poor readers were found to draw fewer inferences of both types than their skilled peers even when visual word recognition skills, background knowledge, and memory of relevant text-based information were statistically controlled. They appear to have difficulties selecting relevant information from the text and to integrate this information with their general knowledge to draw inferences (Cain et al., 2001).

Cain and Oakhill (1999) emphasized that inference making is essential for successful text comprehension rather than a mere by-product of the comprehension process. They found that younger skilled readers who were matched to older poor readers based on their comprehension abilities performed better on drawing inferences than their comprehension-age matches. This finding can be interpreted in favor of a causal link between the inability to draw inferences and poor reading comprehension. Longitudinal studies by Cain et al. (2004) and Oakhill et al. (2003) further supported the assumption of a causal link between inference drawing and reading comprehension. The ability of young readers to draw inferences predicted their reading comprehension at three time points during the ages of 7 to 8, 9 to 10, and 10 to 11 years even after controlling for word reading skills and verbal ability measures. Furthermore, Cain et al. (2004) ruled out the possibility that limited working memory capacity accounted for their results. Measures of both working memory and inference making explained a unique amount of variance in reading comprehension.
In the same study, Cain et al. (2004) obtained comparable results for measures of comprehension monitoring, which assess “a reader’s ability to detect inconsistencies in text, such as scrambled sentences, contradictory sentences, or statements that conflict with external information (world knowledge)” (p. 33). They found that poor readers were less able to detect conflicting information in a text compared to skilled readers. As with drawing inferences, individual differences in comprehension monitoring explained unique variance in children’s reading comprehension. The differential effect of monitoring on reading comprehension was further supported by Van der Schoot, Vasbinder, Horsley, Reijntjes, and van Lieshout (2009). In their study, poor readers were less likely than skilled readers to reinterpret an ambiguous word that was initially misinterpreted. In contrast to skilled readers, poor readers showed no reading time increase on disambiguating information and provided more incorrect responses to comprehension questions, indicating that they had not noticed their interpretation error. This finding suggests that poor readers are less successful in monitoring their comprehension process.

Establishing Coherence

Establishing coherence in a text is another essential cognitive ability for the construction of a coherent mental model of the text. The process requires the reader to form mental links between adjacent propositions (local coherence) and distant propositions (global coherence) in the text. Studies from the last decade demonstrated that individual differences in reading comprehension are related to individual differences in skills of establishing coherence relations. Hannon and Daneman (2004) found that less-skilled readers established less mental links of local coherence between two intrasentential propositions compared to skilled readers. When presented with incoherent propositions such as *Amanda was bouncing all over because of too many sedatives* (p. 197), poor readers noticed the anomaly less often than skilled readers. In contrast, poor readers in a study by Long and Chong (2001, Exp. 1) exhibited particular difficulties with the establishment of global instead of local coherence. They were less likely to recognize inconsistencies when critical and inconsistent information were separated by several filler sentences (violation of global coherence) compared to texts in which critical and inconsistent information were separated by just one filler sentence (violation of local coherence). To rule out that the failure to detect violations of global coherence was not due to a mere inability to reactivate distant text information, they conducted a second experiment. Using a probe-verification paradigm, they showed that critical information could be reactivated easily at various points throughout the text. This
finding suggests that the integration and not the reactivation of distant information poses difficulties for poor readers.

A possible source for difficulties in establishing coherence relations might be an inability to recognize or interpret linguistic markers of coherence such as captions, connectives, and referential markers that explicitly signal relationships between propositions. Bridge and Winograd (1982) administered a cloze task to 9th graders who read short text passages in which 15 cohesive devices were omitted. The readers’ task was to supply the omitted referential markers (such as personal pronouns), lexical markers (such as repetitions, synonyms, or superordinates of referents), or connectives (e.g., then, and, or so). They found that poor readers were less likely to supply the intended cohesive markers, and they had more difficulties in detecting signals in the text that could help identify the missing words.

Comprehending Connectives

Connectives such as therefore, and, or nevertheless are specific type of cohesive marker signaling the kind of local coherence relation that needs to be established between adjacent propositions. Texts passages can include a wide variety of coherence relations (for a taxonomy, see Sanders, Spooren, & Nordmann, 1992) and connectives signaling them. Although coherence relations need not be signaled by connectives explicitly, connectives can function as ‘processing instructions’ (Kamalski, Sanders, & Lentz, 2008, p. 324) that facilitate text comprehension by indicating the particular link that must be established between propositions (e.g., Murray, 1995). Conversely, poor comprehension of connectives can hinder text comprehension, in particular when an appropriate coherence relation becomes difficult to infer without the aid of a connective. Geva and Ryan (1985) presented skilled (high reading level) and less-skilled (medium and low reading level) readers of Grades 5 and 7 with short expository texts in four conditions: without connectives (implicit condition), with connectives (explicit condition), with connectives highlighted by capital letters and underlining (highlighted condition), and with blank slots for which the readers were required to supply the missing connective by choosing one of three alternatives (deep condition). Following each text, readers answered ten true-or-false comprehension questions addressing coherence relations between propositions (structure questions) and ten additional questions concerning intrasentential details. Geva and Ryan found that less-skilled readers performed inferior to more skilled readers on rejecting false structure questions. They performed above chance level only in the highlighted condition. These findings suggest problems with the establishment of local coherence relations in less-skilled readers, even when coherence
relations were explicitly signaled by connectives. The exception was only when the readers’ attention was drawn to the connectives by highlighting. Finally, less-skilled readers performed worse than skilled readers in supplying the missing connectives in the deep condition, indicating poor knowledge of connective meanings. Consistent with the findings by Geva and Ryan (1985), Cain, Patson, and Andrews (2005, Study 2) found that poor 7- to 9-year-old comprehenders had significantly more difficulties supplying appropriate additive, causal, temporal, or adversative connectives in a cloze choice task compared to skilled comprehenders. Given that both groups were matched for word reading skills and vocabulary knowledge, group differences could not be attributed to differences in word level skills.

Yet poor readers might rely even more strongly on cohesive devices than skilled readers, because they lack the ability to infer relations of local and global coherence without the aid of explicit signals in the text. Hall et al. (2014) found that readers with low cognitive ability had more difficulties comprehending texts of low cohesion than comprehending texts of high cohesion, whereas readers with high cognitive ability performed well on both types of texts. A likely interpretation of this finding is that low reading comprehension is associated with difficulties in establishing coherence relations without the aid of explicit linguistic signals in the text.

**Reading Difficulties Characterized by Deficits in Cognitive Component Skills of Reading Comprehension**

In the preceding sections, I argued that reading comprehension depends on the successful accomplishment of several cognitive component skills at the word, sentence, and text level. Numerous studies have shown that each of these skills make a unique contribution to reading comprehension and that reading comprehension difficulties arise when one or more of these skills are deficient (e.g., Cain & Oakhill, 2004; Vellutino et al., 2004). This evidence bears important implications for the diagnosis of reading difficulties and the construction of intervention programs. To create target-oriented and successful reading interventions specially geared to improve deficient component skills, diagnostic tools are required to assess individual differences in cognitive component skills of reading comprehension selectively. However, despite extensive evidence in favor of distinct cognitive component skills of reading comprehension, the common approach is to identify poor readers on the basis of widespread but obsolete diagnostic guidelines and definitions, which do not address the structure and interdependencies of cognitive component skills of reading comprehension.
Instead, they simply classify readers into two groups: those with and those without a reading disability.

Poor readers are usually diagnosed with developmental dyslexia (specific reading disability) when they have below-age reading abilities and when their reading abilities are lower than would have been expected from their general cognitive development (American Psychiatric Association, 2013; World Health Organization, 2010). This so-called aptitude-achievement discrepancy has some serious weaknesses. First, it lacks the diagnostic capability of identifying the specific type of cognitive deficit that is causing the reading difficulties. As a result, categorizing a poor reader as dyslexic provides no indication as to the type and extent of required intervention (Coltheart & Jackson, 1998). Second, poor readers showing no aptitude-achievement discrepancy (i.e., their general cognitive abilities are as poor as their reading skills) are not diagnosed with dyslexia, because their poor reading abilities are assumed to stem from a more general cognitive deficit (Rutter & Yule, 1975; Yule, Rutter, Berger, & Thompson, 1974). As a result, these poor readers are often excluded from reading research and even worse from reading interventions (Catts et al., 2003; Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992). Several studies (e.g., Siegel, 1988; Stanovich & Siegel, 1994) demonstrated that dyslexics and “garden-variety poor readers” (Stanovich, 1988, p. 590) performed similarly on several reading-related tasks. Thus, treating them differently based on an arbitrary cut-off criterion seems baseless. A more promising approach is to characterize poor readers on the basis of their specific deficits in cognitive component skills of reading comprehension so that appropriate intervention programs can be implemented tailored to their individual needs (for further discussion see Chapter VI).

Assessing Individual Differences in Cognitive Component Skills of Reading Comprehension with ProDi-L

The empirical research reported in this dissertation is based on data that were collected in the context of a more comprehensive study on reading and listening comprehension in German 1st to 4th graders that was funded by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF, grants 01 GJ 0985, 01 GJ 0986, 01 GJ 1206A, 01 GJ 1206B). Individual differences in cognitive component skills of reading comprehension were assessed via ProDi-L (Prozessbezogene Diagnostik des Leseverstehens bei Grundschulkindern [Process-based assessment of reading skills in primary school children]; Richter, Naumann, Isberner, Neeb, & Knoepke, in press), a German-speaking test
battery for 1st to 4th graders. ProDi-L is a computerized process-oriented reading test that is characterized by six well-defined reading tasks assessing the skills of (1) phonological recoding, (2) orthographical decoding, (3) access to word meanings, (4) syntactic integration, (5) semantic integration, and (6) establishing local coherence. ProDi-H (Prozessbezogene Diagnostik des Hörverstehens bei Grundschulkindern [Process-based assessment of listening skills in primary school children]), the auditory equivalent to ProDi-L, was used to assess individual differences in corresponding listening comprehension skills. The test items are different for both tests but parallel with respect to construction criteria and item features. The tasks include items of varying difficulty to allow measurement of individual differences in cognitive component skills of reading and listening comprehension. Item difficulties are varied by careful modification of item features, which are known to facilitate or impede the addressed cognitive processes according to psycholinguistic research. To minimize the influence of cognitive processes nonspecific to reading and listening comprehension, all tasks require processing of only written and spoken language stimuli. Responses are provided by pressing one of two response buttons for yes and no responses. Response accuracy and latency are recorded for each item (Richter, Isberner, Naumann, & Kutzner, 2012; Richter, Naumann, Isberner, & Kutzner, 2011).

ProDi-L and ProDi-H follow a process-oriented approach to reading and listening assessment, hence they are suitable to assess individual differences in cognitive component skills of reading and listening comprehension and to investigate their interrelations and development throughout primary school years. In addition, the parallel tasks and materials of ProDi-L and ProDi-H allow for the investigation of commonalities and differences between reading and listening comprehension skills and for interrelations between them.

**Advantages of ProDi-L Over Extant Psychological Reading Assessment Tests**

A wide variety of psychological tests are currently available for assessing individual differences in reading comprehension in German. However, the majority of these tests are subject to several limitations and shortcomings, and they lack requirements for a sufficient, comprehensive, and differentiated diagnosis of possible reading deficits. First, most of the available reading tests barely differentiate between cognitive component skills involved in reading comprehension. Instead, they focus on the assessment of visual word recognition skills (e.g., SLRT, Landerl, Wimmer, & Moser, 2006, 2nd Ed.; SLRT-II, Moll & Landerl, 2010; WLLP-R, Schneider, Blanke, Faust, & Küspert, 2011) or they assess text
comprehension by means of global, product-oriented measures (e.g., FLVT 5-6, Souvignier, Trenk-Hinterberger, Adam-Schwebe, & Gold, 2008; HLP 1-4, May & Arntzen, 2000; LGVT 6-12, Schneider, Schlagmüller, & Ennemoser, 2007; SLS 2–9, Wimmer & Mayringer, 2014; VSL, Walter, 2013), which assess the quality of mental representations resulting from text processing rather than assessing the quality of the processes involved in reading.

Second, many of the commonly used reading tests involve tasks that require reading-unrelated cognitive skills such as picture recognition and selection (e.g., ELF E 1-6, Lenhard & Schneider, 2006; HAMLET 3-4, Lehmann, Peek, & Poerschke, 1997; WLLP-R) or choosing the correct answer from several alternatives that are required to be kept active in short term memory in a multiple-choice task (e.g., ELFE 1-6; FLVT 5-6; HAMLET 3-4; LESEN 6-7, Bäuerlein, Lenhard, & Schneider, 2012a; LESEN 8-9, Bäuerlein, Lenhard, & Schneider, 2012b; LGVT 6-12; VSL). As a consequence, measures of reading ability might be distorted at least to a certain degree by individual differences in cognitive skills that are unrelated to reading.

Finally, current reading tests often use the number of accurately solved test items as an indicator of reading comprehension (e.g., the number of correctly answered comprehension questions, or the number of words correctly read aloud), but this approach neglects the diagnostic value of reading latencies (e.g., HAMLET 3-4; HLP 1-4; VSL). Reading latencies are indicative of the degree of automatization of cognitive processes, whereas accuracy is indicative of their reliability. Only processes that are both reliable and automatized are considered to be efficient (Perfetti, 1985). Few reading tests (e.g., ELFE 1-6) incorporate measures of reading latency unlike the tests that merely focus on the fluency of single-word reading (see so-called speed tests, e.g., IEL-1, Diehl & Hartke, 2012; LDL, Walter, 2010; SLRT-II; one minute of reading, Deno, 1985; Deno, Fuchs, Marston, & Shin, 2001).

In sum, a reliable and differentiated diagnosis of reading deficits requires a test that assesses individual differences in cognitive component skills of reading comprehension separately in a process-oriented fashion. The tasks should be defined to address particular cognitive component skills without assessing cognitive skills unrelated to reading. Finally, both accuracy and reading latencies should be used as measures of efficiency. These requirements are met by ProDi-L. The six subtests of ProDi-L are briefly summarized in the following section.
The Six Subtests of ProDi-L

The six subtests of ProDi-L are embedded in a story of an extraterrestrial named Reli (Figure 1.1) who came to earth to learn the earthling’s language. The children’s task is to tell whether he did something wrong. Reli introduces the tasks in short animated video clips. Each task consists of several test items of varying item difficulty. A practice phase consisting of two items precedes each test phase. In the practice phase, children receive feedback from Reli. If they provide an incorrect answer, both practice items are repeated until they are answered correctly.

![Figure 1.1. The extraterrestrial Reli walks the children through the tasks.](image)

**Phonological Comparison Task**

The efficiency of phonological recoding skills is assessed via a *phonological comparison task*. Children are presented with a spoken non-word (e.g., *zelifo*). Subsequently, a written non-word (e.g., *zelefè*, Figure 1.2) is displayed in the middle of the computer screen. The children’s task is to decide whether spoken and written non-words are the same or different. This task requires grapheme-to-phoneme translation skills, because written non-words must be translated into a phonological code first to be compared with spoken non-words. When spoken and written non-words are the same, children must press a green button on the keyboard for a *yes* response. When they are different children must press a red button for a *no* response. In ProDi-H, both non-words are presented auditorily. The task is designed for Grades 1 to 4 in both modalities.
Introduction and Aims

Lexical Decision Task

The efficiency of orthographical decoding skills is assessed via a *lexical decision task*. Written words (e.g., *Donner*/*thunder*, Figure 1.3) or non-words (e.g., *Maum*) are presented in the middle of the screen. After the presentation of each word or non-word children indicate whether or not they know the item by pressing the green button for *yes* or the red button for *no*. Fast and accurate responses to this task require efficient orthographical decoding skills. Even though regular words and non-words can be recognized (or, in case of non-words, rejected) correctly via both phonological recoding and orthographical decoding, fast responses to these items require automatized access to orthographical word forms stored in sight vocabulary. In addition, the test also involves irregular word forms (e.g., *Jeans* or *Cowboy*) and a subset of pseudo-homophones (e.g., *Hechse* instead of *Hexe*/witch). Pseudo-homophones are non-words that sound like real words but have a different orthography. Fast and accurate responses to irregular words and pseudo-homophones require efficient orthographical decoding skills, because phonological recoding leads to false negative responses for irregular words and false positive responses for pseudo-homophones. Words and non-words are presented auditorily in the ProDi-H lexical decision task. The task is designed for Grades 1 to 4 in both modalities.

*Figure 1.2*. Example of the presentation of a non-word in the ProDi-L phonological comparison task.
Introduction and Aims

Figure 1.3. Example of the presentation of a real word in the ProDi-L lexical decision task.

Semantic Classification Task

The efficiency of access to word meanings is assessed via a semantic classification task. Children are presented with a spoken category name (e.g., Insekten/insects). Subsequently, a written word is presented in the middle of the screen (e.g., Ameise/ant, see Figure 1.4). The task is to decide whether or not the written word is a member of the auditory-presented category by pressing the green button for yes or the red button for no. Fast and accurate decisions about the category membership of a word require sufficient semantic knowledge of the word and efficient access to its meaning. To prevent correct responses on the basis of associative relations between category names and words, the test also contains category-word pairs that are associated but not related by hyponymy (e.g., Sportarten–Schiedsrichter/sports–referee). Decisions based on associative relations should lead to false positive responses for these items. Participants are presented with spoken category names and words in the ProDi-H auditory semantic verification task. The task is designed for Grades 1 to 4 in both modalities.

Figure 1.4. Example of the presentation of a member from the insects category in the ProDi-L semantic classification task.
Introduction and Aims

Grammaticality Judgment Task

The efficiency of syntactic integration skills is assessed via a grammaticality judgment task. Children are presented with short written sentences in the middle of the screen. These sentences are either grammatically correct (e.g., *Die Sterne leuchten am Himmel/*The stars are shining in the sky) or contained a word-order violation (e.g., *Steffi gut Klavier spielt/*Steffi well plays the piano, see Figure 1.5), a tense-form violation (e.g., *Der Drachen ist in die Luft flog/*The kite has flew up in the air), or a case-marking violation (e.g., *Der Frosch fängt der-GEN Fliege mit seiner klebrigen Zunge/*The frog catches the-GEN fly with his sticky tongue). The task is to decide whether or not a sentence is grammatically correct by pressing the green button for yes or the red button for no. Accurate and fast responses to the ungrammatical sentences require sufficient grammatical knowledge about word-order, case marking, and verb-tense form inflection in German. Spoken sentences are presented in the ProDi-H auditory grammaticality judgment task. The ProDi-L task is designed for Grades 3 and 4, the ProDi-H task for Grades 1 to 4.

Semantic Verification Task

The efficiency of semantic integration skills is assessed via a semantic verification task. Short written sentences are presented in the middle of the screen. These sentences contain true assertions of the world (e.g., *Die meisten Vögel können fliegen/*Most birds can fly, see Figure 1.6) or false assertions (e.g., *Treppen sind ein rotes Gemüse/*Stairs are red vegetables). The task was to decide whether the assertions are true by pressing the green button for yes or the red button for no. Fast and accurate responses require efficient semantic integration skills, that is, the ability to establish coherent mental representations of the
sentence content and to validate them against world knowledge. Spoken sentences are presented in the ProDi-H auditory semantic verification task. The ProDi-L task is designed for Grades 2 to 4, the ProDi-H task for Grades 1 to 4.

Figure 1.6. Example for a true assertion in the ProDi-L semantic verification task.

Semantic Verification Task with Sentence Pairs

The efficiency of establishing local coherence is assessed via a semantic verification task with sentence pairs. Short written sentence pairs are presented in the middle of the screen. After reading the first sentence, children press the spacebar and the second sentence appears below the first sentence. The second sentence is either coherent with the first sentence (e.g., *Lena war zu lange in der Sonne. Darum bekam sie einen Sonnenbrand/Lena stayed in the sun for too long. Therefore, she got a sunburn) or incoherent (e.g., *Katrin muss ins Krankenhaus. Sie ist nämlich ganz gesund/*Katrin has to go to the hospital because she feels totally fine, see Figure 1.7). The task is to decide whether or not the sentences go together by pressing the green button for yes or the red button for no. To provide fast and accurate responses, the reader must derive coherent mental representations from both sentences and then integrate them into a conjoint mental model by establishing a mental link of coherence between them. Spoken sentence pairs are presented in the ProDi-H auditory semantic verification task with sentence pairs. The ProDi-L task is designed for Grades 3 and 4, the ProDi-H task for Grades 1 to 4.
Introduction and Aims

Research Questions

In the preceding sections, I have reviewed and discussed extant literature and empirical studies on the cognitive component skills of reading comprehension at the word level (i.e., phonological recoding, orthographical decoding, and access to word meanings), at the sentence level (syntactic and semantic integration), and at the text level (establishing coherence relations, drawing inferences, and comprehension monitoring). Although, researchers widely accept that reading comprehension depends on the successful accomplishment of these cognitive component skills, several questions concerning interdependence among these skills, their corresponding skills in listening comprehension, and their development during reading acquisition remain unanswered. The aim of this dissertation is to advance a more comprehensive understanding of reading comprehension by empirically exploring a selective subset of open research questions.

The first two empirical studies of this dissertation aim to further explore the relationship of visual word recognition skills and reading comprehension in developing readers. The first study presented in Chapter II investigates the extent that phonological recoding and orthographical decoding skills uniquely contribute to reading comprehension in beginning German readers and how both reading comprehension skills change over the course of reading development. Most (developmental) models of visual word recognition have been designed to explain processes of visual word recognition in English readers. However, because German is characterized by higher orthographic consistency than English (e.g., Landerl, Wimmer, & Frith, 1997; Wimmer & Goswami, 1994), it seems likely to find cross-linguistic differences in the contributions of both visual word recognition skills to reading comprehension. Given that grapheme-to-phoneme conversion is highly consistent and reliable
• Introduction and Aims •

in German, not only beginning but also advanced German readers are likely to rely (at least to a certain degree) on phonological recoding during written word recognition instead of shifting completely from phonological recoding to orthographical decoding (as proposed, for example, by Frith’s three-stage developmental model, 1985, 1986). However, an overall weak relationship can be expected between phonological recoding skills and reading comprehension even in beginning German readers. Several studies found that phonological recoding skills are easily acquired and perfected early (by the end of Grade 1) in reading development because of the high orthographic consistency in German (e.g., Landerl & Wimmer, 2008; Landerl et al., 1997; Seymour et al., 2003; Wimmer & Goswami, 1994). Consequently, individual differences in German primary school children’s phonological recoding skills should diminish early during reading acquisition (Pfost, 2015) and thus account for only little variance in reading comprehension (at least this should hold for accuracy measures of phonological recoding; large individual differences have been found in speed measures, Landerl & Wimmer, 2008). This assumption is consistent with Gough and Tunmer’s (1986) simple view of reading \( R = D \times C \), which implies that the power of decoding skills \( D \) to predict reading comprehension \( R \) should decrease with increasing decoding skills. Accordingly, phonological recoding skills should be weakly related to reading comprehension in beginning and advanced German readers. By analogy, the contribution of orthographical decoding skills to reading comprehension should initially increase with increasing reading experience and growing sight vocabulary but should eventually decrease when orthographical decoding skills approach ceiling. The first empirical study addresses all of these issues in a cross-sectional study with German 2nd to 4th graders.

The second empirical study presented in Chapter III aims to test the simple view of reading by Gough and Tunmer (1986) by using a stringent methodology that overcomes several shortcomings and limitations of extant studies. The simple view of reading is one of the most prominent and influential theories in the field of reading research. It states that reading comprehension \( R \) is a function of the multiplicative combination of decoding skills \( D \) and more general linguistic (listening) comprehension skills \( C \): \( R = D \times C \). The simple view of reading is appealing and convincing because of its mere simplicity, and because several studies have provided evidence in favor of the view (e.g., Hoover & Gough, 1990; Joshi & Aaron, 2000; Kendeou et al., 2009). However, most of these studies are subject to various methodological shortcomings and limitations. First, decoding \( D \) is usually operationalized as the “ability to pronounce (or silently apprehend the pronunciation of) pseudowords” (Gough & Tunmer, 1986, p. 7). In this vein, phonological recoding skills are
assessed as an indicator of decoding. Although this approach might be appropriate for beginning readers who primarily use phonological recoding to recognize written words that are not yet part of their sight vocabulary, it might not hold for skilled readers who are capable of recognizing most orthographical word forms directly. A second methodological limitation concerns measures of listening comprehension ($C$) and reading comprehension ($R$). According to the simple view of reading, listening and reading comprehension involve basically the same cognitive processes after written word forms have been decoded. Thus, listening and reading comprehension should be assessed via the same tasks using parallel materials differing only with respect to modality of presentation (Tunmer & Chapman, 2012). Yet, many studies have tested the simple view of reading with different tasks and materials to assess reading and listening comprehension skills (e.g., Joshi & Aaron, 2000; Kendeou et al., 2009). As a consequence, the possibility that their findings (either consistent or inconsistent with the view) were confounded by task differences cannot be ruled out. The second empirical study of this dissertation aims to test the simple view of reading by using optimized and stringent measures of decoding, listening comprehension, and reading comprehension. A second aim of the study is to test the generalizability of the simple view of reading to German readers from Grades 3 and 4. To my knowledge, the only published study that has tested the simple view of reading in a German sample with German materials was conducted by Marx and Jungmann (2000). In their results, $D$ and $C$ explained a substantial amount of variance in $R$. However, researchers should be careful in interpreting their findings as support for the simple view of reading, because they had not tested the key assumption of the simple view that the product of $D \times C$ is a better predictor of reading comprehension than their additive combination $D + C$. Thus, whether or not the simple view of reading in its most simplistic form can be generalized to German readers is still an open question.

The third empirical study presented in Chapter IV further explores the establishment of local coherence relations in German developing readers. More specifically, the study focuses on the processing of positive-causal coherence relations, which link a cause and its consequence (e.g., *Lena war zu lange in der Sonne. Darum bekam sie einen Sonnenbrand*/Lena stayed in the sun for too long. Therefore, she got a sunburn) and negative-causal coherence relations, which add a contrastive meaning or negation to the causal link (e.g., *Sandra war nicht müde. Trotzdem ging sie ins Bett*/Sandra was not tired. Nevertheless, she went to bed) (Sanders et al., 1992). Previous studies have shown that the processing of negative-causal coherence relations is cognitively more demanding, that is, more error-prone and slower than the processing of positive-causal coherence relations (e.g., Goldman &
The cumulative cognitive complexity approach by Evers-Vermeul and Sanders (2009) and Spooren and Sanders (2008) attributes these differences to the internal cognitive complexity of positive-causal and negative-causal coherence relations. Negative-causal coherence relations are assumed to be cognitively more complex than positive-causal coherence relations, because negative-causal coherence relations add a contrast or negation to the causal link. Consequently, they evoke more cognitive processing effort and are acquired later than positive-causal coherence relations. Several studies with English or Dutch samples and materials reported findings that can be interpreted in support of the cumulative cognitive complexity approach (e.g., Bloom, Lahey, Hood, Lifter, & Fiess, 1980; Evers-Vermeul & Sanders, 2009; Goldman & Murray, 1992; Katz & Brent, 1968; Millis & Just, 1994, Exp. 4; Spooren & Sanders, 2008). However, comparatively little is known about the development and processing of positive-causal and negative-causal coherence relations and connectives in German. Thus, the first aim of the third empirical study was to explore how positive-causal and negative-causal coherence relations are processed in German children and adults. The second aim of the study was to directly compare the processing of both coherence relations in spoken and written language processing using parallel tasks and materials for both modalities, which has not been tested in previous studies to date. According to the simple view of reading formula $R = D \times C$, cognitive processes of reading and listening comprehension should be basically the same after written word forms have been decoded. Accordingly, establishing mental relations of local coherence should be highly comparable for written and spoken language processing. However, approaches such as Perfetti’s (1985) verbal efficiency hypothesis suggest that little cognitive resources remain for higher-level comprehension processes such as establishing coherence relations in beginning readers who still need sufficient cognitive resources to recognize written word forms. Hence, beginning readers should exhibit more difficulties establishing coherence relations (and cognitively complex negative-causal coherence relations in particular) in written compared to spoken language comprehension. The final aspect of the third study addresses the lack of clarity in how the processing of positive-causal and negative-causal coherence relations develops throughout reading acquisition. Findings by Cain et al. (2005) suggest that adequate comprehension of connectives and coherence relations still develops during primary school years. According to the cumulative cognitive complexity approach, this assumption is likely to hold for the most complex coherence relations in particular, namely negative-causal coherence relations. Thus, the final aim of the study was to
investigate how the processing of positive-causal and negative-causal coherence relations develops throughout primary school years.

The fourth empirical study of this dissertation presented in Chapter V aims to demonstrate the construct validity of process-oriented tests that assess individual differences in cognitive component skills of reading comprehension selectively. Psychological reading tests need to assess the complex structure of cognitive component skills that constitute reading comprehension to be able to ascertain meaningful diagnoses of reading deficits and develop adaptive, target-oriented intervention programs. The study demonstrates the usefulness of process-oriented reading tests by using the example of the ProDi-L grammaticality judgment task for German 3rd and 4th graders. Explanatory item response models provide evidence of the construct validity of the test by demonstrating that theoretically motivated variations of item features (i.e., grammaticality, syntactic complexity, and type of grammatical violation), which are assumed to facilitate or impede the process of syntactic integration, are able to predict empirically observed item difficulties (a similar procedure was used by Neeb, Naumann, Knoepke, Isberner, & Richter, 2015 to demonstrate the construct validity of the phonological comparison task of ProDi-L and ProDi-H). The analyses of the ProDi-L grammaticality judgment task are accompanied by analyses of the ProDi-H auditory grammaticality judgment task, which allows for a direct comparison of syntactic integration skills in listening and reading comprehension of German 3rd and 4th graders. The research questions of all four empirical studies of this dissertation are further elaborated in the Chapters II to V.

Finally, Chapter VI provides an overview of the cognitive component skills of reading comprehension (i.e., phonological recoding, orthographical decoding, access to word meanings, syntactic and semantic integration, establishing coherence, inference drawing, and comprehension monitoring) and argues for the usefulness of characterizing poor readers by their deficits in these components skills instead of categorizing them based on whether or not they have dyslexia. The basic premise of Chapter VI is that characterizing poor readers according to a single deficit is flawed. Instead, sources of poor reading comprehension are shown to be multifaceted and heterogeneous and to vary between poor readers. Accordingly, diagnostic tests that aim to detect individual sources of poor reading comprehension and corresponding reading intervention programs should consider the complex structure of cognitive component skills of reading comprehension to allow for a meaningful diagnoses and an adaptive, target-oriented intervention that addresses the sources of poor reading.
Chapter II

Study 1

Phonological Recoding, Orthographic Decoding, and Comprehension Skills During Reading Acquisition

A version of this chapter is published as:

Phonological Recoding, Orthographic Decoding, and Comprehension Skills During Reading Acquisition

Julia Knoepke, Tobias Richter, Maj-Britt Isberner¹, Johannes Naumann², & Yvonne Neeb³

Abstract. To become skillful readers, children have to acquire the ability to translate printed words letter by letter into phonemic representations (phonological recoding) and the ability to recognize the written word forms holistically (orthographical decoding). Whereas phonological recoding is the key for learning to read and useful for recognizing unknown or low-frequent words, orthographical decoding is often more efficient and takes less time, thus facilitating reading processes on the sentence and text level. Several studies with English-speaking children provided evidence for the relevance of the two routes but the question whether and to what extent both word recognition skills contribute to reading comprehension in young German readers requires further clarification. Based on data from a cross-sectional study with German primary school children we investigated whether and to what extent both types of word recognition skills are associated with sentence (N=666) and text comprehension skills (N=149) and how these relationships develop from Grade 2 to 4. The results indicate that both phonological recoding skills and orthographical decoding skills are important for reading comprehension skills. Their relative weight does not change across grade levels.

Keywords: Orthographical decoding, phonological recoding, reading comprehension, reading acquisition, reading development

Introduction

Research on reading comprehension has focused on the central role of word recognition skills in skillful reading (e.g., Gough & Tunmer, 1986; Hoover & Gough, 1990; Perfetti, 1985). According to the dual route cascaded model (DRC) proposed by Coltheart et al. (2001) word recognition can be achieved via two different routes: a phonological route by which every

¹ University of Kassel
² Goethe University Frankfurt
³ Maiden name: Kutzner; German Institute for International Educational Research
word has to be recoded letter by letter based on grapheme-phoneme-correspondence rules, and an orthographical route by which written words can be directly mapped onto mental representations of word forms without an intermediary step of grapheme-to-phoneme-translation. Several studies with English-speaking children and adults provided evidence for the relevance of the two routes (e.g., Paap & Noel, 1991; Shankweiler, Lundquist, Dreyer, & Dickson, 1996) and the implication of the dual route model that both phonological recoding skills and orthographical decoding skills should be fundamental for children’s reading comprehension (Cunningham & Stanovich, 1990; Shankweiler et al., 1999; Tunmer & Chapman, 2012). However, most of the available evidence comes from studies with English-speaking children, which differs from languages such as German in orthographic transparency. Thus, the question whether and to what extent phonological recoding skills and orthographical decoding skills also contribute to the reading comprehension of children learning to read in German requires further clarification. A second and related issue refers to the time course of development of both skills during reading acquisition. Frith (1986) assumes that children learning to read proceed from a letter-by-letter recoding strategy to an orthographical decoding strategy, which becomes the dominant strategy in skillful readers. In the present study we addressed these questions by examining the relative contributions of phonological recoding and orthographical decoding skills to reading comprehension skills in primary school children from Grades 2 to 4.

In the following sections we will highlight the dominant role of successful word recognition as precursor of reading comprehension. Afterwards, we will discuss the role of phonological recoding and orthographical decoding in visual word recognition and its implications for individual differences in reading comprehension skill. Subsequently, we will turn to the time course of development of word reading skills. Finally, we will present our study and discuss the findings and implications with respect to cognitive models of visual word recognition (e.g., DRC model, Coltheart et al., 2001; strong phonological model, Frost, 1998) and the developmental model of word recognition proposed by Frith (1986).

The Role of Word Recognition in Reading Comprehension

To become skillful readers children have to acquire several cognitive skills at the word, sentence, and text level (Müller & Richter 2014; Perfetti, 2001; Richter & Christmann, 2009). They have to learn to recognize written words, to retrieve their meanings from the mental lexicon, to integrate word meanings into sentence interpretations, and to build and continuously update a coherent mental model of the text. An important line of research has
focused on the crucial role of well-functioning word-level processes for good reading comprehension. As pointed out by Perfetti in his *verbal efficiency hypothesis* (1985) and by Perfetti and Hart in their *lexical quality hypothesis* (2001; 2002; see also Perfetti’s *DVC Decoding, Vocabulary, and Comprehension triangle*, 2010), reliable representations of word forms and meanings and their rapid and efficient retrieval is at the core of skillful reading comprehension. Efficient processes at the word level are assumed to release cognitive resources that can be used at higher levels of processing such as the sentence and text level. Another approach stressing the unique role of word recognition in reading comprehension is the *simple view of reading* (Gough & Tunmer, 1986; Hoover & Gough, 1990). According to this view, reading comprehension ($R$) is defined as the product of decoding skills ($D$) and linguistic comprehension ($C$): $R = D \times C$. Thus, Gough and Tunmer assume that reading comprehension can be perfectly predicted by a reader’s ability to decode words from written text and his or her general ability to comprehend language. Thus, the only predictor specific to reading comprehension that differs from spoken language comprehension is an individual’s ability to recognize written words. In support of this assumption, several studies provided evidence for a strong correlation of word recognition abilities and reading comprehension in children using various tasks to measure word recognition (e.g., word and pseudoword reading: Golinkoff & Rosinski, 1976; Hoover & Gough, 1990; Joshi & Aaron, 2000; Shankweiler et al., 1999; letter and word identification: Kendeou et al., 2009; lexical decision task: Knoepke, Richter, Isberner, Neeb, & Naumann, 2013; Richter et al., 2013).

**Word Recognition: Phonological Recoding and Orthographical Decoding**

The question of how to conceptualize and to operationalize word recognition skills is a matter of debate (Kirby & Savage, 2008; Knoepke et al., 2013; Tunmer & Champman, 2012; see also Hoover & Gough, 1990). Gough and Tunmer (1986) stated that “word recognition skill (in an alphabetic orthography) is fundamentally dependent upon knowledge of letter-sound correspondence rules” (p. 7). However, this definition incorporates only one of two possible routes of word recognition assumed by the DRC model (Coltheart, 2005; Coltheart et al., 2001), namely the route of phonological recoding or the *non-lexical route*. Via this phonological route, words are translated letter by letter into a phonological code by means of grapheme-phoneme-correspondence rules. Based on the phonological code the respective lexical entry is retrieved from the mental lexicon in a way similar to auditory word recognition. This route is most important for beginning readers because it enables the reader to recode new and unknown word forms based on single grapheme-phoneme mappings.
However, more experienced readers have built up a *sight vocabulary* (Ehri, 2005a) that allows a more rapid and efficient word recognition. The reader recognizes orthographic word forms as a whole and maps them directly onto his or her lexical entries without the preliminary step of grapheme-to-phoneme translation. This route of orthographical decoding is called the *lexical route* (Coltheart, 2005; Coltheart et al., 2001). Orthographical decoding allows effortless recognition of familiar words and words with irregular spelling which are already part of the *sight vocabulary*. According to the DRC model, both routes start to operate in parallel when encountering a word. The route that recognizes the word faster and more reliably, that is, the more efficient route, gains stronger activation and accesses the lexical entry: In particular when known words with high frequency are processed, the orthographical route is more efficient. In contrast, low-frequent and unknown words are more likely to be recognized via the phonological, non-lexical route (see Paap & Noel, 1991; for an application of the DRC model to German see Ziegler et al., 2000).

The assumption of the DRC model that skillful readers use phonological information only for reading low-frequent or unknown words implies that orthographic decoding skills should be far more important in these readers than phonological decoding skills. In contrast to this view, the connectionist triangle model (Plaut et al., 1996; see also Seidenberg & McClelland, 1989) and the strong phonological model (Frost, 1998) suggest a slightly different conclusion. According to the triangle model, phonological information is stored in the mental lexicon and used along with orthographic and meaning representations whenever words are processed. Thus, word recognition regularly involves phonological information. The relative importance of phonological and orthographic information is assumed to depend on the word characteristics (frequency and consistency) and reading proficiency. Importantly, however, the triangle model implies that phonological information is always used to some degree in word recognition. Similarly, the strong phonological model (Frost, 1998) posits that phonological information is accessed automatically and early in word processing, suggesting that it plays a regular role in word recognition (for meta-analytic evidence from masked priming studies, see Rastle & Brysbaert, 2006).

**Development of Phonological Recoding and Orthographical Decoding**

Beginning readers must recode an unknown written word at least once letter by letter before information about its orthographic form can be added to the mental lexicon. As a consequence, phonological recoding is the prerequisite for the development of the orthographical decoding route. Thus, beginning readers should basically rely on the
phonological route because almost every word in its written form is new and unknown to them. Only recognizing the same written word over and over again allows the reader to build up a representation of the word form as part of the mental lexicon. These orthographical representations can then be used for word recognition as well, either in a direct way (lexical route according to the DRC model, Coltheart et al., 2001) or in concert with phonological and meaning representations (according to the triangle model, Plaut et al., 1996). Thus, the assumption seems reasonable that word recognition in beginning readers is primarily accomplished via the phonological recoding route and in experienced readers via the orthographical decoding route (Hoover & Gough, 1990; Tunmer & Chapman, 2012).

A prominent approach describing the development of word-recognition skills during reading acquisition is the three-stage developmental model by Frith (1986). She assumes three reading strategies of word recognition, which are acquired in a serial order. It is important to note that the strategies in Frith’s model are not under the strategic control of the reader: Rather, they describe processes that occur automatically and usually without conscious awareness of the reader, at least when they are sufficiently routinized (Ehri, 2005b). The first strategy children employ is the logographic strategy. Often prior to entering reading instruction, children are able to recognize a small sample of words based on their graphic features, that is, children are not aware of the grapheme structure of words but read them in an iconic fashion. The second strategy children use when they enter reading instruction is the alphabetic strategy. They systematically recode words letter by letter, translating each grapheme into its corresponding phoneme (this strategy is consistent with the phonological recoding route; for a more differentiated distinction of alphabetic strategies see Ehri, 2005b). The final and most mature strategy children acquire when learning to read is the orthographic strategy. Children recognize words as whole orthographic units without translating graphemes into phonemes first (this strategy is consistent with the orthographical decoding route assumed in the DRC model, Coltheart et al., 2001). Frith further assumes that the acquisition of the three strategies is serial, that is, children proceed from one strategy to the next whereby each strategy is built on the previous one. Accordingly, children proceed from the logographic strategy to the alphabetic strategy and subsequently to the orthographic strategy.

The developmental model of reading proposed by Frith (1986) differs from the dual route model (Coltheart et al., 2001) and—even more so—from the triangle model (Plaut et al., 1996) and the strong phonological model (Frost, 1998) in its implications for the relative importance of phonological recoding and orthographical decoding skills. Although, as discussed earlier, experienced readers should assign more weight to the orthographical
decoding route because it is more efficient, dual route models such as the DRC model (Coltheart et al., 2001) nevertheless predict readers to make use of both routes: In case of unknown or low-frequent words readers are expected to use the phonological recoding route and in case of familiar, high-frequent, and irregular words they should employ the route of orthographical decoding. The triangle model (Plaut et al., 1996) and the strong phonological model (Frost) even assign a more prominent role to phonological recoding skills in more experienced readers. The reason is that these models predict that phonological information is regularly used whenever words are processed. In line with these assumptions, the importance of both orthographic and phonological skills was confirmed in several studies with children (e.g., Richter et al., 2013; Shankweiler et al., 1999; Tunmer & Chapman, 2012), adolescents, and adults (Paap & Noel, 1991; Shankweiler et al., 1996), which found both phonological recoding and orthographical decoding abilities to be highly predictive of reading comprehension. However, given that orthographical decoding is usually more efficient, it seems reasonable to assume that readers might at least gradually shift from a rather phonologically based to a rather orthographically based word recognition strategy during reading acquisition.

The Present Study

The present study followed three related aims. The first aim was to investigate whether both phonological recoding skills and orthographical decoding skills are predictive of reading comprehension skills in German primary school children and which unique contributions both skills make to reading comprehension skill. All models of word recognition discussed so far and the developmental three-stage model (Frith, 1986) were originally proposed for English-speaking children. However, German is a language with higher orthographical consistency than English (e.g., Landerl et al., 1997; Wimmer & Goswami 1994), that is, most German words have a regular spelling perfectly consistent with the German grapheme-to-phoneme-correspondence rules. As a consequence, the phonological recoding route might be more efficient than it is in languages such as English which are characterized by a comparatively high amount of irregularly spelled words (e.g., Seymour et al., 2003). It would therefore not be surprising to find phonological skills to be more strongly predictive of reading comprehension in German than they are in English (see the discussion in Ziegler et al., 2000).

A second aim of the present study was to examine the potential shift from a rather
phonologically based recoding strategy in beginning readers to a rather orthographically based decoding strategy in advanced readers as predicted by Frith (1986). If such a shift occurs, we would expect to find a decrease in the strength of the relationship of phonological recoding skills and reading comprehension skills with increasing grade level. At the same time, we would expect the strength of the relationship of orthographical decoding skills and reading comprehension skills to increase. However, if children do not shift from a phonological to an orthographical word recognition strategy but instead rely on both word recognition skills once they are sufficiently developed (Coltheart, 2005; Coltheart et al., 2001), we would expect to find phonological recoding skills to remain highly predictive of reading comprehension skills across all grade levels. A somewhat different prediction is implied by the strong phonological model (Frost, 1998) and by the connectionist triangle model (Plaut et al., 1996; see also Seidenberg & McClelland, 1989) according to which phonological processing is always involved in visual word identification.

Another issue to be addressed by the present study was whether skills of word recognition gradually become less predictive of reading comprehension skills in children with enhanced phonological recoding and orthographical decoding skills. The simple view of reading formula $R = D \times C$ (Gough & Tunmer, 1986; Hoover & Gough, 1990) predicts that on a high level of word recognition skills ($D$) reading comprehension skills ($R$) should depend exclusively on general (higher-level) comprehension skills ($C$) and vice versa. Thus, the improvement of phonological recoding and orthographical decoding skills should reduce their predictive power for reading comprehension because variance in reading comprehension skills might then be better accounted for by general comprehension skills at the sentence and text level (such as inference skills and meta-cognitive strategies, e.g., Oakhill et al., 2003). In contrast to this view, studies with normally developing adolescents and adults found that even for older and more advanced readers both types of skills are highly predictive of reading comprehension (Paap & Noel, 1991; Shankweiler et al., 1996). Thus, the third aim of the present study was to explore whether the relationships of phonological recoding and orthographical decoding skills with reading comprehension are linear across the complete range of individual differences in word-level skills or whether the relationships follow a quadratic trend with a stronger relationship in the lower range of word-level skills.
Method

Participants

Participants were 992 primary school children recruited from 21 schools (72 classes) in Cologne, Frankfurt am Main and Kassel (Germany). Of these children, 967 took part in the sentence comprehension task and 214 children took part in the text comprehension task; 189 children participated in both tasks.

Participants of the sentence comprehension task. The data of 100 children (10.3%) were excluded from the analyses of the sentence comprehension data because data were missing for more than 20% of the trials in at least one of the tasks included in the analysis. Moreover, the data of 201 non-native German speaking children (20.8%) were also excluded from the analyses. Of the remaining 666 children (325 boys and 329 girls, for 12 children gender information was missing), 232 children were in Grade 2 (age: $M=8.41$; $SD=0.39$; $Min=7.25$; $Max=10.25$), 190 children were in Grade 3 (age: $M=9.44$; $SD=0.56$; $Min=8.17$; $Max=11.83$), and 244 children were in Grade 4 (age: $M=10.42$; $SD=0.42$; $Min=9.00$; $Max=12.42$).

Participants of the text comprehension task. The data of 22 children (10.3%) were excluded from the analyses of the sentence comprehension data because data were missing for more than 20% of the trials in at least one of the tasks included in the analysis. The data of 43 non-native German speaking children (20.1%) were also excluded from the analyses. Of the remaining 149 children (68 boys and 74 girls, for 7 children gender information was missing), 58 children were in Grade 2 (age: $M=8.29$; $SD=0.35$; $Min=7.58$; $Max=9.08$), 40 were in Grade 3 (age: $M=9.38$; $SD=0.41$; $Min=8.33$; $Max=10.42$), and 51 were in Grade 4 (age: $M=10.29$; $SD=10.38$; $Min=9$; $Max=11.17$).

Socio-demographic data were collected via a questionnaire completed by the parents and were supplemented by a questionnaire completed by the teacher when information was missing. Children only participated in the study when parents provided written consent.

Variables

The phonological recoding task, the orthographical decoding task, and the sentence comprehension task described in the following paragraphs were taken from the computerized German-speaking test battery ProDi-L (Richter et al., in press; see also Richter et al., 2012).

Phonological recoding skills. Children’s phonological recoding skills were assessed by the computerized phonological comparison task embedded in ProDi-L. Children were
presented with 64 pairs of pseudowords (62 test pairs and 2 practice pairs), that is, meaningless strings of phonemes or graphemes that were congruent with the phonological and orthographical rules of German. They were asked to indicate whether the two pseudowords matched (e.g., *risamo* – *risamo*) or mismatched (e.g., *tebedika* – *tebudiki*). The first pseudoword was presented orally over headphones, the second pseudoword was presented subsequently in written form on the notebook screen. Children responded by pressing a green button on the keyboard for *match* or a red button for *mismatch*. Half of the pseudoword pairs consisted of matching, the other half of mismatching pseudowords. They consisted of one to four open syllables with a simple consonant-vowel structure. The mismatching pseudowords contained either one or two mismatching vowels, which occurred in stressed or unstressed syllables. The orally presented stimuli were recorded by a male speaker.

**Orthographical decoding skills.** Children’s ability to orthographically decode words was assessed by a computerized lexical decision task taken from ProDi-L. Children were presented with 92 written words (e.g., *Traktor/tractor*) and pseudowords (e.g., *Spinfen*). Their task was to indicate whether the presented item was a real word or not. Half of the presented items were existing German words and the other half were pseudowords. Children responded by pressing a green button on the keyboard for *yes, this is a real word* or a red button for *no, this is not a real word*. Both words and pseudowords varied in length (number of word characters: $M=5.68$; $SD=1.08$; $Min=3$; $Max=10$; number of pseudoword characters: $M=6.09$; $SD=2.02$; $Min=3$; $Max=12$) and frequency (log-transformed frequency of words: $M=1.81$; $SD=1.03$; $Min=0.00$; $Max=3.77$; Mannheim Corpus of the CELEX data base for written German; Baayen, Piepenbrock & Gulikers, 1995). The words which were used to construct the pseudowords (e.g., the pseudoword *Maum* was created by replacing the first letter of *Baum/tree*) were matched in frequency with the real words used in this task (log-transformed frequency: $M=1.66$; $SD=0.97$). Pseudowords varied in the degree to which they resembled real German words. Pseudowords similar to existing words (e.g., *Nand*) were based on words with a regular German spelling (e.g., *Sand/sand*) and pseudowords dissimilar to existing words (e.g., *Koveau*) were based on words with an irregular German spelling (e.g., *Niveau/level*). Some of the pseudowords were pseudohomophones, which sound like an existing German word but have a different orthography (e.g., *Heckse* instead of *Hexe/witch*).

**Sentence comprehension skills.** Children’s ability to comprehend sentences was assessed by a computerized sentence verification task taken from ProDi-L. Children were presented with 46 written declarative sentences (44 test sentences and 2 practice sentences), which either contained a true statement about the world (e.g., *Züge fahren auf*
Schienen/Trains run on rails) or a false statement (e.g., Treppen sind ein rotes Gemüse/Stairs are red vegetables). Children were asked to verify the statements by pressing the green button on the keyboard for yes, the statement is correct or the red button for no, the statement is not correct. Half of the sentences contained true and the other half false statements. The sentences varied in the number of propositions and length (number of characters: $M=34.87$; $SD=12.37$; $Min=15$; $Max=61$). The true statements also varied in predictability (e.g., predictable: Giraffen haben lange Hälse/Giraffes have long necks; less predictable: Zitronen sind gesunde Früchte/Lemons are healthy fruits), and the false sentences varied in the degree to which they violated general world knowledge (e.g., Schnecken sind schnell/Snails are fast), that is, expressed propositions that were sensible but incongruent with well-known facts, or rather semantic knowledge (e.g., Husten ist blau/Cough is blue), that is, expressed propositions that were incongruent with basic semantic features of the focal word (for a discussion of the distinction between world knowledge violations and semantic anomalies, see Isberner & Richter, 2014).

**Text comprehension skills.** Children’s text comprehension skills were assessed by the sub test *Text Comprehension* of ELFE 1–6 (computerized version, Lenhard & Schneider, 2006), which is a standardized reading comprehension test battery for German primary school children from Grade 1 to Grade 6. Children were presented with 20 short texts and were asked to answer questions concerning the content of each text by choosing one of four multiple-choice items.

**Procedure**

Phonological recoding skills, orthographical decoding skills, and sentence comprehension skills were assessed in the context of a cross-sectional study investigating processes of listening and reading comprehension in German primary school children with various measures on the word, sentence, and text level (ProDi-L: Prozessbezogene Diagnostik des Leseverstehens bei Grundschulkindern, Richter et al., in press; see also Richter et al., 2012; Richter et al., 2013). Children were tested together in classrooms of the participating schools. All written items were presented on notebook computers (font: Verdana, visual angle: 1.5 degrees). The three tasks were embedded in a story of an extraterrestrial named Reli who came to earth to learn the earthlings’ language. He asked the children to help him by indicating when he did something wrong. Reli introduced the tasks in short animated video clips and walked the children through them. All test items were presented in randomized order. Prior to each task, children saw two practice items for which they received feedback.
from Reli. When children gave an incorrect answer, the practice items were repeated until they answered all practice items correctly. As measures of efficiency log-transformed response times (measured from stimulus onset to the press of the response button), which are assumed to reflect the degree of routinisation of cognitive processes, and response accuracies, which are assumed to reflect the reliability of cognitive processes, were recorded. The log-transformation served to normalize the distribution of the otherwise skewed response time data. For each child and each task, the mean was calculated across all items for response accuracies and log-transformed response times respectively. In addition, the log-transformed reaction times were screened and adjusted for outliers by a two-step procedure. First, reaction times 3 standard deviations or more below the item-specific mean were discarded from the analysis. Second, reaction times 2 standard deviations or more below or above the person-specific mean were discarded from the analysis. From the two measures, an integrated test score was calculated by dividing children’s mean accuracy by their mean log-transformed response times. These integrated test scores reflect both the degree of reliability and routinisation, that is, the overall efficiency of a cognitive process. In the following analyses we will therefore primarily focus on the results found for integrated test scores while taking accuracy and response time results into account to support their interpretation.

Subsequent to the tasks of the cross-sectional study, the computerized ELFE subtest Text Comprehension was conducted. The ELFE test scores, which were calculated by counting the number of children’s correct responses, served as measure of text comprehension skills. The tasks were presented in two separate sessions, which lasted approximately 45 min, and were carried out on different days.

Results

Mean log-transformed response times, mean accuracies and integrated test scores as measures of sentence comprehension and the ELFE test scores as measures of text comprehension were analyzed using Hierarchical Linear Models (HLM, Raudenbush & Bryk, 2002) with intercepts randomly varying between school classes. All models were estimated with the software package lme4 for R (Bates, Maechler, & Bolker, 2011). All significance tests were based on a Type I error probability of 0.05 (two-tailed). Descriptive statistics for the sentence-comprehension measures are reported in Table 2.1 and for the text-comprehension measures in Table 2.2.
Table 2.1:

Descriptive Statistics for Response Time (Log-transformed), Accuracy, and Integrated Test Scores as Dependent Variables in the Phonological Recoding Task, Orthographical Decoding Task, and the Sentence Comprehension Task (N=666)

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Grade 2</th>
<th>Grade 3</th>
<th>Grade 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M (SD)</td>
<td>n</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Phonological Recoding Task</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated Test Scores(^a)</td>
<td>666</td>
<td>0.115 (0.02)</td>
<td>232</td>
<td>0.107 (0.02)</td>
</tr>
<tr>
<td>Response Accuracy(^b)</td>
<td>666</td>
<td>0.853 (0.12)</td>
<td>232</td>
<td>0.812 (0.14)</td>
</tr>
<tr>
<td>Response Time(^c)</td>
<td>666</td>
<td>7.432 (0.31)</td>
<td>232</td>
<td>7.574 (0.33)</td>
</tr>
</tbody>
</table>

| Orthographical Decoding Task |            |               |             |               |             |               |
| Integrated Test Scores\(^a\) | 666        | 0.117 (0.02)  | 232  | 0.105 (0.01)  | 190  | 0.118 (0.01)  | 244  | 0.127 (0.01)  |
| Response Accuracy\(^b\) | 666        | 0.857 (0.09)  | 232  | 0.797 (0.09)  | 190  | 0.866 (0.08)  | 244  | 0.905 (0.07)  |
| Response Time\(^c\) | 666        | 7.374 (0.38)  | 232  | 7.616 (0.36)  | 190  | 7.373 (0.33)  | 244  | 7.146 (0.28)  |

| Sentence Comprehension Task |            |               |             |               |             |               |
| Integrated Test Scores\(^a\) | 666        | 0.113 (0.01)  | 232  | 0.108 (0.01)  | 190  | 0.113 (0.01)  | 244  | 0.118 (0.01)  |
| Response Accuracy\(^b\) | 666        | 0.932 (0.07)  | 232  | 0.918 (0.07)  | 190  | 0.934 (0.08)  | 244  | 0.943 (0.06)  |
| Response Time\(^c\) | 666        | 8.264 (0.40)  | 232  | 8.528 (0.39)  | 190  | 8.251 (0.32)  | 244  | 8.024 (0.29)  |

Note. \(^a\)Response Accuracy/Response Time, \(^b\)relative frequency, \(^c\)log-transformed.

Separate models were estimated for sentence comprehension and text comprehension measures as dependent variables. For both the sentence comprehension and the text comprehension data, we first estimated a model for the integrated test scores which reflect measures of the efficiency of component processes of reading. To interpret the results for the integrated test scores, we subsequently estimated models for the response accuracies and the response times. Because the only measure that was available for the ELFE sub test Text Comprehension was the number of correct responses, the outcome variable was the same in all three models. The predictors were entered into the model in two steps. 1) In the first step,
phonological recoding skills, orthographical decoding skills, and grade level were included as grand-mean centered predictors. Moreover, grade-level interaction terms with phonological recoding skills and orthographical decoding skills were included into the model. These interaction terms accounted for possible developmental changes in the extent that phonological recoding and orthographical decoding skills predict sentence and text comprehension skills. 2) In the second step, phonological recoding skills and orthographical decoding skills were squared and added to the model. The quadratic terms allowed to test for the assumption that the linear correlations between both word recognition measures and sentence and text comprehension skills might be more strongly pronounced for children with poor word recognition skills as compared to children with more advanced word recognition skills.

Table 2.2:

Descriptive Statistics for Response Time (Log-transformed), Accuracy, and Integrated Test Scores as Dependent Variables in the Phonological Recoding Task and the Orthographical Decoding Task, and the ELFE Test Scores of the Text Comprehension Task (N=149)

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Grade 2</th>
<th>Grade 3</th>
<th>Grade 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M (SD)</td>
<td>n</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Phonological Recoding Task</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated Test Scores(a)</td>
<td>149</td>
<td>0.114 (0.02)</td>
<td>58</td>
<td>0.108 (0.02)</td>
</tr>
<tr>
<td>Response Accuracy(b)</td>
<td>149</td>
<td>0.843 (0.12)</td>
<td>58</td>
<td>0.807 (0.13)</td>
</tr>
<tr>
<td>Response Time(c)</td>
<td>149</td>
<td>7.412 (0.29)</td>
<td>58</td>
<td>7.515 (0.26)</td>
</tr>
<tr>
<td>Orthographical Decoding Task</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated Test Scores(a)</td>
<td>149</td>
<td>0.118 (0.02)</td>
<td>58</td>
<td>0.107 (0.01)</td>
</tr>
<tr>
<td>Response Accuracy(b)</td>
<td>149</td>
<td>0.860 (0.09)</td>
<td>58</td>
<td>0.807 (0.08)</td>
</tr>
<tr>
<td>Response Time(c)</td>
<td>149</td>
<td>7.358 (0.40)</td>
<td>58</td>
<td>7.561 (0.42)</td>
</tr>
<tr>
<td>Text Comprehension Task</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELFE Test Score(d)</td>
<td>149</td>
<td>11.604 (4.88)</td>
<td>58</td>
<td>9.241 (4.08)</td>
</tr>
</tbody>
</table>

Note. \(a\)Response Accuracy/Response Time, \(b\)relative frequency, \(c\)log-transformed, \(d\)number of correct responses.
Table 2.3:  
**Fixed Effects and Variance Components in the HLM with Response Accuracy, Response Time (Log-transformed), and Integrated Test Scores as Dependent Variables for Sentence Comprehension Skills and Response Accuracy, Response Time (Log-transformed), and Integrated Test Scores as Measures of Phonological Recoding Skills and Orthographical Decoding Skills**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sentence Comprehension Task</th>
<th>Fixed Effects</th>
<th>Variance Components</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Integrated Test Score a</td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B (SE)</td>
<td>B (SE)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Response Accuracy b</td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B (SE)</td>
<td>B (SE)</td>
</tr>
<tr>
<td></td>
<td>Response Time c</td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B (SE)</td>
<td>B (SE)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.113*</td>
<td>0.113*</td>
<td>0.934*</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Phonological Recoding Skill d</td>
<td>0.078*</td>
<td>0.135*</td>
<td>0.088*</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Orthographical Decoding Skill d</td>
<td>0.372*</td>
<td>0.377*</td>
<td>0.301*</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Grade Level d</td>
<td>0.000</td>
<td>0.001</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Grade level X Phonological Recoding Skill</td>
<td>-0.002</td>
<td>-0.006</td>
<td>-0.027</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Grade level X Orthographical Decoding Skill</td>
<td>-0.046</td>
<td>-0.010</td>
<td>-0.039</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Phonological Recoding Skill 2</td>
<td>2.495*</td>
<td>0.214</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>(0.77)</td>
<td>(0.13)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Orthographical Decoding Skill 2</td>
<td>-4.032*</td>
<td>-1.523*</td>
<td>-0.89</td>
</tr>
<tr>
<td></td>
<td>(1.11)</td>
<td>(0.27)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Grade level X Phonological Recoding Skill 2</td>
<td>0.876</td>
<td>0.137</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.90)</td>
<td>(0.15)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Grade level X Orthographical Decoding Skill 2</td>
<td>-3.539*</td>
<td>-0.952*</td>
<td>-0.077</td>
</tr>
<tr>
<td></td>
<td>(1.17)</td>
<td>(0.28)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Variance Components</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School Class</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>

*Note.* a Response Accuracy/Response Time, b proportion of correct responses (person means), c log-transformed (person means), d grand-mean centered.

* p < .05.
Finally, we included the interaction terms of grade level with the squared predictors into the model to account for possible developmental changes in the quadratic relationships. In all models, the intercepts were allowed to vary randomly between school classes to account for the nested data structure (students nested within classes). The parameter estimates for the fixed and random effects in the HLMs are provided in Table 2.3 for the sentence comprehension measures as dependent variables and in Table 2.4 for the text comprehension measures as dependent variables.

Sentence Comprehension Task

*Integrated test scores.* The HLM for the integrated test scores revealed significant main effects for phonological recoding skills and orthographical decoding skills (Table 2.3, Model 1). Children with more efficient phonological recoding processes ($\beta=0.08; t(660)=3.8; p<.05$) and more efficient orthographical decoding processes ($\beta=0.37; t(660)=16.3; p<.05$) also exhibited more efficient sentence comprehension skills. The positive relationship with sentence comprehension skills was more strongly pronounced for orthographical decoding skills than for phonological recoding skills. Including the squared predictor variables and their interactions with grade level did not change the significance of the linear effects (Table 2.3, Model 2). However, Model 2 revealed significant main effects for the quadratic terms of phonological recoding skills ($\beta=2.49; t(656)=3.2; p<.05$) and orthographical decoding skills ($\beta=-4.03; t(656)=-3.6; p<.05$) as well as a significant interaction of grade level and squared orthographical decoding skills ($\beta=-3.54; t(656)=-3.0; p<.05$). The quadratic regression for the squared predictor variables is displayed in Figure 2.1a and b. The positive relationship between orthographical decoding skills and sentence comprehension skills was strongest in the lower range of orthographical decoding skills and became weaker with increasing orthographical decoding skills. Moreover, this pattern became more pronounced with increasing grade level. In contrast, for phonological recoding skills, the quadratic relationship with sentence comprehension skills showed a reverse pattern: Unexpectedly, the positive relationship with sentence comprehension skills was weaker in the lower range of phonological recoding skills and became stronger with increasing phonological recoding skills. In order to clarify the sources of the effects for the integrated measures, we estimated two further models with the accuracy and reaction time data.

*Accuracy data.* The HLM for the mean accuracy data revealed significant main effects for phonological recoding and orthographical decoding skills (Table 2.3, Model 1). Children who responded with high accuracy in the phonological recoding task ($\beta=0.09; t(660)=3.6;$
Study 1

$p<.05$) and with high accuracy in orthographical decoding task ($\beta=0.30; t(660)=9.0; p<.05$) also responded with higher accuracy to the sentences in the sentence comprehension task. This relationship was stronger for orthographical decoding skills than it was for phonological recoding skills. Including the quadratic terms and their interaction with grade level in the model did not change the significance of the linear effects (Table 2.3, Model 2). However, Model 2 revealed a significant main effect for squared orthographical decoding skills ($\beta=-1.52; t(656)=-5.6; p<.05$) and a significant interaction of grade level and squared orthographical decoding skills ($\beta=-0.95; t(656)=-3.4; p<.05$). The relationship between orthographical decoding skills and sentence comprehension skills was strongest in the lower range of orthographical decoding accuracy and became weaker the more the accuracy values approached 100%. This pattern became more pronounced with increasing grade level.

Response latencies. The HLM for the mean log-transformed response times (Table 2.3, Model 1) again revealed significant main effects for phonological recoding skills ($\beta=0.10; t(660)=2.8; p<.05$), orthographical decoding skills ($\beta=0.75; t(660)=24.1; p<.05$), and grade level ($\beta=-0.06; t(660)=-4.8; p<.05$) as well as a significant interaction of grade level and orthographical decoding skills ($\beta=-0.09; t(660)=-2.4; p<.05$). The main effect for grade level indicates an overall increase in response speed with increasing grade level. Children who provided faster responses in the phonological recoding and the orthographical decoding task also responded faster in the sentence comprehension task. This relationship was slightly more pronounced for the response times in the orthographical decoding task compared to the phonological recoding task but it decreased slightly with grade level. Including the quadratic terms and their interactions with grade level into the model did not change the significance of the linear effects (Table 2.3, Model 2). In contrast to the integrated test scores and the accuracy data, none of the squared predictor variables or their interactions reached significance.

The results of the sentence comprehension task can be summarized as follows: Consistent with the assumptions of the DRC model (Coltheart et al., 2001) and the triangle model (Plaut et al., 1996) orthographical decoding skills were predictive of sentence comprehension skills at all grade levels. Second, phonological recoding skills were also associated with sentence comprehension skills. This relationship was weaker than the one of orthographical decoding skills and sentence comprehension but it was nevertheless substantial and did not decrease from Grade 2 to Grade 4. The consistently strong relationship of phonological recoding skills with sentence comprehension is compatible with the triangle model as well as the strong phonological model (Frost, 1998) but not so much with the DRC
model and the developmental model by Frith (1986). The described pattern was found in the integrated test scores and both of its components, that is, the accuracy data and the reaction time data.

Figure 2.1. Linear and quadratic relationships between (a) phonological recoding skills (integrated test scores) and (b) orthographical decoding skills (integrated test scores) with sentence comprehension skills (integrated test scores).
## Table 2.4:

**Fixed Effects and Variance Components in the HLM with the ELFE Test Scores as Dependent Variable for Text Comprehension Skills and Response Accuracy, Response Time (Log-transformed), and Integrated Test Scores as Measures of Phonological Recoding Skills and Orthographical Decoding Skills**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Integrated Test Score&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Response Accuracy&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Response Time&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 1</td>
</tr>
<tr>
<td></td>
<td>$B$ ($SE$)</td>
<td>$B$ ($SE$)</td>
<td>$B$ ($SE$)</td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(0.48)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>Phonological Recoding Skill&lt;sup&gt;d&lt;/sup&gt;</td>
<td>50.778*</td>
<td>90.796*</td>
<td>4.930</td>
</tr>
<tr>
<td></td>
<td>(20.30)</td>
<td>(28.46)</td>
<td>(2.68)</td>
</tr>
<tr>
<td>Orthographical Decoding Skill&lt;sup&gt;d&lt;/sup&gt;</td>
<td>201.715*</td>
<td>206.836*</td>
<td>38.698*</td>
</tr>
<tr>
<td></td>
<td>(25.36)</td>
<td>(31.17)</td>
<td>(4.46)</td>
</tr>
<tr>
<td>Grade Level&lt;sup&gt;d&lt;/sup&gt;</td>
<td>-0.190</td>
<td>-0.253</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
<td>(0.58)</td>
<td>(0.40)</td>
</tr>
<tr>
<td>Grade level X Phonological Recoding Skill</td>
<td>7.719</td>
<td>43.743</td>
<td>-0.932</td>
</tr>
<tr>
<td></td>
<td>(23.50)</td>
<td>(34.28)</td>
<td>(3.13)</td>
</tr>
<tr>
<td>Grade Level X Orthographical Decoding Skill</td>
<td>-4.435</td>
<td>-7.622</td>
<td>6.202</td>
</tr>
<tr>
<td></td>
<td>(27.69)</td>
<td>(36.61)</td>
<td>(4.87)</td>
</tr>
<tr>
<td>Phonological Recoding Skill&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1970.453*</td>
<td>26.803</td>
<td>-7.308*</td>
</tr>
<tr>
<td></td>
<td>(814.98)</td>
<td>(15.81)</td>
<td>(3.63)</td>
</tr>
<tr>
<td>Orthographical Decoding Skill&lt;sup&gt;2&lt;/sup&gt;</td>
<td>-1818.701</td>
<td>17.278</td>
<td>-3.717</td>
</tr>
<tr>
<td></td>
<td>(1495.19)</td>
<td>(44.30)</td>
<td>(2.51)</td>
</tr>
<tr>
<td>Grade level X Phonological Recoding Skill&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1733.629</td>
<td>29.316</td>
<td>2.557</td>
</tr>
<tr>
<td></td>
<td>(926.87)</td>
<td>(18.91)</td>
<td>(3.54)</td>
</tr>
<tr>
<td>Grade level X Orthographical Decoding Skill&lt;sup&gt;2&lt;/sup&gt;</td>
<td>-2648.430</td>
<td>-45.782</td>
<td>-2.272</td>
</tr>
<tr>
<td></td>
<td>(1635.44)</td>
<td>(49.72)</td>
<td>(2.71)</td>
</tr>
<tr>
<td><strong>Variance Components</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School Class</td>
<td>0.000</td>
<td>0.375</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.61)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

*Note.*<sup>a</sup>Response Accuracy/Response Time, <sup>b</sup>proportion of correct responses (person means), <sup>c</sup>log-transformed (person means), <sup>d</sup>grand-mean centered.

* $p < .05.$
Moreover, the strength of the positive relationship of orthographical decoding skills with sentence comprehension was strongest in the lower range of orthographical decoding skills. This pattern was found for the integrated test scores and also for the accuracy data but not for the reaction time data. In contrast, the relationship between phonological recoding skills and sentence comprehension skills was more strongly pronounced in the upper range of phonological recoding skills. This pattern was found only for the integrated test scores. Thus, the assumption derived from the simple view of reading (Gough & Tunmer, 1986) that the relationship of word recognition skills and reading comprehension skills is weakest in the upper range of word recognition skills was supported only partially by the data.

**Text Comprehension Task**

*Integrated test scores.* The HLM for the ELFE test scores and the integrated test scores as measure of phonological recoding and orthographical decoding skills revealed significant main effects for phonological recoding and orthographical decoding skills (Table 2.4, Model 1). Children with more efficient phonological recoding processes ($\beta=50.78; t(143)=2.5; p<.05$) and more efficient orthographical decoding processes ($\beta=201.71; t(143)=8.0; p<.05$) also reached higher test scores in the text comprehension task (see Figure 2.2a and b). The positive relationship with text comprehension skills was more strongly pronounced for orthographical decoding skills than for phonological recoding skills. Including the squared predictor variables and their interactions with grade level did not change the significance of the linear effects (Table 2.4, Model 2). However, Model 2 revealed a significant main effect for the quadratic term of phonological recoding skills ($\beta=1970.45; t(139)=2.4; p<.05$). The quadratic regression is depicted in Figure 2.2a. Again, unexpectedly, the positive relationship between phonological recoding skills and text comprehension skills was weaker in the lower range of phonological recoding skills. Again, in order to clarify the sources of the effects for the integrated measures, we estimated two further models with the accuracy and reaction time data.

*Accuracy data.* The HLM for the ELFE test scores and the mean accuracies as measure of phonological recoding and orthographical decoding skills revealed a significant main effect for orthographical decoding skills (Table 2.4, Model 1). Children who provided highly accurate responses in the orthographical decoding task also provided highly accurate responses in the text comprehension task ($\beta=38.70; t(143)=8.7; p<.05$). Including the quadratic terms and their interaction with grade level changed the significance of one of the linear effects (Table 2.4, Model 2). Now, Model 2 revealed a significant main effect for
phonological recoding skills ($\beta=10.14; t(139)=2.4; p<.05$). Children who responded with higher accuracy in the phonological recoding task also responded with higher accuracy in the text comprehension task. The positive relationship with text comprehension skills was more strongly pronounced for orthographical decoding skills than for phonological recoding. None of the squared predictor variables or their interactions reached significance.

**Response latencies.** The HLM for the ELFE test scores and the mean log-transformed response times as measure of phonological recoding and orthographical decoding skills revealed significant main effects for grade level and orthographical decoding skills (Table 2.4, Model 1). Overall, children reached higher test scores in the text comprehension task with increasing grade level ($\beta=1.38; t(143)=2.9; p<.05$). Children who provided faster responses in the orthographical decoding task also provided more accurate responses in the text comprehension task ($\beta=-5.72; t(143)=-4.1; p<.05$). Including the quadratic terms and their interaction with grade level did not change the significance of the linear effects (Table 2.4, Model 2). However, Model 2 revealed a significant main effect for squared phonological recoding skills ($\beta=-7.31; t(139)=-2.0; p<.05$), indicating that the relation between phonological recoding skills and text comprehension skills was strongest in the upper range of response times in the phonological recoding task.

The results for the text comprehension task can be summarized as follows: Again, phonological recoding skills as well as orthographical decoding skills were predictive of text comprehension across all grade levels, as was predicted by the DRC model (Coltheart et al., 2001) and the triangle model (Plaut et al., 1996). Children with better phonological recoding skills and better orthographical decoding skills exhibited better text comprehension skills. This pattern was found in the integrated test scores and the accuracy data for phonological recoding and orthographical decoding skills and for orthographical decoding skills in the reaction time data. The positive relationship was strongest for orthographical decoding skills and text comprehension skills as predicted by the DRC model in particular. Nevertheless, the relationship of phonological recoding skills and text comprehension was substantial in all grade levels. This finding coheres well with the triangle model and the strong phonological model (Frost, 1998). Finally, the strength of the positive relationship of phonological recoding skills and text comprehension skills was weaker in the lower range of phonological recoding skills. This was found for the integrated test scores. In contrast, the negative relationship of phonological recoding times and text comprehension skills was weakest for faster phonological recoding times. This finding was predicted by the simple view of reading (Gough & Tunmer, 1986). However, the assumption that the relationship of word recognition
skills and reading comprehension skills is weakest in the upper range of word recognition skills was supported only partially by the data.

Figure 2.2. (a) Linear and quadratic relationships between phonological recoding skills (integrated test scores) and text comprehension (ELFE test scores) and (b) linear relationship between orthographical decoding skills (integrated test scores) with text comprehension skills.
Discussion

The first aim of the present study was to investigate whether and to what extent phonological recoding skills and orthographical decoding skills are both predictive of reading comprehension skills in German primary school children. We found both phonological recoding and orthographical decoding skills to be predictive of sentence and text comprehension skills at all grade levels. The one exception was that text comprehension skills were not predicted by the speed of phonological recoding processes but only by their accuracy and the integrated test scores. Given that the integrated test scores capture both the speed and the reliability aspect of the efficiency of phonological recoding processes, our results indicate that, overall, phonological recoding skills are predictive of text comprehension as well. Consistent with the DRC model of word recognition (Coltheart, 2005; Coltheart et al., 2001), orthographical decoding skills appeared to be more strongly predictive of sentence and text comprehension skills than phonological recoding skills. Apparently, by the end of Grade 2, many German primary school children have already built a sufficient sight vocabulary that enables them to access the lexical entries of many words directly and efficiently from their written forms via the lexical route. Nevertheless, phonological recoding skills made a significant contribution to reading comprehension across all grade levels. Given that the test items of the ProDi-L sentence comprehension test and the ELFE text comprehension test did not contain very rare words or words likely to be unknown to primary school children (at least not to those in upper grade levels), we infer that phonological skills are relevant for word recognition in primary school children beyond the restricted category of unknown or low-frequent words. Thus, the results indicate that the children regularly made use of phonological information whenever words were recognized. This conclusion is in line with the triangle model (Plaut et al., 1996) as well as the strong phonological model (Frost, 1998). Furthermore, because of the high orthographical consistency in German words (Landerl et al., 1997; Wimmer & Goswami, 1994), it seems not surprising that phonological decoding abilities play a prominent role during reading comprehension in beginning and more experienced German readers as well (see discussion in Ziegler et al., 2000).

The second aim of the present study was to investigate a potential shift from a rather phonologically based recoding strategy in beginning readers to a rather orthographically based decoding strategy as predicted by Frith’s (1986) three-stage developmental model of reading. Such a shift would be indicated by interactions of grade level with phonological recoding and orthographical decoding skills. However, only one interaction with grade level reached significance and its pattern runs counter the predictions implied by Frith’s model:
The positive relationship between sentence comprehension speed and orthographical decoding speed decreases from Grade 2 to 4. This interaction was in contrast to Frith’s predictions. No further grade-level interactions with one of the linear predictors reached significance. In sum, phonological recoding and orthographical decoding skills were both predictive of reading comprehension at all grade levels. However, it is possible that we might have observed evidence in favor of a shift if we had investigated a broader range of grade levels. In contrast to children at the end of Grade 2, first graders and early second graders might have relied more strongly on a phonological strategy and less strongly on an orthographic strategy. Nevertheless, according to Frith we would then expect to find orthographical decoding skills to be exclusively predictive of reading comprehension in Grades 2 to 4. Although orthographical decoding skills were more strongly predictive of reading comprehension than phonological recoding skills, they were not exclusively predictive of reading comprehension. One explanation supporting Frith’s model might be that children gradually shift from one strategy to the next with an intervening phase where the two strategies overlap for some time (see Frith). Our results might reflect this phase of overlap where children rely on both the phonological and the orthographical strategy before finally shifting to the orthographical strategy. However, it seems very unlikely that such an overlap would persist for two years without any indication of a change. Thus, we found no evidence in favor of the three-stage developmental model of reading proposed by Frith for German primary school children from Grades 2 to 4.

The final issue we addressed was whether the strength of the relationship of phonological recoding and orthographical decoding skills with reading comprehension depends on the level of children’s word recognition skills. To this end, we included phonological recoding and orthographical decoding skills as squared predictors into the multilevel regression analyses. We found that orthographical decoding skills predicted sentence comprehension more strongly in children with poorer decoding skills and less strongly in children with highly developed decoding skills (this was found for integrated test scores and accuracy data). In addition, this effect was more strongly pronounced for older than for younger children. One possible interpretation of this pattern of results is that the more efficient orthographic decoding functions, the more variance in reading comprehension skills has to be attributed to other skills. This explanation is consistent with the simple view of reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), which states that an improvement in word recognition \( D \) reduces its predictive power of reading comprehension \( R \) by leaving the remaining variance in \( R \) to be explained by listening comprehension skills \( C \) (see e.g.,
It is also consistent with the lexical quality hypothesis (Perfetti & Hart, 2001) and the verbal efficiency hypothesis (Perfetti, 1985) according to which efficient word recognition skills are a necessary but by no means sufficient condition for successful reading comprehension. Previous studies have identified several skills that affect reading comprehension performance in primary school children beyond single word recognition such as semantic integration processes, comprehension monitoring, working memory (Oakhill et al., 2003), inference making (Cain & Oakhill, 1999), and grammatical sensitivity (Willow & Ryan, 1986). Thus, we would assume that in children with highly developed orthographical decoding skills, such higher-order reading skills might play a more important role in reading comprehension. Remarkably, the squared predictor orthographical decoding skill explained a significant amount of variance only in sentence comprehension skills but not in text comprehension skills. One explanation for this difference could be the fact that different types of tasks were used to assess children’s sentence and text comprehension skills. We will return to this issue later when we address potential limitations of the present study.

One unexpected finding requiring further clarification is that, in contrast to orthographical decoding skills, phonological recoding skills were more strongly associated with sentence and text comprehension skills in children with highly developed phonological recoding skills compared to children with poorly developed phonological recoding skills (this was found for integrated test scores in both the sentence and text comprehension task and for response times in the text comprehension task). How can this unexpected finding be explained? As can be seen from Tables 2.1 and 2.2, the majority of children have already developed fairly high phonological recoding skills and only few children appear to have extensive difficulties with phonologically recoding pseudowords. This is not surprising. As discussed earlier, it seems reasonable to assume that by the end of Grade 2 the majority of children have already acquired sufficient phonological recoding skills. However, it seems that within this major group of children with highly developed phonological recoding skills there is a smaller but nevertheless systematic amount of variance left to account for differences in reading comprehension. In contrast, the variance in the smaller group of children with very poor phonological recoding skills might be rather unsystematic. A possible explanation is that children with poor phonological recoding skills use word recognition strategies other than phonological recoding to compensate for their poor word recognition performance. For example, they might recognize many words based on some salient graphical features rather than grapheme-to-phoneme translations (logographic strategy, Frith, 1986). It might also be the case that they use sentence- or text-level skills such as context or world knowledge in a
top-down fashion to infer the words they have difficulties to recognize (as assumed by the interactive compensatory model by Stanovich, 1980). Overall, we assume that children with poor phonological recoding skills probably rely on other skills below or beyond the word level to compensate for poor word recognition abilities.

In sum, we found that both word recognition skills, phonological recoding and orthographical decoding, are associated with reading comprehension skills throughout all grade levels. If we assume a causal relationship between the two word-level and reading-comprehension skills, namely that efficient phonological recoding and orthographical decoding are at the core of skilled reading comprehension even at Grades 3 and 4, our results have some practical implications with respect to reading education. First, because both skills, phonological recoding and orthographical decoding, make individual and separable contributions to reading comprehension, early reading acquisition should be supported by practical exercises aimed at fostering these skills. Furthermore, the fact that we found significant relationships of phonological recoding and orthographical decoding skills with reading comprehension skills even in third and fourth graders highlights the possibility that fostering both word level skills even in older and more advanced readers might be fruitful to enhance their reading comprehension skills. Another implication concerns children who exhibit word recognition difficulties. Here, it is essential to find out which word level skill exactly is impaired to what extent and to create an optimal individual support plan to ensure target-oriented training for the impaired readers.

The results of the present study need to be interpreted with its limitations in mind. First, we differentiate between only two skills of visual word recognition: phonological recoding and orthographical decoding skills. However, we have to consider the possibility that there are processing units on a level between single graphemes or whole orthographical word forms, which might assist word recognition during reading, such as morphemes and syllables. For example, several studies have established syllable-frequency effects or syllable-length effects in Spanish (e.g., Barber, Vergara, & Carreiras, 2004), French (e.g., Ferrand & New, 2003), and German (e.g., Conrad & Jacobs, 2004), suggesting that syllables play a role in visual word recognition. Due to the fact that the phonological recoding task in our study does not differentiate between the recognition of single graphemes and whole syllables, we cannot rule out the possibility that our phonological recoding measures reflect syllable recognition skills to some extent. As a result, it is possible that children’s ability to efficiently recognize written syllables account for a unique portion of variance in reading comprehension. This issue requires further clarification in future research.
The second potential limitation concerns the comparability of the sentence- and text-comprehension tasks. Both tasks assess comprehension but are likely to differ in terms of cognitive requirements and measures. Whereas children had to verify whether sentence contents made sense in the sentence comprehension task by providing yes/no-answers under mild time pressure, the ELFE subtest text comprehension required the processing of text passages, of a multiple-choice-question following each passage, and the identification of the correct response out of four possibilities. Thus, performing the text comprehension task might have involved more complex linguistic cognitive processes (such as the establishment of local and global coherence, see van Dijk & Kintsch, 1983) and extra-linguistic cognitive processes (such as keeping information active in working memory and comparing several alternative answers) than the sentence comprehension task. In part, these differences are simply due to the fact that text comprehension per se is a more complex task than sentence comprehension. However, it must be noted that due to the multiple-choice format of the text comprehension task the text comprehension data might also to some extent reflect offline comprehension processes and strategies that are not part of text comprehension itself. Nevertheless, the results we found for both the sentence and the text comprehension task were fairly comparable for all grade levels, schools, and school classes.

A third limitation of the present study is its cross-sectional design. In fact, the best way to investigate developmental questions (such as the applicability of Frith’s, 1986, three-stage developmental model of reading to German primary school children) is by means of longitudinal designs. We cannot rule out the possibility that the absence of a potential shift from a rather phonological to a rather orthographical word recognition strategy as predicted by Frith (1986) in our data is due to accidental grade level differences. Moreover, Frith points out that each child progresses from one strategy to the next at his or her own pace independent of age or grade level. Thus, a longitudinal investigation might possibly reveal evidence in favor of her theory, which we failed to track down with a cross-sectional design. However, our findings are perfectly in line with several recent studies demonstrating that both phonological recoding and orthographical decoding are highly associated with reading comprehension in children, adolescents, and even adults (e.g., Paap & Noel, 1991; Richter et al., 2013; Shankweiler et al., 1996; Shankweiler et al., 1999; Tunmer & Chapman, 2012). Furthermore, our findings appear to be consistent throughout all three grade levels as well as for different schools and school classes. Therefore, it seems unlikely that our findings were simply due to accidental grade level differences.

To conclude, our results consistently demonstrate the significant role that both
phonological recoding and orthographical decoding skills play in successful reading comprehension throughout the elementary school years, somewhat surprisingly even in Grades 3 and 4. Educators should take both routes of word recognition processes into account when designing reading curricula and interventions for poor readers.

Acknowledgments. The research presented in this article was supported by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF, grants 01 GJ 0985, 01 GJ 0986, 01 GJ 1206A, 01 GJ 1206B). We would like to thank David Nitz for programming assistance and the numerous student research assistants who were involved in collecting data for this project. Researchers who are interested in the material used in the visual and auditory grammaticality judgment tasks are invited to send an e-mail to the first or the second author.
Chapter III

Study 2

Leseverstehen = Hörverstehen X Dekodieren? Ein stringenter Test der Simple View of Reading bei deutschsprachigen Grundschulkindern [Reading comprehension = listening comprehension X decoding? A stringent test of the simple view of reading in German primary school children]

A version of this chapter is published as:

Einleitung

Für das theoretische Verständnis von Leseschwierigkeiten in der Grundschule und eine entsprechende zielgerichtete Diagnostik, Prävention und Förderung ist die Frage zentral, auf welchen kognitiven Teilfähigkeiten das Leseverstehen beruht. Eine Theorie, die auf diese Frage eine frappierend einfache Antwort gibt, ist die sogenannte einfachen Sicht auf das Lesen \((\text{Simple View of Reading}, \text{ Gough & Tunmer, 1986; Hoover & Gough, 1990})\). Nach der Simple View of Reading lassen sich individuelle Unterschiede im Leseverstehen als das Produkt von zwei Teilfähigkeiten beschreiben, nämlich der Fähigkeit der Dekodierung geschriebener Wörter und einer allgemeinen Fähigkeit des Sprachverstehens, die auch dem Hörverstehen zu Grunde liegt. Leseschwierigkeiten entstehen demnach aus Schwierigkeiten bei der visuellen Worterkennung (Dyslexie), aus Schwierigkeiten im allgemeinen Sprachverstehen (Hyperlexie) oder aus Schwierigkeiten in beiden Bereichen.

war, die Simple View of Reading bei deutschsprachigen Grundschulkindern der dritten und vierten Klassen in methodisch stringenter Weise zu überprüfen.\textsuperscript{4}

\textbf{Die Simple View of Reading}


Die in der Simple View of Reading enthaltene Annahme, dass sowohl die Fähigkeit der visuellen Worterkennung als auch allgemeine Fähigkeiten des Sprachverstehens notwendige Bedingungen für das Leseverstehen darstellen, impliziert, dass sich das Leseverstehen als Funktion einer multiplikativen Verknüpfung (d.h. als Produkt) der beiden

\textsuperscript{4} Die Daten wurden im Rahmen des Verbundprojektes „Prozessbezogene Diagnostik des Lese- und Hörverstehens im Grundschulalter“ erhoben, das als Teil der „Förderinitiative Sprachdiagnostik und Sprachförderung (FiSS)“ vom Bundesministerium für Bildung und Forschung (Förderkennzeichen 01GJ0985) gefördert wurde.
Variablen darstellen lässt (vgl. Gough & Tunmer, 1986): \( R = D \times C \). Die Implikationen einer multiplikativen Verknüpfung lassen sich besonders gut veranschaulichen, wenn man hypothetisch davon ausgeht, dass die Fähigkeiten der visuellen Worterkennung und des allgemeinen Sprachverstehens von 0 (= Fähigkeit nicht vorhanden) bis 1 (= Fähigkeit perfekt ausgeprägt) skaliert sind (Hoover & Gough, 1990). Unter dieser Voraussetzung lässt sich aus der multiplikativen Verknüpfung einerseits ableiten, dass bei perfekter Ausprägung einer der beiden Fähigkeiten die verbleibende Varianz im Leseverstehen vollständig von der anderen Fähigkeit determiniert wird (z.B. \( R = 1 \times C = C \)). Andererseits ist bei einer multiplikativen Verknüpfung das Verstehen von Schriftsprache nichtmöglich, wenn eine der beiden Fähigkeiten überhaupt nicht beherrscht wird (z.B. \( R = 0 \times C = 0 \)). Um Leseverstehen überhaupt zu ermöglichen, müssen beide Fähigkeiten zumindest in minimalem Maße vorhanden sein; sie können, anders als bei einer additiven Beziehung, das vollständige Fehlen der jeweils anderen Fähigkeit nicht kompensieren.

Die Annahme, dass das Leseverstehen auf dem Produkt der Fähigkeiten des Dekodierens und Hörverstehens aufbaut, hat eine Reihe von Untersuchungen zu der Frage angestoßen, in welcher Form \( D \) und \( C \) als Prädiktoren von \( R \) zusammenwirken. In einer Längsschnittuntersuchung mit bilingualen (englisch-spanischsprachigen) Kindern vom Kindergartenalter bis zur vierten Klasse zeigten Hoover und Gough (1990), dass das Einbeziehen des Produkts aus \( D \) und \( C \) ein Inkrement gegenüber der linearen Kombination liefert (bis zu 7% mehr Varianzaufklärung in der dritten Klasse). Zudem konnten die Autoren ihre aus der Annahme einer multiplikativen Beziehung abgeleitete Vorhersage stützen, dass bei schlechten Lesern/-innen, die entweder gut dekodieren oder gut verstehen können, die jeweils andere Fähigkeit defizitär ist.

Eine weitere Vorhersage, die sich aus der Simple View of Reading ergibt, besteht darin, dass ein gutes Niveau im Leseverstehen garantiert ist, wenn im individuellen Fall gute Dekodier- und Hörverstehensfähigkeiten vorliegen. Fälle, in denen die visuelle Worterkennung und das Hörverstehen effizient funktionieren, aber trotzdem kein adäquates Leseverständnis erreicht werden kann, widersprechen der Simple View of Reading (Gough & Tunmer, 1986). Eine solche Probandengruppe jedoch fanden Georgiou, Das und Hayward (2009) mit 50 kanadischen indigenen Kindern der dritten und vierten Klasse, die sich trotz altersgerechter Fähigkeiten im Dekodieren und Hörverstehen durch schlechte Lesefähigkeiten auszeichneten. Georgiou et al. vermuteten daher, dass das Leseverstehen in dieser Gruppe nicht als das Produkt aus Dekodier- und Hörverstehensfähigkeit modelliert werden kann. Im Sinne dieser Annahme zeigte in ihrer Untersuchung das Produkt von Dekodier- und

Aus diesem letzten Befund ergibt sich die Frage, ob die visuelle Worterkennung und das Hörverstehen (unabhängig davon, ob sie additiv und/oder multiplikativ zusammenwirken) alleine ausreichend sind, um die komplexe Fähigkeit des Leseverstehens und die zahlreichen daran beteiligten Prozesse umfassend zu beschreiben, oder ob weitere Fähigkeiten herangezogen werden müssen (wie z.B. die Benennungsgeschwindigkeit, Joshi & Aaron, 2000, Studie 2; Johnston & Kirby, 2006). Wie im folgenden Abschnitt gezeigt wird spielt dabei die Operationalisierung der theoretischen Konstrukte der Simple View of Reading eine entscheidende Rolle.

**Die Operationalisierung von Fähigkeiten der visuellen Worterkennung und des Hör- und Leseverstehens**

Angesichts der vielfältigen kognitiven Teilprozesse, die den Fähigkeiten der visuellen Worterkennung ($D$), dem Hör- ($C$) und dem Leseverstehen ($R$) zu Grunde liegen (Bredel & Reich, 2008; Müller & Richter, 2013; Richter & Christmann, 2009), kommt der Frage, wie diese gemessen werden sollen, eine besondere Bedeutung zu (vgl., Kendou et al., 2009; Vellutino, Tunmer, Jaccard, & Chen, 2007). Nur dann, wenn die drei theoretischen Konstrukte tatsächlich in einer kognitionspychologisch fundierten und konsistenten Weise operationalisiert werden, ist ein sinnvoller Test der Simple View of Reading möglich.
Operationalisierung der visuellen Worterkennung


vermehrten Nutzung orthographischer Vergleichsprozesse bei der Leseentwicklung allmählich und nicht bei allen Kindern zur selben Zeit erfolgt (Frith, 1986). Außerdem benutzen auch geübte Leser/-innen beide Routen, wobei die direkte orthographische eher bei bekannten und häufigen Wörtern, die indirekte eher bei unbekannten und seltenen Wörtern relevant ist.


*Operationalisierung des Lese- und Hörverstehens*

Im Vergleich zur visuellen Worterkennung stellt das Verstehen gesprochener und geschriebener Sprache auf der Diskurs- bzw. Textebene eine noch deutlich komplexere kognitive Leistung dar, die auf einer Vielzahl kognitiver Teilprozesse auf Wort-, Satz- und Textebene beruht (Richter & Christmann, 2009). Daraus ergeben sich zwei Anforderungen an die Operationalisierungen des Hörverstehens ($C$) und des Leseverstehens ($R$) im Rahmen von Tests zur Überprüfung der Simple View of Reading. Erstens sollten die eingesetzten Messmethoden so gestaltet sein, dass sie die relevanten Teilprozesse auf der Wort-, Satz- und Textebene auch tatsächlich abbilden. Damit sind z.B. Methoden, die etwa nur mit einzelnen Sätzen als Testmaterial arbeiten (z.B. Braze et al., 2007), nicht geeignet, weil satzübergreifende Prozesse der Kohärenzbildung nicht abgedeckt werden. Zweitens besteht eine noch wichtigere Anforderung darin, dass Hör- und Leseverstehen über strikt parallele Testverfahren erhoben werden (vgl. z.B. Hoover & Gough, 1990, S. 131). Sowohl $C$ als auch $R$ liegt laut der Simple View of Reading die gleiche allgemeine Fähigkeit des Sprachverstehens zugrunde. Der einzige Unterschied ist das Medium, welches das linguistische Material transportiert, was der Theorie zufolge unterschiedliche Teilprozesse bei der Worterkennung, aber nicht auf darüber hinausgehenden Prozessebenen bedingt. Mögliche modalitätsspezifische Verarbeitungsprozesse auf höheren Prozessebenen, die beispielsweise
durch Textsortenwissen oder Interpunktion angeregt werden und das Textverstehen zusätzlich beeinflussen könnten, werden in der Simple View of Reading nicht berücksichtigt.


**Ist die Simple View of Reading sprachspezifisch?**

Ein letzter, aber gleichwohl für die Reichweite der Simple View of Reading wichtiger Punkt betrifft die Sprachspezifität des Modells. Sämtliche der uns bekannten Untersuchungen zur Überprüfung der Simple View of Reading wurden in alphabetischen Schriftsystemen durchgeführt (für Überlegungen zur Geltung der Theorie in nicht-alphabetischen Schriftsystemen s. Gough, 1996). Die überwiegende Mehrzahl der bislang durchgeführten Untersuchungen wurde mit englischsprachigen Probanden/-innen und Materialien durchgeführt (z.B. Braze et al., 2007; Georgiou et al., 2009; Hoover & Gough, 1990; Johnston & Kirby, 2006; Joshi & Aaron, 2000; Kendeou et al., 2009; Tunmer & Chapman, 2012;


**Ziele der aktuellen Untersuchung**

Die vorliegende Untersuchung verfolgte das generelle Ziel einer Überprüfung der Vorhersagen der Simple View of Reading für eine Stichprobe deutschsprachiger Grundschulkinder. In der einzigen uns bekannten Studie zur Simple View of Reading im deutschsprachigen Raum (Marx & Jungmann, 2000) wurde die Annahme einer
multiplikativen Verknüpfung von visueller Worterkennung und Hörverstehen keiner Prüfung unterzogen. Sofern die Simple View of Reading uneingeschränkt sprachübergreifend verallgemeinerbar ist, sollten wir zeigen können, dass ein beträchtlicher Teil der Varianz individueller Unterschiede im Leseverstehen anhand der Fähigkeiten der Kinder in der visuellen Worterkennung und im Hörverstehen aufgeklärt werden kann. Darüber hinaus sollten die Beiträge der visuellen Worterkennung und des Hörverstehens zum Leseverstehen nicht strikt additiv sein. Vielmehr impliziert die Annahme der Simple View of Reading, dass sich Leseverstehen als das Produkt der beiden Variablen darstellen lässt, eine Interaktion von visueller Worterkennung und Hörverstehen, die über die Haupteffekte der beiden Variablen hinaus einen Beitrag zur Varianzaufklärung im Leseverstehen leisten sollte.

Bei der Durchführung der Untersuchung wurde ein besonderes Augenmerk auf die Einhaltung messmethodischer Standards gelegt, die aus unserer Sicht für eine stringente Überprüfung der Simple View of Reading unerlässlich sind. So wurden zur Erfassung von Fähigkeiten des visuellen Worterkennens (D) psychometrisch gut erprobte lexikalische Entscheidungsaufgaben eingesetzt, die sowohl die direkte als auch die indirekte Route der visuellen Worterkennung abdecken (vgl. Richter et al., 2012). Als Lese- und Hörverstehenstests wurden Textverifikationsaufgaben verwendet, die im Hinblick auf die Aufgabenstellung und die psycholinguistisch relevanten sprachlichen Merkmale der Testitems strikt parallelisiert waren.

Die Aufgaben und Testitems waren so gestaltet, dass zu ihrer Beantwortung die wichtigsten Typen verstehensrelevanter kognitiver Prozesse gefordert waren (von Worterkennung über semantische und syntaktische Integration bis hin zur Herstellung satzübergreifender Sinnzusammenhänge), aber keine verstehensfremden kognitiven Leistungen in Anspruch genommen wurden. Bei allen Testitems wurden sprachliche Merkmale variiert, die aus kognitionspsychologischer Sicht die angezielten kognitiven Prozesse erleichtern oder erschweren sollten. Schließlich wurden im Rahmen einer computergestützten Vorgabe für jede der drei Aufgaben sowohl Reaktionszeiten als auch die Antwortrichtigkeit (Akkuratheit) gemessen, um mit dem Routinisierungsgrad und der Zuverlässigkeit beide Aspekte der Effizienz kognitiver Teilprozesse des Leseverstehens erfassen zu können.
Methode

Design und Stichprobe

Es handelte sich bei der Untersuchung um eine querschnittliche korrelative Untersuchung, an der insgesamt 124 Grundschüler/-innen der dritten und vierten Klassenstufe aus Köln und Frankfurt am Main teilnahmen. Davon besuchten 50 Kinder die dritte Klasse (18 Mädchen und 31 Jungen aus sechs Klassen) und 74 Kinder die vierte Klasse (40 Mädchen und 34 Jungen aus fünf Klassen). Die Kinder der dritten Klassen waren im Durchschnitt 9.42 Jahre (SD=0.45) und die Kinder der vierten Klasse 10.41 Jahre (SD=0.37) alt. Insgesamt 92 Kinder waren deutsche Muttersprachler/-innen (34 Kinder der dritten und 58 Kinder der vierten Klasse). Fünfzehn Kinder wuchsen mit einer anderen Muttersprache oder bilingual mit Deutsch als einer der beiden Muttersprachen auf (7 Kinder der dritten Klasse und 8 Kinder der vierten). Von 17 Kindern fehlte diese Information. Alle Kinder nahmen freiwillig an der Untersuchung teil. Vor der Teilnahme wurde die schriftliche Erlaubnis der Eltern eingeholt.

Ablauf


Die Daten wurden im Rahmen einer größeren Querschnittsuntersuchung mit weiteren Maßen des Lese- und Hörverstehens erhoben (Richter et al., 2012; Richter et al., 2011). Die Testungen wurden an zwei unterschiedlichen Schultagen durchgeführt und beanspruchten an jedem Untersuchungstag maximal eine Unterrichtsstunde (45 Minuten).
Erhobene Variablen

Visuelle Worterkennung (Dekodieren). Um die Dekodierfähigkeit der Kinder zu erfassen, verwendeten wir eine lexikalische Entscheidungsaufgabe mit 94 Testitems (46 Wörter, 46 Pseudowörter sowie zwei Übungsitems). Für jedes Testitem musste das Kind entscheiden, ob es sich um ein ihm bekanntes Wort handelte oder nicht. Dabei wurden als Indikatoren der Effizienz von Worterkennungsprozessen sowohl die Reaktionszeit in logarithmierter Form (Routinisierungsgrad) als auch die Akkuratheit (Zuverlässigkeit) erfasst und per Mittelwertsbildung über alle Items zu Testwerten zusammengefasst. Die durchschnittliche Länge der Wörter betrug 5.68 Zeichen (SD=1.08; Min=3; Max=10), die der Pseudowörter 6.09 Zeichen (SD=2.02; Min=3; Max=12). Die logarithmierte Frequenz der Wörter varierte von 0.00 bis 3.77 mit einer mittleren Frequenz von 1.81 (SD=1.03; Mannheim-Korpus der CELEX-Datenbank für geschriebene Sprache; Baayen et al., 1995). Die Ausgangswörter, auf deren Grundlage die Pseudowörter konstruiert wurden, waren im Hinblick auf die logarithmierte Wortfrequenz mit den Wortstimuli parallelisiert (M=1.66; SD=0.97). Die Pseudowörter entsprachen den deutschen phonologischen Regeln, unterschieden sich jedoch hinsichtlich ihrer Nähe zu tatsächlichen Wörtern des Deutschen. Die Pseudowörter mit einer hohen Wortähnlichkeit wurden auf der Grundlage eines regulär gebildeten deutschen Wortes gebildet, dessen Wortonset geändert wurde. Die wortähnlichen Pseudowörter entstanden durch das Vertauschen mindestens einer Silbe eines deutschen aber unregelmäßig gebildeten Wortes. Eine dritte Gruppe von Pseudowörtern waren Pseudohomophone, die lautlich mit tatsächlichen deutschen Wörtern übereinstimmten, sich jedoch orthographisch von ihnen unterschieden. Die Stimuli wurden einzeln und nacheinander in randomisierter Reihenfolge dargeboten.

Hörverstehen. Für die Erfassung der Fähigkeiten des Hörverstehens verwendeten wir eine Textverifikationsaufgabe, bei der die Kinder zwei aufeinander folgende Sätze hörten. Ihre Aufgabe war es, zu beurteilen, ob die präsentierten Sätze eines Paares zusammenpassten oder nicht. Damit zielt der Test auf die Erfassung der Effizienz von Kohärenzbildungsprozessen und Prozessen der syntaktischen und semantischen Integration ab, die für das Hörverstehen wesentlich sind. Auch bei diesem Test wurden zwei effizienzbezogene Testwerte ermittelt, die mittlere Reaktionszeit und die mittlere Antwortrichtigkeit. Insgesamt hörten die Kinder 72 Satzpaare (35 kohärente und 35 nicht kohärente Satzfolgen sowie 2 Übungspaare). Die Satzpaare hatten eine durchschnittliche Länge von 58.44 Zeichen (SD=10.59; Min=31; Max=99) bei einer mittleren Wortzahl von 9.61 Wörtern (SD=1.81; Min=5; Max=15) und eine durchschnittliche logarithmierte Frequenz...

Leseverstehen. Für die Erfassung der Fähigkeiten des Leseverstehens verwendeten wir eine Textverifikationsaufgabe, deren Struktur und Testitems (35 kohärente und 35 nicht kohärente Satzfolgen sowie 2 Übungspaare) strikt parallel zur Hörverstehensaufgabe gehalten waren. Wiederum mussten die Kinder nach jedem Satzpaar entscheiden, ob die beiden Sätze zusammenpassten oder nicht, und die mittlere logarithmierte Reaktionszeit und die mittlere Antwortrichtigkeit wurden erfasst. Der einzige Unterschied bestand darin, dass die Satzpaare visuell dargeboten wurden. Zunächst wurde der erste Satz eines Satzpaares in der Bildschirmmitte präsentiert. Nachdem die Kinder den ersten Satz gelesen hatten, drückten sie eine blau markierte Taste auf der Laptop-Tastatur, woraufhin der zweite Satz des Satzpaares direkt unter dem ersten Satz erschien. Die Satzpaare hatten eine durchschnittliche Länge von 58.58 Zeichen (SD=10.68; Min=35; Max=96) mit einer durchschnittlichen Wortzahl von 9.54 Wörtern (SD=2.03; Min=6; Max=17) und einer durchschnittlichen logarithmierten Frequenz der Inhaltsworte von 2.01 (SD=0.42; Min=0.56; Max=2.90; Mannheim-Korpus der CELEX Datenbank für geschriebene Sprache). Die Gestaltung der Testitems erfolgte nach den gleichen Kriterien wie die Gestaltung der Testitems im Hörverstehentest.
Ergebnisse

Um die Annahmen der Simple View of Reading und die Art des Zusammenspiels von Fähigkeiten der visuellen Worterkennung und des Hörverstehens bei der Prädiktion des Leseverstehens zu prüfen, wurden Regressionsmodelle mit drei verschiedenen Kriteriumsvariablen und entsprechend gebildeten Prädiktoren geschätzt: (1) mit den logarithmierten Reaktionszeiten des Leseverstehenstests als Maß des individuellen Routinisierungsgrads der beteiligten kognitiven Teilprozesse, (2) mit dem mittleren Akkuratheitswert als Maß für die individuelle Zuverlässigkeit dieser Prozesse und (3) mit einem integrierten Testwert, der als Quotient aus Akkuratheit und logarithmierter Reaktionszeit gebildet wurde (mittlere Akkuratheit geteilt durch mittlere logarithmierte Reaktionszeit). Dieser integrierte Testwert lässt sich als zusammenfassender Indikator der Effizienz der kognitiven Prozesse auffassen, die an der jeweiligen Aufgabe beteiligt sind (vgl. Hale et al., 2011; Richter & van Holt, 2005). Deskriptiv-statistische Kennwerte und Interkorrelationen aller Variablen sind in Tabelle 3.1 aufgeführt.


Reaktionszeiten. Die Ergebnisse für die logarithmierten Reaktionszeiten zeigten in allen drei Schritten der Regressionsanalyse signifikante Haupteffekte für die Prädiktoren Worterkennung und Hörverstehen. Die additive Kombination der Prädiktoren in Modell 1 erklärte dabei 28% der Varianz in den Reaktionszeiten der Schüler/-innen. Die in den Modellen 2 und 3 einbezogenen Interaktionsterme, einschließlich des Produkts von
Worterkennung und Hörverstehen, erreichten kein signifikantes Niveau und trugen nicht signifikant zur Verbesserung der Varianzaufklärung bei.

**Akkuratheit.** Für die Akkuratheit fanden sich vergleichbare Ergebnisse. Die Prädiktoren Worterkennung und Hörverstehen erklärten in Modell 1 etwa 38% der Varianz in der Akkuratheit im Leseverständenstest. Das Produkt aus Worterkennung und Hörverstehen war nicht signifikant und trug nicht zur Verbesserung der Varianzaufklärung bei. Dasselbe gilt für die übrigen Interaktionsterme im Modell.

**Integrierte Testwerte.** Für die integrierten Testwerte als Kriteriumsvariable fanden sich in den Modellen 1 und 2 ebenfalls signifikante Haupeffekte für Worterkennung und Hörverstehen mit einer Varianzaufklärung von 35%. Auch hier konnte das Produkt aus beiden Fähigkeiten in Modell 2 keinen signifikanten Beitrag zur Varianzaufklärung leisten. Im Gegensatz zu den Reaktionszeit- und Akkuratheitsmodellen zeigte sich jedoch in Modell 3 ein signifikanter Effekt der Dreifachinteraktion von Worterkennung, Hörverstehen und Klassenstufe, während der Prädiktor Worterkennung allein nicht mehr signifikant wurde. Durch die zusätzlichen Prädiktoren verbesserte sich die Varianzaufklärung um 2.7% (s. Tabelle 3.2).

Tabelle 3.1:

*Study 2*

**Deskriptive Statistiken und Interkorrelationen aller Variablen**

<table>
<thead>
<tr>
<th>M/Schule</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Klasse</td>
<td>0.19</td>
<td>0.99</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Worterkenntnis Reaktionszeit</td>
<td>0.00</td>
<td>1.00</td>
<td>-0.35*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Worterkenntnis Akkuratheit*</td>
<td>0.00</td>
<td>1.00</td>
<td>0.32*</td>
<td>0.71</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Worterkenntnis integriert*</td>
<td>0.00</td>
<td>1.00</td>
<td>0.45*</td>
<td>-0.31*</td>
<td>0.93*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Höverstehn Reaktionszeit*</td>
<td>0.00</td>
<td>1.00</td>
<td>0.06</td>
<td>0.22*</td>
<td>0.24*</td>
<td>0.14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Höverstehn Akkuratheit*</td>
<td>0.00</td>
<td>1.00</td>
<td>0.26*</td>
<td>0.02</td>
<td>0.51*</td>
<td>0.47*</td>
<td>0.32*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Höverstehn integriert*</td>
<td>0.00</td>
<td>1.00</td>
<td>0.26*</td>
<td>-0.02</td>
<td>0.49*</td>
<td>0.47*</td>
<td>0.15</td>
<td>0.99*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Leseverstehn Reaktionszeit</td>
<td>7.76</td>
<td>0.54</td>
<td>-0.01</td>
<td>0.43*</td>
<td>0.44*</td>
<td>0.25*</td>
<td>0.37*</td>
<td>0.31*</td>
<td>0.25*</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Leseverstehn Akkuratheit</td>
<td>0.74</td>
<td>0.18</td>
<td>0.22*</td>
<td>-0.01</td>
<td>0.49*</td>
<td>0.47*</td>
<td>0.12</td>
<td>0.57*</td>
<td>0.58*</td>
<td>0.41*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10 Leseverstehn integriert</td>
<td>0.10</td>
<td>0.02</td>
<td>0.25*</td>
<td>-0.14</td>
<td>0.41*</td>
<td>0.44*</td>
<td>0.03</td>
<td>0.53*</td>
<td>0.55*</td>
<td>0.17</td>
<td>0.97*</td>
<td>1</td>
</tr>
</tbody>
</table>

### Tabelle 3.2:

**Parameterschätzungen der geschachtelten Regressionsmodelle für mittlere logarithmierte Reaktionszeiten, Akkuratethitswerte und integrierte Testwerte im Leseverstehen als Kriteriumsvariablen**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Reaktionszeit</th>
<th>Akkuratheit</th>
<th>Integrierte Testwerte</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Modell 1</td>
<td>Modell 2</td>
<td>Modell 3</td>
</tr>
<tr>
<td>Konstante</td>
<td>7.752*</td>
<td>7.755*</td>
<td>7.768*</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.044)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>Klassenstufe</td>
<td>0.068</td>
<td>0.069</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.046)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>Worterkenntnis</td>
<td>0.224*</td>
<td>0.225*</td>
<td>0.212*</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.046)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>Hörverstehen</td>
<td>0.146*</td>
<td>0.139*</td>
<td>0.144*</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.050)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>Worterkenntnis X Hörverstehen</td>
<td>-0.015</td>
<td>0.000</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.056)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Klassenstufe X Worterkenntnis</td>
<td>0.046</td>
<td>0.046</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.048)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Klassenstufe X Hörverstehen</td>
<td>0.012</td>
<td>0.012</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.054)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Klassenstufe X Worterkenntnis X Hörverstehen</td>
<td>-0.027</td>
<td>0.023</td>
<td>0.004*</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.056)</td>
<td>(0.016)</td>
</tr>
</tbody>
</table>


*p < .05 (einsseitig).


Damit stützen die Ergebnisse insgesamt die aus der Simple View of Reading ableitbare Annahme, dass sowohl effiziente, d.h. gut routinierte und zuverlässige, visuelle Worterkennungsprozesse als auch effiziente Prozesse des allgemeinen Sprachverständnisses für ein gutes Leseverständnis wichtig sind. Für die spezifische Annahme, dass die Effizienz der visuellen Worterkennung und des allgemeinen Sprachverständnisses multiplikativ
miteinander verknüpft sind, was ihre Beziehung zum Leseverstehen angeht, haben sich dagegen kaum Belege ergeben. Lediglich für den zusammenfassenden Effizienzindikator, der aus Akkuratheit und Reaktionszeit zusammengesetzt war, ergab sich ein schwacher Hinweis darauf, dass in der vierten (nicht aber der dritten) Klasse eine multiplikative Beziehung vorliegen könnte, die den Vorhersagen der Simple View of Reading entspricht.

**Diskussion**


Eine zweite bemerkenswerte Abweichung gegenüber den meisten der bislang durchgeführten Untersuchungen zur Simple View of Reading besteht darin, dass in der vorliegenden Untersuchung durch individuelle Unterschiede in Worterkennung und Hörverstehen zwar rund ein Drittel der Varianz im Leseverstehen aufgeklärt werden kann, dieser Anteil erklärter Varianz aber wesentlich geringer ist als in vielen der bisherigen Untersuchungen zur Simple View of Reading, in welchen zum Teil 75% Varianzaufklärung und mehr erreicht werden (z.B. Braze et al., 2007; Hoover & Gough, 1990; Johnston & Kirby, 2006; Tunmer & Chapman, 2012).


dass der Beitrag des Produktes aus visueller Worterkennung und Hörverstehen zum Leseverstehen bzw. die Höhe der durch \( D \) und \( C \) aufgeklärten Varianz in Abhängigkeit vom Alter variiert. In diesem Fall müsste die Simple View of Reading um eine entsprechende Entwicklungskomponente erweitert werden.

Chapter IV

Study 3
Processing of Positive-causal and Negative-causal Coherence Relations in Primary School Children and Adults: A Test of the Cumulative Cognitive Complexity Approach in German

A version of this chapter is published as:
DOI: 10.1017/S0305000915000872
Copyright: © Cambridge University Press 2016
Abstract. Establishing local coherence relations is central to text comprehension. Positive-causal coherence relations link a cause and its consequence, whereas negative-causal coherence relations add a contrastive meaning (negation) to the causal link. According to the cumulative cognitive complexity approach, negative-causal coherence relations are cognitively more complex than positive-causal ones. Therefore, they require greater cognitive effort during text comprehension and are acquired later in language development. The present cross-sectional study tested these predictions for German primary school children from Grades 1 to 4 and adults in reading and listening comprehension. Accuracy data in a semantic verification task support the predictions of the cumulative cognitive complexity approach. Negative-causal coherence relations are cognitively more demanding than positive-causal ones. Moreover, our findings indicate that children’s comprehension of negative-causal coherence relations continues to develop throughout the course of primary school. Findings are discussed with respect to the generalizability of the cumulative cognitive complexity approach to German.

Keywords: Coherence relation, cumulative cognitive complexity approach, text comprehension

Introduction

Text comprehension may be regarded as building a coherent mental representation of the text and its contents. Accordingly, the ability to establish local and global coherence, that is, linking the content of adjacent and distant statements of a text, is central to text
comprehension (Graesser, Millis, & Zwaan, 1997; Van Dijk & Kintsch, 1983). Two of the most intensely investigated coherence relations are positive-causal coherence relations (4.1), which link a cause and its consequence, and negative-causal coherence relations (4.2), which add a contrastive meaning or a negation to the causal link (Sanders et al., 1992):

(4.1) Tom ate too many cherries. Therefore, he now has a stomachache.

(4.2) Anna is ill. Nevertheless, she goes to school.

According to the cumulative cognitive complexity approach (Evers-Vermeul & Sanders, 2009; Spooren & Sanders, 2008), positive-causal and negative-causal relations differ in internal complexity and, therefore, vary in the cognitive effort they require during text comprehension (a detailed explanation is provided in the next sections). Furthermore, the order of acquisition of coherence relations is also assumed to depend on their internal complexity. Less complex relations, such as positive-causal coherence relations, are assumed to be acquired earlier than negative-causal coherence relations. Studies investigating the processing of coherence relations on English- and Dutch-speaking children and adults report findings that are consistent with the cumulative cognitive complexity approach (e.g., Bloom et al., 1980; Evers-Vermeul & Sanders, 2009; Goldman & Murray, 1992; Spooren & Sanders, 2008).

The present study addressed three research questions about the scope and explanatory power of the cumulative complexity approach. First, most of the existing studies have been conducted with English- and Dutch-speaking participants, but little is known about the processing of these coherence relations in German. Second, the extent that the processing of coherence relations and its development run in parallel in listening and reading comprehension (as predicted, for example, by the simple view of reading, Gough & Tunmer, 1986; Hoover & Gough, 1990) is still unknown. To our knowledge, no study to date has employed the method of parallel tasks and materials to investigate the processing of positive-causal and negative-causal coherence relations in both written and spoken text comprehension. Finally, although several studies demonstrated early usage of positive-additive, positive-causal, and negative-additive coherence relations and connectives in preschoolers’ speech production (e.g., Bloom et al., 1980; Van Veen, Evers-Vermeul, Sanders, & Van den Bergh, 2009, 2013), findings by Cain et al. (2005) suggest that comprehension of connectives and coherence relations still develops during primary school. According to the cumulative cognitive complexity approach, this might hold, in particular, for
the most complex type of coherence relations, namely negative-causal coherence relations which rarely occur in preschoolers’ spontaneous speech production.

Against this background, the present study (a) investigated the processing of positive-causal and negative-causal coherence relations in German primary school children and adults, (b) directly compared the processing of positive-causal and negative-causal coherence relations in written and spoken language, and (c) examined developmental changes as indicated by age differences in the processing of these coherence relations during the first four years of primary school. In the following section, we will outline the core assumptions of the cumulative cognitive complexity approach and discuss previous studies testing its assumptions. The cumulative cognitive complexity approach and questions raised by existing research motivated the predictions tested in the present study, whose rationale and findings will be presented afterwards.

Positive-causal and Negative-causal Coherence Relations

Coherence relations are cognitive meaning relations that link the mental representations of adjacent and distant statements of a text by adding a distinct aspect of meaning that is ‘more than the sum of the parts’ (Sanders et al., 1992, p. 2; see also Sanders & Noordman, 2000). According to Sanders et al., coherence relations can be characterized along four dimensions. The first dimension, termed basic operation, refers to the strength with which two statements are connected. They can be connected either strongly (causal) or weakly (additive). The second dimension distinguishes between a semantic and a pragmatic source of coherence, i.e. statements can be coherent with respect to their propositional content (semantic) or to the illocutionary meaning of the speech act (pragmatic). The third dimension refers to the order of the two statements for causal relations, which are the basic order (cause-consequence) or the non-basic order (consequence-cause). No such distinction is made for additive relations. The final dimension is polarity, characterizing coherence relations as either positive (containing no negation or contrast) or negative (adding a negative or contrastive meaning to the additivity or causality of the coherence relation).

One of the most intensely investigated types of coherence relations are causal links between statements. According to the taxonomy of Sanders et al. (1992), these coherence relations can be further distinguished along the polarity dimension into positive-causal and negative-causal relations. Positive-causal relations, such as expressed in (4.1), link two statements, one of which designates the cause of the effect described in the other one. In contrast, negative-causal relations, such as the relation expressed in (4.2), link two statements,
These coherence relations can be (but need not be) signaled by means of positive-causal connectives such as *because* or *therefore* or negative-causal connectives such as *although* or *nevertheless*. Connectives can function as ‘processing instructions’ (Kamalski et al., 2008, p. 324), which guide the reader’s expectations on how to integrate the upcoming statement into a coherent mental representation (Murray, 1995). Several studies underline the essential role connectives play in text comprehension. Connectives were found to decrease processing times, enhance text memory, and facilitate on-line integration of information and inference generation (e.g., Degand, Lefèvre, & Bestgen, 1999; Maury & Teisseranc, 2005; Millis, Golding, & Barker, 1995; Millis & Just, 1994; Sanders, Land, & Mulder, 2007; Sanders & Noordman, 2000; Van Silfhout, Evers-Vermeul, Mak, & Sanders, 2014), especially in readers with little prior knowledge of the text content (e.g., Kamalski et al., 2008).

**The Cumulative Cognitive Complexity Approach**

Connectives and the coherence relations signaled by them are assumed to differ in the cognitive processing effort they require during text comprehension. One prominent account of the cognitive processing of coherence relations is the cumulative cognitive complexity approach (Evers-Vermeul & Sanders, 2009; Spooren & Sanders, 2008). Based on the taxonomy of Sanders et al. (1992), the cumulative cognitive complexity approach assumes that coherence relations differ in their internal cognitive complexity depending on the dimensions of their characterization. Along the basic operation dimension, causal relations can be regarded as cognitively more complex than additive relations, because they involve a cause-consequence link in addition to the additive combination of the statements (Spooren & Sanders, 2008). In other words, causal relations are specified with respect to causality, whereas additive relations are underspecified with regard to this feature (Evers-Vermeul & Sanders, 2009). Likewise, along the dimension of polarity, negative relations are assumed to be cognitively more complex than positive relations, because they add a contrastive component to the relation (Spooren & Sanders, 2008). This characterization can also be extended to the combination of the two dimensions, resulting in a complexity hierarchy of coherence relations (Evers-Vermeul & Sanders, 2009; Spooren & Sanders, 2008): Positive-additive coherence relations should be cognitively the least complex, because they do not involve a causal link or a contrast. Positive-causal and negative-additive coherence relations
are assumed to be more complex than positive-additive coherence relations, because they involve either a causal link (positive-causal relations) or a contrastive link (negative-additive relations). Finally, negative-causal relations should be the most complex, because they involve both a causal link and a contrast. Based on this hierarchy, Evers-Vermeul and Sanders (2009) and Spooren and Sanders (2008) assumed that the cognitive processing of coherence relations is more demanding for complex compared to less complex relations and that complex relations are acquired later than less complex ones. Accordingly, positive-additive coherence relations should be easier to process than positive-causal and negative-additive coherence relations, and these, in turn, should be easier to process than negative-causal coherence relations. In addition, positive-additive coherence relations should be acquired first and negative-causal coherence relations last, with positive-causal and negative-additive coherence relations between the two.

Several studies with English- and Dutch-speaking children and adults report findings that appear to support the cumulative cognitive complexity approach. In a self-paced reading study, Millis and Just (1994, Exp. 4) found that English readers provided less accurate responses to comprehension questions and responded more slowly when reading negative-causal sentence pairs linked by although compared to positive-causal sentence pairs linked by because. Goldman and Murray (1992) found that English students exhibited more difficulties in appropriately filling in negative connectives into blank slots between sentences compared to positive-additive and positive-causal connectives (but see also Sanders & Noordman, 2000, who found faster processing times for causal compared to additive relations).

With respect to the acquisition order of coherence relations, several studies (most of them focusing on language production) provided evidence that English-speaking pre-school children (e.g., Bloom et al., 1980) and primary-school children (e.g., Katz & Brent, 1968; Wing & Scholnick, 1981) use additive relations and connectives before causal ones and positive relations and connectives before negative ones. Furthermore, their knowledge of coherence relations (negative coherence relations in particular) appears to be still developing during the primary school years (e.g., Cain & Nash, 2011; Katz & Brent, 1968; Wing & Scholnick, 1981). In Dutch, Evers-Vermeul and Sanders (2009) analyzed spontaneous language-production data of 12 children aged 1;5 to 5;6 to explicitly test the cumulative cognitive complexity approach. As expected, they found that additive relations are acquired before causal relations and that positive relations are acquired before negative relations. Spooren and Sanders (2008) obtained comparable results when investigating Dutch primary school children’s production and comprehension of coherence relations.
Open Question 1: Cross-linguistic Generalization of Cumulative Cognitive Complexity to German

Although the studies discussed in the preceding section lend support to the cumulative cognitive complexity approach, open questions remain regarding its generalizability and its explanatory power. First, the cumulative cognitive complexity approach appears to explain processing differences and acquisition orders of coherence relations and connectives in English and Dutch speakers, but it still lacks cross-linguistic validation beyond these two languages. Sanders et al.’s (1992) taxonomy of cognitive coherence relations can be applied to German, and the semantics of the connectives signaling these coherence relations in German are similar to those in English and Dutch. Thus, the cumulative cognitive complexity approach is also expected to apply to German-speaking children. However, the existing studies with this population provide only surprisingly limited support for the approach.

Kail and Weissenborn (1984), for example, investigated the acquisition order of the substitutive and the contrastive use of the adversative *aber/sondern* – ‘but’ in German eight-to ten-year-olds and *mais* – ‘but’ in French children of the same age. They found that the substitutive use of these connectives (4.3) is acquired earlier than the contrastive use (4.4).

(4.3)  Kathy’s father didn’t sell the old car but he gave it to Kathy.
(4.4)  Joe is an Indian but he didn’t win the rodeo.

(examples from Kail & Weissenborn, 1984, pp. 147-148)

They argued that the contrastive use is acquired later, because the comprehension of contrastive links requires more complex inferences. To comprehend a sentence containing a contrastive adversative (4.4) one must infer that Indians usually win rodeos, whereas no such inference is necessary to comprehend sentences containing a substitutive adversative. The substitutive use of adversatives is associated with a negative-additive relation, whereas the contrastive use is associated with a negative-causal relation. These findings by Kail and Weissenborn (1984) support the cumulative cognitive complexity approach, but only for effects within the category of negative coherence relations.

Additional but rather weak support for the cumulative cognitive complexity approach can be found in an eye-tracking study by Köhne and Demberg (2013) who investigated the function of positive-causal and negative-causal connectives to restrict possible discourse referents during on-line text processing in German adults. In Experiment 1, based on the visual world paradigm, participants listened to short spoken texts that expressed either a
positive-causal or a negative-causal coherence relation signaled by their respective connectives. At the same time, participants viewed a visual scene with one target object matching the final word in the text and three distractor objects. Köhne and Demberg found that as soon as the connective occurred participants focused significantly less often on those objects that could be excluded based on the meaning of the connective. However, they recognized the target objects more slowly in negative-causal compared to positive-causal sentence pairs and responded less accurately to comprehension questions following negative-causal sentence pairs. In line with the assumptions of the cumulative cognitive complexity approach, the authors suggested that processing negative-causal coherence relations is cognitively more demanding than processing positive-causal coherence relations. However, in contrast to Experiment 1, the results of a second eye-tracking experiment with written texts (Köhne & Demberg, Exp. 2) failed to support the cumulative cognitive complexity approach. No significant reading time differences were found, no increase in response accuracy to comprehension questions, and no decrease in reaction times for positive-causal compared to negative-causal coherence relations.

Finally, Dragon, Berendes, Weinert, Heppt, and Stanat (2015) found that German second and third graders showed considerable difficulties in comprehending negative-causal sentence pairs compared to positive-causal sentence pairs. They presented the children with coherent sentence pairs containing appropriately used positive-causal and negative-causal connectives or with sentence pairs that had been made incoherent by exchanging positive-causal and negative-causal connectives. When asked to indicate whether the sentence pairs made sense, children systematically rejected coherent negative-causal sentence pairs and accepted incoherent sentence pairs. Dragon et al. suggested that most children ignored the negative-causal connectives (or even all connectives) and judged the sentence pairs only on the basis of their semantic and situational compatibility independent of the connective meaning. These findings indicate that German second and third graders have problems comprehending negative-causal connectives. This interpretation is in line with the assumption implied by the cumulative cognitive complexity approach that an adequate understanding of the cognitively complex negative-causal coherence relations and connectives is mastered comparatively late in language development. However, the findings by Dragon et al. are restricted to oral language processing and thus cannot be generalized to written language processing. Furthermore, only accuracy of the sensibility judgments was recorded but not response latency. Longer response latencies for negative-causal compared to positive-causal
sentence pairs might lend further support to the assumptions of the cumulative cognitive complexity approach.

Open Question 2: Processing of Coherence Relations in Reading vs. Listening Comprehension

The existing relevant literature also lacks evidence on the extent that coherence relations and their respective connectives are processed similarly in listening and reading comprehension and on whether the ability to comprehend positive-causal and negative-causal coherence relations develops in parallel or at a different pace during elementary school years. According to the simple view of reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), reading comprehension \( R \) can be defined as the product of two cognitive component skills, namely decoding \( D \) and general linguistic (or listening) comprehension \( C \). The formula \( R = D \times C \) implies that the cognitive processes of text comprehension should be basically the same for listening and reading comprehension after written word forms have been decoded. On the basis of this account, similar findings should be expected for the processing of positive-causal and negative-causal coherence relations in written and spoken language processing. However, children might conceivably be able to comprehend complex coherence relations (such as negative-causal ones) at a later age when they read a text compared to when they listen to it. Young readers in the primary school years need to spend a considerable amount of cognitive resources on lower-level reading processes such as decoding (e.g., Knoepke & Richter, in press; Perfetti, 1985). Consequently, they might lack the cognitive resources required to comprehend cognitively complex coherence relations.

To our knowledge, no study to date has examined the comprehension of positive-causal and negative-causal coherence relations in both spoken and written language processing using the method of parallel tasks and materials. The existing studies used either written (e.g., Cain et al., 2005; Cain & Nash, 2011) or spoken text materials (e.g., Kail & Weissenborn, 1984; Katz & Brent, 1968). Few studies used both written and spoken text materials, but none directly compared performance in both modalities. For example, Köhne and Demberg (2013) used spoken text materials in their first and written materials in their second eye-tracking experiment. However, the tasks were different in both experiments. In Experiment 1, participants were required to identify a target object in a visual scene, whereas in Experiment 2, they only read texts. Thus, the results in both experiments are not directly comparable.
Open Question 3: Developmental Changes in the Processing of Coherence Relations During the Primary School Years

A third question that has been fairly neglected by previous research is how the processing of positive-causal and negative-causal coherence relations and connectives develops throughout years of primary school. Many studies that have investigated the acquisition order of coherence relations and connectives focused on the first time that specific connectives occur in preschooler’s speech production (e.g., Bloom et al., 1980; Evers-Vermeul & Sanders, 2009; Van Veen et al., 2009, 2013). Although they found adequate production of positive-additive, positive-causal, and negative-additive connectives and coherence relations, Cain et al. (2005) argued that comprehension of connectives and coherence relations might still be incomplete and developing during years of primary school. According to the cumulative cognitive complexity approach, this might hold, in particular, for the most complex coherence relations, namely negative-causal coherence relations, which occur rarely in preschoolers’ spontaneous speech production (e.g., Katz & Brent, 1968; Spooren & Sanders, 2008).

The Present Study

The present study pursued three related aims. The first aim was to replicate the findings from previous studies that the processing of negative-causal coherence relations requires more cognitive effort than the processing of positive-causal coherence relations in German primary school children and adults. The second aim was to directly compare via parallel tasks and experimental materials the processing of these coherence relations in written and spoken language comprehension. The third aim was to investigate age-related differences in the processing of positive-causal and negative-causal coherence relations during the first four years of primary school.

To this end, we used a computerized semantic verification task in a visual and auditory version that was similar to the task used by Cain and Nash (2011, Exp. 2 and 3; see also the paradigm used in Murray, 1997, Exp. 2). German primary school children and adults were presented with spoken (all grade levels and adults) and written sentence pairs (Grades 3 and 4 and adults) that contained positive-causal or negative-causal coherence relations signaled by either an appropriate connective resulting in a coherent sentence pair such as (4.5) or an inappropriate connective resulting in an incoherent sentence pair such as (4.6).
The participants’ task was to judge the coherence of the sentence pairs. Mastery of this task requires the comprehension of the coherence relation between both statements, and mastery of negative-causal sentence pairs requires the comprehension of the connective. Response latencies and accuracy were recorded. This paradigm has already been shown to be sensitive to differences in the cognitive processing of coherence relations and to detect developmental changes in children’s ability to establish local coherence (Cain & Nash, 2011, Exp. 2 and 3).

According to the cumulative cognitive complexity approach, we expected that children at all grade levels and adults need more time and make more errors when judging the coherence of the cognitively more complex coherent and incoherent negative-causal sentence pairs compared to the less complex positive-causal sentence pairs.

Two different patterns of outcomes seem possible from the analysis of differences between reading and listening comprehension. The simple view of reading implies that processing of positive-causal and negative-causal coherence relations should not differ between written and auditory text presentation and that the ability to comprehend these relations should develop at the same pace in both modalities. In contrast, theories and findings emphasizing that reading comprehension may be hampered by inefficient word-level processes (e.g., Perfetti, 1985) suggest that younger readers might be able to comprehend negative-causal relations presented auditorily while struggling with the same relations when they are presented in writing. Thus, younger readers might have more difficulties processing the cognitively more complex negative-causal coherence relations in the visual version of the task compared to the auditory version.

Furthermore, based on Cain et al.’s (2005) suggestion that children’s comprehension of connectives and coherence relations might still be incomplete and developing during the primary school years, we expected to find age-related differences in the comprehension of both positive-causal and negative-causal coherence relations in primary school children. Because of their higher cognitive complexity, we particularly expected developmental trends in the processing of negative-causal coherence relations to lag behind developmental trends in the processing of positive-causal coherence relations.
Finally, Cain and Nash (2011, Exp. 3) found longer reading times for sentences preceded by an inappropriate connective (incoherent sentence pair) compared to sentences preceded by an appropriate connective (coherent sentence pair). They suggested that the longer processing times for sentences following an inappropriate connective reflect a laborious attempt to repair the current mental model of the text when inconsistent information cannot be integrated. Based on their findings, we also expected to find shorter response latencies for coherent (containing an appropriate connective) compared to incoherent sentence pairs (containing an inappropriate connective) in our study.

Method

Participants

Participants were 422 German primary school children from Grades 1 to 4, recruited from ten schools (32 classes) in the Frankfurt am Main area, Cologne, and Kassel, and 78 undergraduate students of psychology from the University of Kassel (Germany), who received course credit for their participation.

Children. Two hundred and seventy-seven primary school children participated in the auditory version of the task and 269 children in the visual version of the task. Of these children, 124 from Grades 3 and 4 took part in both the auditory and the visual version. Because of limitations to the time available for testing, 96 children failed to complete the auditory version of the task and 107 children could not complete the visual version of the task. By mischance, the untimely abort of the program led to errors preventing the adequate saving of their data. However, the errors did not result in a systematic exclusion of data. In addition, the data of 38 (auditory version of the task) and 37 (visual version of the task) non-native German speakers were excluded from the analysis. Overall, 143 native German speakers (66 boys, 70 girls, 7 missing gender information) from Grades 1 to 4 completed the auditory version of the task and 125 (72 boys, 53 girls) completed the visual version of the task. Sample characteristics are provided in Table 4.1. Sociodemographic data were collected via a questionnaire completed by the parents. This questionnaire was supplemented by a teacher questionnaire that was used in the event of missing or incomplete parents’ questionnaires. Children only participated in the study when parents had given their written consent.
Table 4.1:

<table>
<thead>
<tr>
<th></th>
<th>Auditory Version</th>
<th>Visual Version</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grade 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>n=35</td>
<td>n=35</td>
</tr>
<tr>
<td>Age</td>
<td>7;5 (0;5)</td>
<td>9;5 (0;5)</td>
</tr>
<tr>
<td>Age</td>
<td>6;9 8;3</td>
<td>8;5 10;7</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Grade 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>n=18</td>
<td>n=18</td>
</tr>
<tr>
<td>Age</td>
<td>8;2 (0;5)</td>
<td>10;4 (0;5)</td>
</tr>
<tr>
<td>Age</td>
<td>7;3 8;11</td>
<td>9;0 11;3</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Grade 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>n=31</td>
<td>n=57</td>
</tr>
<tr>
<td>Age</td>
<td>9;5 (0;5)</td>
<td>10;5 (0;5)</td>
</tr>
<tr>
<td>Age</td>
<td>8;5 10;7</td>
<td>8;5 10;6</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Grade 4</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>n=59</td>
<td>n=68</td>
</tr>
<tr>
<td>Age</td>
<td>10;5 (0;4)</td>
<td>10;4 (0;5)</td>
</tr>
<tr>
<td>Age</td>
<td>9;10 11;4</td>
<td>9;0 11;3</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adults</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>n=60</td>
<td>n=60</td>
</tr>
<tr>
<td>Age</td>
<td>23;0 (5;3)</td>
<td>23;0 (5;3)</td>
</tr>
<tr>
<td>Age</td>
<td>18;5 41;8</td>
<td>18;5 41;8</td>
</tr>
</tbody>
</table>

Adults. All 78 psychology students provided their written consent and completed both the auditory and the visual version of the task. Sociodemographic data were collected via a questionnaire. Seven participants (9.0%) were removed because of missing sociodemographic data and an additional 11 non-native German speakers (14.1%) were removed. The data from the remaining 60 (47 female, 13 male) participants were analyzed.

Materials

We used the ProDi-L visual semantic verification task with sentence pairs (ProDi-L: Prozessbezogene Diagnostik des Leseverstehens bei Grundschulkindern [Process-based assessment of reading skills in primary school children], Richter et al., in press) and its parallel auditory counterpart ProDi-H.

Visual semantic verification task. The ProDi-L visual semantic verification task with sentence pairs contained 70 written experimental sentence pairs and two additional practice sentence pairs of varying length (length in characters for the sentence pair: $M=58.53$; $SD=10.68$; $Min=35$; $Max=96$; length in characters for the second sentence: $M=30.57$; $SD=6.90$; $Min=18$; $Max=52$). Half of the sentence pairs were coherent, the other half incoherent. The participants’ task was to judge the coherence of the sentence pairs (Does the second sentence go with the first one?) as fast and as accurately as possible by pressing a green button for coherent sentence pairs (yes) and a red button for incoherent sentence pairs (no). Of the total of 70 sentence pairs, 30 (29 for the adults) were chosen as critical items.
according to the aims of the present study. These critical sentence pairs contained a positive-causal or a negative-causal coherence relation signaled by either an appropriate connective (coherent sentence pair) or an inappropriate connective (incoherent sentence pair): six coherent positive-causal sentence pairs (4.7), six (five for the adults) incoherent positive-causal sentence pairs (4.8), twelve coherent negative-causal sentence pairs (4.5) (repeated here as (4.9)), and six incoherent negative-causal sentence pairs (4.6) (repeated here as (4.10)).

(4.7) Lena war zu lange in der Sonne. Darum bekam sie einen Sonnenbrand.
     ‘Lena stayed in the sun for too long. Therefore, she got a sunburn.’
(4.8) *Roland hat verschlafen. Darum kommt er pünktlich zur Schule.
     *’Roland overslept. Therefore, he arrived at school on time.’
(4.9) Sandra war nicht müde. Trotzdem ging sie ins Bett.
     ‘Sandra was not tired. Nevertheless, she went to bed.’
(4.10) *Das Wetter war gut. Trotzdem setzte Laura eine Sonnenbrille auf.
     *’The weather was good. Nevertheless, Laura put on her sunglasses.’

Note that the incoherence in the incoherent negative-causal sentence pairs as illustrated in (4.10) was created by linking two actually compatible sentences with a negative-causal connective.

In addition to these critical items, the semantic verification task contained 40 (41 for adults) sentence pairs, which were coherent and positive-additive (4.11), coherent and positive-causal without a connective (4.12), or incoherent sentence pairs containing semantic violations of varying degree (4.13 - 4.15).

(4.11) Felix baut ein Haus mit Legosteinen. Dabei hört er Musik.
     ‘Felix is building a house of Lego bricks. While doing this, he is listening to music.’
(4.12) Klara möchte ein Buch lesen. Sie geht in die Bücherei.
     ‘Klara wants to read a book. She goes to the library.’
(4.13) *Katie geht gerne Schlittschuhlaufen. Dabei trägt er blaue Handschuhe.
     *’Katie loves to go skating. While doing this, he wears blue gloves.’
(4.14) *Sina isst ein Stück Kuchen. Trotzdem fallen die Blätter im Herbst.
     *’Sina is eating a piece of cake. Nevertheless, the leaves fall in autumn.’
The critical sentence pairs contained the positive-causal connectives *deshalb*, *daher*, *darum*, or *denn* – therefore and the negative-causal connective *trotzdem* – nevertheless. Based on their frequencies in the childLex database (Schroeder, Würzner, Heister, Geyken, & Kliegl, 2015), primary school children are likely to be familiar with these connectives in Grades 1 and 2 (frequency per 1 million words for *deshalb*: 300.5; *daher*: 34.8; *darum*: 99.2; *denn*: 889.4; *trotzdem*: 257.6) and Grades 3 and 4 (*deshalb*: 309.1; *daher*: 45.0; *darum*: 150.8; *denn*: 704.9; *trotzdem*: 303.1).

Some of the filler sentence pairs also contained positive-causal and negative-causal connectives. However, these items differed from the critical sentence pairs, because little or no referential overlap exists between the first and the second statements as illustrated in (4.14). Most of them could easily be rejected as incoherent based on the lack of semantic overlap. In contrast, participants needed to consider the local coherence relations to verify the critical sentence pairs and consider the connective to make the correct judgment on negative-causal sentence pairs. Stimulus material and instructions were the same for children and adults.

*Auditory semantic verification task.* The items of the auditory semantic verification task of ProDi-H differed from the items in the visual version but were strictly parallel with respect to construction criteria and length (character length of the whole sentence pair: $M=58.44$; $SD=10.59$; $Min=31$; $Max=99$). Half of the sentence pairs were recorded by a professional male speaker (a radio journalist), the other half by a trained female speaker (an actress).

**Procedure**

Both semantic verification tasks were administered in the context of a larger cross-sectional study. The study investigated processes of listening and reading comprehension with various measures on the word, the sentence, and the text level (Richter et al., 2012; Richter et al., 2013). Children were tested together in classrooms of the participating schools at the end of the school year. Adults were tested in groups of one to six in a laboratory at the University of Kassel Psychology Department. Tasks embedded in a story of an extraterrestrial named Reli who came to earth to learn the earthlings’ language were presented on notebook computers.
Reli walked the participants through the tasks in a manner suitable for children. He introduced the instructions via headphones and gave feedback during the two practice trials preceding the presentation of each task. When participants provided an incorrect answer to one of the practice trials, the practice trials were repeated until the response was correct. In the visual version of the task, the first sentence of the sentence pair appeared in the middle of the notebook screen. When the participants finished reading, they pressed the space bar and then the second sentence appeared below the first one (see Figure 4.1). In the auditory version of the task, participants were presented with the spoken sentence pairs via headphones. Items were presented in a random order. Response accuracy and log-transformed response latencies
Study 3

from presentation onset (of the second sentence in the visual version and of the whole sentence pair in the auditory version) to the pressing of the response button were recorded. The visual semantic verification task was presented first for children and adults who participated in both the visual and the auditory version.

Results

Only critical items were included in the analysis. Responses that were unusually slow or fast (3 SD or more below the item-specific mean and 2 SD or more below or above the person-specific mean after standardizing each item by its item-specific mean) were excluded from the analyses, which resulted in a loss of 4.8% of the children’s data in the visual version of the task and 6.8% in the auditory version of the task for the critical sentence pairs. In the sample of adults, 4.8% of the data were excluded from the analysis in the visual version of the task and 5.2% in the auditory version. Log-transformed response latencies were analyzed using Linear Mixed Models (LMM, Baayen, Davidson, & Bates, 2008) with crossed random effects for items nested within participants and participants nested within items. For the accuracy data, we estimated Generalized Linear Mixed Models (GLMM) with a logit link function (Dixon, 2008). Given that the presentation order of the items was randomized between participants, the results of the mixed-models analyses were assumed to be unbiased despite the small proportion of missing data.

All models were estimated and tested with the software packages lme4 (Bates et al., 2014) and lmerTest for R (Kuznetsova, Brockhoff, & Christensen, 2014). All significance tests were based on a Type I error probability of .05. Separate models were estimated for the auditory and the visual version of the task and for accuracy and log-transformed response latencies as dependent variables. In addition, separate models were estimated for children and adults to model linear trends in the processing of coherence relations throughout primary school years by including grade level for children as a discrete but interval-scaled predictor. Because participants and items were sampled from a larger population, intercepts for persons and items were allowed to vary randomly. To control for main effects of item length and position of item presentation throughout the course of the experiment, the number of characters and the position of each item for each participant were included as grand mean-centered predictor variables. Main effects (fixed effects) were estimated for variables of theoretical interest. Polarity of the coherence relation and coherence were included as contrast-coded predictor variables (polarity: 1=positive-causal, -1=negative-causal;
coherence: 1=coherent, -1=incoherent). Grade level was included as grand mean-centered predictor variable to model linear developmental trends from Grades 1 to 4 in children. The age of adult participants was included as grand mean-centered predictor variable to control for a possible increase in response latency because of increasing age. In addition, interaction effects were estimated for polarity and coherence and for grade level and age with polarity and coherence to account for possible developmental differences in the processing of the various types of sentence pairs, although we expected to find no interaction effects for adults’ age with polarity and coherence. Finally, log-transformed response latency was included as grand mean-centered predictor variable in the GLMMs and accuracy as dummy-coded predictor variable (0=correct, 1=incorrect) in the LMMs to control for a potential speed-accuracy trade-off. In this way, the intercepts (and all main and interaction effects) were estimated for correct responses in the LMMs and for mean response latency in the GLMMs. The parameter estimates for the fixed and random effects are provided in Table 4.2 for the GLMMs and in Table 4.3 for the LMMs. In the following sections, we focus on the variables of theoretical interest, namely polarity, coherence, grade level (for children), and their interactions.

Visual Semantic Verification Task

Response accuracy (children). The GLMM for the accuracy data revealed significant main effects of polarity ($\beta=1.05; \ z=10.64; \ p<.001$), coherence ($\beta=0.34; \ z=3.52; \ p<.001$), log-transformed response latencies ($\beta=0.40; \ z=4.87; \ p<.001$), and grade level ($\beta=0.52; \ z=2.59; \ p=.01$). An overall increase was found in response accuracy from Grades 3 to 4. Furthermore, the model revealed three significant two-way interactions of polarity and coherence ($\beta=0.28; \ z=3.84; \ p=.005$), polarity and log-transformed response latency ($\beta=0.39; \ z=5.83; \ p<.001$), and coherence and log-transformed response latency ($\beta=-0.26; \ z=-3.96; \ p<.001$). However, these interactions were qualified by a significant three-way interaction of polarity, coherence, and log-transformed response latency ($\beta=0.18; \ z=2.78; \ p=.01$). The three-way interaction is depicted in Figure 4.2a. Figure 4.2a shows that there was a positive relationship between response latency and accuracy for positive-causal coherent, positive-causal incoherent, and negative-causal incoherent sentence pairs. In contrast, longer response latencies were associated with less accurate responses in negative-causal coherent sentence pairs.

Response latency (children). The LMM for response latency revealed significant main effects of coherence ($\beta=-0.09; \ t(112)=-4.10; \ p<.001$) and accuracy ($\beta=-0.08; \ t(112)=-2.62; \ p=.01$). Furthermore, the significant two-way interaction of coherence and accuracy ($\beta=0.07; \
\( t(112)=2.28; \ p=.02 \) and the three-way interaction of polarity, coherence, and accuracy (\( \beta=-0.07; \ t(112)=-2.31; \ p=.02 \)) indicate a positive relationship between response latency and accuracy (i.e., a speed-accuracy trade-off) for all sentence pairs except for negative-causal coherent sentence pairs (Figure 4.3). Inaccurate responses of the latter pairs were associated with longer response latencies than accurate responses.

**Figure 4.2.** Interaction of log-transformed response latency, polarity, and coherence with model-based estimated probability of correct responses as the dependent variable in the visual semantic verification task (top) and in the auditory semantic verification task (bottom) for primary school children.
Response accuracy (adults). The GLMM for response accuracy revealed a significant main effect of polarity ($\beta=1.08$; $z=3.98$; $p<.001$), indicating that adults responded with higher accuracy to positive-causal sentence pairs (99%) compared to negative-causal sentence pairs (93-95%). None of the other main effects or interaction effects were significant.

Response latency (adults). The LMM for response latency revealed significant main effects of polarity ($\beta=-0.08$; $t(47)=-3.54$; $p=.002$), coherence ($\beta=-0.08$; $t(47)=-3.44$; $p=.002$), and age ($\beta=0.01$; $t(47)=2.85$; $p=.01$). Adults provided faster responses for coherent compared to incoherent sentence pairs and younger participants provided slightly faster responses than older participants. Furthermore, a significant two-way interaction of polarity and response accuracy was found ($\beta=-0.09$; $t(47)=-2.21$; $p=.03$), indicating a speed-accuracy trade-off for positive-causal sentence pairs, whereas longer response latencies were associated with less accurate responses for negative-causal sentence pairs.

The results of the visual semantic verification task can be summarized as follows: The analysis revealed an overall increase in children’s response accuracy from Grades 3 to 4. As predicted by the cumulative cognitive complexity approach, children provided more accurate responses to coherent and incoherent positive-causal sentence pairs compared to incoherent negative-causal sentence pairs. They also responded less accurately to coherent negative-causal sentence pairs than to both types of positive-causal sentence pairs but only when response latencies were long. Somewhat unexpectedly, response accuracy was negatively associated with response latency for negative-causal coherent sentence pairs. A similar three-way interaction between polarity, coherence, and accuracy was found for response latency as dependent variable. In contrast to response accuracy, children’s response latencies cannot be clearly interpreted as being consistent with the cumulative cognitive complexity approach. Only inaccurate responses to coherent negative-causal sentence pairs were slower than coherent positive-causal sentence pairs.

Adults provided less accurate and slower responses for negative-causal compared to positive-causal sentence pairs. This pattern of results is consistent with the cumulative cognitive complexity approach. Furthermore, there was a speed-accuracy trade-off for positive-causal sentence pairs, whereas the reverse pattern was observed for negative-causal sentence pairs. Finally, response speed decreased with increasing age.

As expected, adults also needed more time to respond to incoherent compared to coherent sentence pairs. For children, positive-causal incoherent sentence pairs were associated with longer response latencies than positive-causal coherent sentence pairs, whereas negative-causal incoherent sentence pairs were associated with longer response
latencies than coherent pairs only when the responses were accurate.

Figure 4.3. Interaction of response accuracy, polarity, and coherence with model-based estimated response latency as the dependent variable in the visual semantic verification task for primary school children.

Auditory Semantic Verification Task

Response accuracy (children). The GLMM for response accuracy in the auditory semantic verification task revealed significant main effects of polarity ($\beta=1.36; z=7.60; p<.001$), grade level ($\beta=0.41; z=5.90; p<.001$) and log-transformed response latency ($\beta=0.65; z=3.49; p<.001$). The significant effect of grade level indicates that older children provided more accurate responses than younger children.

Furthermore, the analysis revealed three significant interaction effects of polarity and log-transformed response latency ($\beta=0.53; z=3.14; p=.002$), coherence and log-transformed response latency ($\beta=-1.04; z=-6.23; p<.001$), and a three-way-interaction of polarity, coherence, and log-transformed response latency ($\beta=-0.38; z=-2.28; p=.02$). The three-way interaction of polarity, coherence, and log-transformed response latency is depicted in Figure 4.2b. A positive relationship between response latency and accuracy (i.e., a speed-accuracy trade-off) was obtained only for incoherent sentence pairs. In contrast, a more thorough processing resulted in no response accuracy increase for the coherent sentence pairs. As in the visual semantic verification task, slower responses were associated with lower accuracy for negative-causal coherent sentence pairs. Figure 4.2b illustrates that children responded with overall higher accuracy to positive-causal sentence pairs compared to negative-causal sentence pairs.
### Table 4.2:

**Fixed Effects and Variance Components in the GLMMs for Response Accuracy in Children and Adults**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Visual Task</th>
<th>Auditory Task</th>
<th>Visual Task</th>
<th>Auditory Task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta (SE) )</td>
<td>( \beta (SE) )</td>
<td>( \beta (SE) )</td>
<td>( \beta (SE) )</td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.52 (0.13)*</td>
<td>1.38 (0.19)*</td>
<td>3.90 (0.35)*</td>
<td>3.99 (0.36)*</td>
</tr>
<tr>
<td>Characters(^a)</td>
<td>0.01 (0.02)</td>
<td>-0.00 (0.02)</td>
<td>0.04 (0.03)</td>
<td>0.03 (0.03)</td>
</tr>
<tr>
<td>Polarity(^b)</td>
<td>1.05 (0.10)*</td>
<td>1.36 (0.18)*</td>
<td>1.08 (0.27)*</td>
<td>0.84 (0.29)*</td>
</tr>
<tr>
<td>Coherence(^b)</td>
<td>0.34 (0.10)*</td>
<td>-0.25 (0.18)</td>
<td>0.10 (0.27)</td>
<td>0.28 (0.29)</td>
</tr>
<tr>
<td>Grade Level(^a)/Age(^a)</td>
<td>0.52 (0.20)*</td>
<td>0.41 (0.07)*</td>
<td>0.09 (0.06)</td>
<td>0.10 (0.06)</td>
</tr>
<tr>
<td>Response Latency(^a)</td>
<td>0.40 (0.08)*</td>
<td>0.65 (0.19)*</td>
<td>0.03 (0.53)</td>
<td>-1.62 (1.45)</td>
</tr>
<tr>
<td>Polarity X Coherence</td>
<td>0.28 (0.10)*</td>
<td>-0.17 (0.18)</td>
<td>-0.07 (0.26)</td>
<td>0.40 (0.28)</td>
</tr>
<tr>
<td>Polarity X Grade Level/Age(^a)</td>
<td>-0.21 (0.11)</td>
<td>-0.09 (0.04)</td>
<td>0.01 (0.05)</td>
<td>0.08 (0.05)</td>
</tr>
<tr>
<td>Coherence X Grade Level/Age(^a)</td>
<td>-0.14 (0.11)</td>
<td>-0.00 (0.04)</td>
<td>-0.09 (0.05)</td>
<td>-0.00 (0.05)</td>
</tr>
<tr>
<td>Polarity X Response Latency(^a)</td>
<td>0.39 (0.07)*</td>
<td>0.53 (0.17)*</td>
<td>1.00 (0.51)</td>
<td>1.00 (1.32)</td>
</tr>
<tr>
<td>Coherence X Response Latency(^a)</td>
<td>-0.26 (0.07)*</td>
<td>-1.04 (0.17)*</td>
<td>-0.08 (0.51)</td>
<td>-0.55 (1.28)</td>
</tr>
<tr>
<td>Polarity X Coherence X Grade Level/Age(^a)</td>
<td>0.04 (0.11)</td>
<td>-0.00 (0.04)</td>
<td>-0.09 (0.05)</td>
<td>-0.02 (0.05)</td>
</tr>
<tr>
<td>Polarity X Coherence X Response Latency(^a)</td>
<td>0.18 (0.07)*</td>
<td>-0.38 (0.17)*</td>
<td>0.31 (0.51)</td>
<td>-0.24 (1.29)</td>
</tr>
<tr>
<td><strong>Variance Components</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects</td>
<td>0.86 (0.93)</td>
<td>0.59 (0.77)</td>
<td>1.31 (1.14)</td>
<td>1.15 (1.07)</td>
</tr>
<tr>
<td>Items</td>
<td>0.15 (0.39)</td>
<td>0.72 (0.85)</td>
<td>0.46 (0.68)</td>
<td>0.83 (0.91)</td>
</tr>
</tbody>
</table>

*Note.* \(^a\)grand mean-centered, \(^b\)contrast-coded. Characters: number of characters for the sentence pair (auditory version) and the second sentence (visual version); Polarity: positive = 1, negative = -1; Coherence: coherent = 1, incoherent = -1; Grade Level/Age: Grade level (children) and age (adults). *\( p < .05 \) (two-tailed).
Table 4.3:

**Fixed Effects and Variance Components in the LMMs for Response Latency in Children and Adults**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Children Visual Task $\beta$ (SE)</th>
<th>Auditory Task $\beta$ (SE)</th>
<th>Adults Visual Task $\beta$ (SE)</th>
<th>Auditory Task $\beta$ (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>7.749 (0.05)*</td>
<td>8.54 (0.03)*</td>
<td>7.44 (0.03)*</td>
<td>8.39 (0.02)*</td>
</tr>
<tr>
<td>Characters$^a$</td>
<td>0.01 (0.00)*</td>
<td>0.01 (0.00)*</td>
<td>0.01 (0.00)*</td>
<td>0.01 (0.00)*</td>
</tr>
<tr>
<td>Polarity$^b$</td>
<td>-0.03 (0.02)</td>
<td>-0.05 (0.02)*</td>
<td>-0.08 (0.02)*</td>
<td>-0.03 (0.02)</td>
</tr>
<tr>
<td>Coherence$^b$</td>
<td>-0.09 (0.02)*</td>
<td>-0.04 (0.02)*</td>
<td>-0.08 (0.02)*</td>
<td>-0.01 (0.02)</td>
</tr>
<tr>
<td>Grade Level/Age$^a$</td>
<td>-0.08 (0.09)</td>
<td>-0.02 (0.02)</td>
<td>0.01 (0.00)*</td>
<td>0.00 (0.00)*</td>
</tr>
<tr>
<td>Response Accuracy$^c$</td>
<td>-0.08 (0.03)*</td>
<td>-0.03 (0.01)</td>
<td>-0.00 (0.04)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>Polarity X Coherence</td>
<td>0.01 (0.02)</td>
<td>0.01 (0.02)</td>
<td>0.00 (0.02)</td>
<td>0.00 (0.02)</td>
</tr>
<tr>
<td>Polarity X Grade Level/Age</td>
<td>-0.01 (0.02)</td>
<td>-0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>Coherence X Grade Level/Age</td>
<td>0.01 (0.02)</td>
<td>0.01 (0.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>Polarity X Response Accuracy</td>
<td>-0.04 (0.03)</td>
<td>0.02 (0.01)</td>
<td>-0.09 (0.04)*</td>
<td>-0.00 (0.01)</td>
</tr>
<tr>
<td>Coherence X Response Accuracy</td>
<td>0.07 (0.03)*</td>
<td>0.05 (0.01)*</td>
<td>-0.03 (0.04)</td>
<td>0.00 (0.01)</td>
</tr>
<tr>
<td>Polarity X Coherence X Grade Level/Age</td>
<td>-0.01 (0.02)</td>
<td>-0.01 (0.00)*</td>
<td>0.00 (0.00)</td>
<td>-0.00 (0.00)</td>
</tr>
<tr>
<td>Polarity X Coherence X Response Accuracy</td>
<td>-0.07 (0.03)*</td>
<td>0.01 (0.01)</td>
<td>-0.05 (0.04)</td>
<td>0.01 (0.01)</td>
</tr>
</tbody>
</table>

**Variance Components**

<table>
<thead>
<tr>
<th></th>
<th>Subjects $\sigma^2$ (SE)</th>
<th>Items $\sigma^2$ (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25 (0.50)</td>
<td>0.06 (0.25)</td>
</tr>
<tr>
<td></td>
<td>0.03 (0.17)</td>
<td>0.00 (0.03)</td>
</tr>
<tr>
<td></td>
<td>0.01 (0.09)</td>
<td>0.01 (0.08)</td>
</tr>
<tr>
<td></td>
<td>0.01 (0.11)</td>
<td>0.01 (0.08)</td>
</tr>
</tbody>
</table>

*Note:*$^a$grand mean-centered, $^b$contrast-coded, $^c$dummy-coded. Characters: number of characters for the sentence pair (auditory version) and the second sentence (visual version); Polarity: positive = 1, negative = -1; Coherence: coherent = 1, incoherent = -1; Grade Level/Age: Grade level (children) and age (adults); Response Accuracy: correct response = 0, incorrect response = 1.

*p < .05 (two-tailed).
As depicted in Figure 4.4, children’s responses to negative-causal sentence pairs were strikingly less accurate than their responses to positive-causal sentence pairs at all grade levels. At Grade 1, their response accuracy to negative-causal sentence pairs was far below chance level (28-31% correct). In other words, they systematically rejected coherent negative-causal sentence pairs as incoherent and accepted incoherent negative-causal sentence pairs as coherent. At Grade 2, they responded to negative-causal sentence pairs slightly below chance level (39-43% correct), at Grade 3 at chance level (51-55% correct), and only at Grade 4 at a level above chance (63-67% correct). Nevertheless, their responses to these items were still remarkably less accurate compared to positive-causal sentence pairs.

Response latency (children). The LMM for response latency revealed significant main effects of polarity ($\beta=-0.05; t(130)=-3.37; p=.002$) and coherence ($\beta=-0.04; t(130)=-2.29; p=.03$). Furthermore, the two-way interaction of coherence and accuracy ($\beta=0.05; t(130)=3.63; p<.001$) was significant, indicating a speed-accuracy trade-off only for incoherent sentence pairs. Finally, the three-way interaction of polarity, coherence, and grade level ($\beta=-0.01; t(130)=-3.08; p=.002$) reached significance. Older children provided faster responses than younger children to coherent and incoherent positive-causal sentence pairs and to incoherent negative-causal sentence pairs. In contrast, no increase was found in the
response speed to coherent negative-causal sentence pairs from Grades 1 to 4 (Figure 4.5).

![Figure 4.5](image)

*Figure 4.5*. Model-based estimated log-transformed response latency with standard error in the auditory semantic verification task by grade level, polarity, and coherence in primary school children.

*Response accuracy (adults).* The GLMM for response accuracy revealed a significant main effect of polarity ($\beta=0.84; z=2.89; p=.004$) indicating higher response accuracy for positive-causal compared to negative-causal sentence pairs. None of the other main or interaction effects reached significance.

*Response latency (adults).* The LMM for response latency revealed a significant main effect of age. Older participants responded slightly more slowly compared to younger participants ($\beta=0.004; t(47)=4.24; p<.001$). None of the other main or interaction effects reached significance.

The results of the auditory semantic verification task can be summarized as follows: The analysis revealed an overall increase of children’s response accuracy from Grades 1 to 4. In line with the cumulative cognitive complexity approach, children’s responses to positive-causal sentence pairs were more accurate than to negative-causal sentence pairs at all grade levels (Figure 4.2b). Their response accuracy in negative-causal sentence pairs was far below chance level at Grade 1, slightly below chance level at Grade 2, and at chance level at Grade 3. Only 4th grade pupils responded with above-chance level accuracy to negative-causal sentence pairs (Figure 4.4). Moreover, the three-way interaction of polarity, coherence, and response latency indicated a speed-accuracy trade-off for the incoherent sentence pairs and
the reverse pattern, i.e. longer latencies associated with less accurate responses, for the coherent negative-causal sentence pairs. Children needed more time to respond to negative-causal incoherent compared to positive-causal incoherent sentence pairs. The pattern was different for coherent sentence pairs. Negative-causal sentence pairs were processed slower than positive-causal sentence pairs only in the higher grade levels. This result was due to the fact that children’s response speed increased with increasing grade level for all types of sentence pairs except for coherent negative-causal sentence pairs (see Figure 4.5).

Adults responded more accurately to positive-causal sentence pairs than to negative-causal sentence pairs, as expected by the cumulative cognitive complexity approach. However, no effect of polarity was found in the response latency data. Finally, response speed decreased with increasing age.

Children but not adults responded faster to coherent than to incoherent sentence pairs, but this effect held only for accurate responses. The same interaction described differently, children responded more accurately to coherent than to incoherent sentence pairs but only when response latency was low. For long response latencies, the pattern was reversed.

**Discussion**

The present study pursued three related aims. The first aim was to replicate the finding that the processing of negative-causal coherence relations requires more cognitive effort than the processing of positive-causal coherence relations. We expected children’s and adults’ responses to the cognitively more complex negative-causal sentence pairs to be slower and less accurate than their responses to the cognitively less complex positive-causal sentence pairs. Our hypothesis was supported by the analyses of response accuracy for children and adults in the visual and the auditory modality. The more complex negative-causal coherence relations resulted in more errors in both the visual and auditory semantic verification judgment task compared to the less complex positive-causal coherence relations in children and adults. The only exception was that children providing slow responses (-2 SD) in the visual semantic verification task made fewer errors when responding to coherent negative-causal sentence pairs compared to incoherent positive-causal sentence pairs. However, as expected, their responses to coherent negative-causal sentence pairs were less accurate compared to coherent positive-causal sentence pairs (Figure 4.2a). In sum, the response accuracy results of the younger children in our study were in line with the findings for German second and third graders in Dragon et al.’s (2015) study.
In contrast to the response accuracy data, response latencies provided evidence in support of the cumulative cognitive complexity approach only in adults and only in the visual version of the task. As predicted by the cumulative cognitive complexity approach, adults needed more time to respond to negative-causal compared to positive-causal sentence pairs in the visual version of the task. However, no significant effect of polarity was found in the auditory modality. Children’s responses to incoherent negative-causal sentence pairs were slower than their responses to incoherent positive-causal sentence pairs but only in the auditory version of the task. The expected difference in response latencies was obtained only at higher grade levels for coherent sentence pairs. Data from the visual version of the task revealed no significant main effect of polarity.

In sum, the accuracy data provide evidence in favor of the cumulative cognitive complexity approach. However, the response latency data are more difficult to interpret. To understand the response latency data, closer scrutiny of their relationships with response accuracy in the children’s data is helpful. We found positive associations of response latency with response accuracy for positive-causal coherent, positive-causal incoherent, and negative-causal incoherent sentence pairs in the visual version of the task and for positive-causal incoherent and negative-causal incoherent sentence pairs in the auditory version of the task. These associations can be interpreted in terms of a speed-accuracy trade-off (Pachella, 1974): Children were able to respond more accurately when they invested more cognitive effort in the task, as indicated by the longer response latencies. In contrast, longer response latencies in the negative-causal coherent sentence pairs were associated with less accurate responses. The unique feature of the negative-causal coherent sentence pairs is that they link together two statements that are inconsistent with common knowledge but nevertheless require a positive response in the verification task. Paradoxically, more accurate responses in the verification task result when children are unaware of the inconsistency between the two statements. When children detected the inconsistency (through validation processes, Isberner & Richter, 2013), they might have also engaged in more elaborative processes in an attempt to repair the inconsistency, which lead to longer response latencies. However, these repair processes seem to have failed more often than they were successful, i.e. children were not able to construct a meaningful mental model of the described situation in coherent negative-causal sentence pairs, as indicated by the lower response accuracy.

The second aim of the present study was to directly compare the processing of positive-causal and negative-causal coherence relations in written and spoken text comprehension. With response accuracy as the dependent variable, children’s response
patterns were highly similar for the visual and the auditory semantic verification task, i.e., children responded with overall higher response accuracy to positive-causal compared to negative-causal sentence pairs as expected by the cumulative cognitive complexity approach. The only exception was that children’s responses to negative-causal coherent sentence pairs were not less accurate than their responses to positive-causal sentence pairs in the visual semantic verification task when children provided fast responses. Furthermore, longer response latencies were associated with less accurate responses in both tasks for negative-causal coherent sentence pairs.

Children’s response latency patterns were somewhat different in the visual and the auditory semantic verification task. The expected effect of polarity in the visual semantic verification task was found only for coherent sentence pairs when inaccurate responses were provided. In the auditory semantic verification task, children’s responses to incoherent positive-causal sentence pairs were faster than their responses to incoherent negative-causal sentence pairs. However, the expected effect of polarity for coherent sentence pairs was found only with increasing grade level, given that response speed for coherent positive-causal sentence pairs increased from Grades 1 to 4 but not for coherent negative-causal sentence pairs (Figure 4.5).

The pattern of adult response accuracy was highly similar for the visual and the auditory semantic verification task. They responded with higher accuracy to positive-causal compared to negative-causal sentence pairs in both tasks. However, the response latency of adults showed the expected effects of polarity and coherence only in the visual semantic verification task.

In sum, patterns of response accuracy were highly comparable for both modalities. This finding is consistent with the predictions based on the simple view of reading that processes of listening and reading comprehension are basically the same after written word forms have been decoded. However, the response latencies suggest another interpretation. The obtained differences in response latencies between both modalities were neither consistent with the predictions implied by the simple view of reading nor with the prediction that negative-causal sentence pairs are more difficult to process in the visual than in the auditory version of the task. An alternative explanation for the obtained differences in the response latency patterns between the two modalities might be that the presentation of the sentence pairs differed slightly between the visual and the auditory version of the task. In the visual version of the task, the sentence pairs remained on the notebook screen until the participant pressed the response button, i.e., participants could reread the sentence pairs if necessary, but
sentence pairs were only presented once in the auditory version of the task. The response
latencies could thus reflect different sets of processes in the visual and the auditory versions
of the task, which nullifies the possibility of comparing the latencies between the two tasks.

The third aim of the present study was to investigate age-related differences in the
processing of negative-causal and positive-causal sentence pairs throughout the first four
years of primary school. In line with the assumptions of the cumulative cognitive complexity
approach and with findings of previous studies with English and Dutch children, we found
that developmental trends in the processing of negative-causal coherence relations lag behind
developmental trends in the processing of positive-causal coherence relations. Children at all
grade levels, even the youngest children at Grades 1 and 2, responded with high accuracy to
positive-causal sentence pairs and improved further with increasing grade level. At the end of
Grade 4, children’s response accuracy for positive-causal sentence pairs approached
maximum scores in the auditory and the visual version of the task. In contrast, children’s
performance on negative-causal sentence pairs was fairly poor in both the visual and the
auditory version of the task. First graders even responded with far below-chance accuracy to
negative-causal sentence pairs. This finding indicates that children, at least at Grade 1,
systematically accept incoherent negative-causal sentence pairs as coherent and
systematically reject coherent negative-causal sentence pairs as incoherent.

How can we explain these systematically incorrect answers of the youngest children in
the auditory semantic verification task? A possible explanation might be that the younger
children simply misunderstood the task. They might have judged the plausibility of the
information conveyed by the sentence pair instead of judging its overall coherence (see e.g.,
Dragon et al., 2015), possibly because of a lack of metalinguistic knowledge (see e.g., Kail &
Weissenborn, 1984). As a result, the current paradigm might not be appropriate for at least the
youngest group of children in our study. To rule out this possibility, we analyzed a subset of
eight incoherent filler items of the auditory semantic verification task. Similar to the critical
items, four of the filler sentence pairs were incoherent because of being linked by an
inappropriate positive-causal connective (4.16) and four of them were incoherent because of
being linked by an inappropriate negative-causal connective (4.17).

(4.16)   *Der Vater schimpft viel. Darum hat der Vater einen Bart.
         *'The father grumbles a lot. Therefore, the father has a beard.’
(4.17)   *Marvin geht in die erste Klasse. Trotzdem mag er kein Gemüse.
         *'Marvin is in first grade. Nevertheless, he does not like vegetables.’
The difference between the incoherent critical items and these incoherent filler items is that the two statements in the filler items are additively related instead of causally, i.e., a positive-additive connective such as and would make the sentence pairs coherent. If children were to simply misunderstand the task and judge the plausibility or consistency of the events described in the two statements independent of the linking connective, we would expect them to systematically (and incorrectly) accept each of the eight incoherent filler items. The reason is that the contents of the sentences involved are not implausible or inconsistent (e.g., Marvin is in first grade and he does not like vegetables). However, if the youngest children’s poor performance on negative-causal items in our auditory semantic verification task were due to difficulties in comprehending negative-causal connectives and coherence relations (in contrast to positive-causal ones) as predicted by the cumulative cognitive complexity approach, children should correctly reject the four filler items linked by an inappropriate positive-causal connective and should struggle to correctly respond to the four filler items linked by a negative-causal connective. A GLMM revealed higher estimated probabilities for correct responses for filler sentence pairs linked by positive-causal connectives compared to filler sentence pairs linked by negative-causal connectives at all grade levels. The youngest children, in particular, performed strikingly poorly on filler items containing negative-causal connectives but not on filler items containing positive-causal connectives. Thus, children did not respond incorrectly to the eight filler items as would have been expected if they had simply misunderstood the task. In contrast, children at all grade levels responded less accurately to filler items containing negative-causal compared to positive-causal connectives. This result suggests that the poor performance on negative-causal items of the youngest children in our auditory semantic verification task was due to difficulties in processing negative-causal connectives and coherence relations as predicted by the cumulative cognitive complexity approach.

In sum, the accuracy results we obtained for children at Grades 1 to 4 (auditory task) and Grades 3 and 4 (visual task) indicate that, as expected, even the younger children performed fairly well on positive-causal coherence relations, whereas their performance on negative-causal coherence relations was poor. Moreover, despite the increase in response accuracy to negative-causal sentence pairs, children’s response accuracy to negative-causal sentence pairs at the end of Grade 4 was still far below the accuracy obtained for positive-causal sentence pairs. These findings strongly suggest that the comprehension of negative-causal coherence relations is still developing throughout the course of primary school, whereas even the youngest children in our study mastered the comprehension of positive-
causal coherence relations. These conclusions are consistent with previous findings on the acquisition order of coherence relations in primary school children (e.g., Cain & Nash, 2011; Katz & Brent, 1968; Spooren & Sanders, 2008; Wing & Scholnick, 1981) and with the assumptions of the cumulative cognitive complexity approach.

Finally, we expected shorter response latencies for coherent compared to incoherent sentence pairs. This pattern was observed in children in the visual version of the task for positive-causal sentence pairs but was observed only for accurate responses to negative-causal sentence pairs. In the auditory version of the task, younger children’s responses to incoherent negative-causal sentence pairs were slower than their responses to coherent sentence pairs, whereas a similar pattern was found for positive-causal sentence pairs in older children. The expected pattern of response latencies was also observed in adults in the visual semantic verification task. The longer response latencies for incoherent sentence pairs thus appear to reflect the cognitively laborious attempt to repair the mental model of the text when encountering and integrating incoherent information (Cain & Nash, 2011). Yet, the expected effect of coherence was not found for adults in the auditory version of the task and only in interaction with polarity and response accuracy or grade level in the children’s data. This null finding might be explained by the sensitivity of response latencies to several strategic or non-strategic on-line and off-line processes of text comprehension, which we could not control for (as elaborated below).

The results presented so far should be interpreted with possible limitations of the present study in mind. First, latencies were recorded only for responses at the end of the second sentence. More detailed analyses of on-line reading times might be more appropriate to reveal processing differences between positive-causal and negative-causal sentence pairs as expected by the cumulative cognitive complexity approach. For example, a recent self-paced reading study by Knoepke, Richter, & Diener (2015) investigated the time course of validating and integrating consistent and inconsistent information in positive-causal and negative-causal sentence pairs. Although they found no overall differences in reading times for the whole sentence pairs, their analyses revealed longer reading times on inconsistent information and at the end of the second statement in negative-causal sentence pairs compared to positive-causal sentence pairs, which contained only consistent information. In accordance with the assumptions of the cumulative cognitive complexity approach, their findings indicate increased cognitive effort in the processing of negative-causal coherence relations, in particular when detecting and integrating inconsistent information into the existing mental model of the text.
A general methodological limitation of the current study is its cross-sectional design. The more appropriate way to investigate developmental changes in language comprehension is to employ a longitudinal design. Given our design of investigating children from four different grade levels in parallel, we cannot be entirely sure whether the developmental changes are due to differences between the different groups. Nevertheless, the accuracy data are in accordance with previous findings on the acquisition order of positive-causal and negative-causal connectives and coherence relations. Thus, it is likely that our findings reflect real developmental trends instead of random group differences.

A third limitation addresses the external validity of our study. The material used to test the cumulative cognitive complexity approach included very short texts containing only one sentence pair each. Arguably, sentence pairs might not be sufficient to investigate processes of establishing coherence, because they are not sufficiently naturalistic. An argument against this concern is that the comprehension and integration of two sentences already requires the establishment of local coherence relations, because local coherence per definition links the mental representations of two adjacent statements in a text. Thus, to study the establishment of local coherence in isolation, two adjacent sentences or statements represent a minimalistic but sufficient approach. Nonetheless, further research is needed to show whether the results we obtained generalize to longer texts.

Finally, we focused on only two types of coherence relations characterized along only two of the four dimensions of the coherence taxonomy proposed by Sanders et al. (1992). Further studies are needed to investigate whether the predictions of the cumulative cognitive complexity approach also hold for coherence relations other than positive-causal or negative-causal coherence relations. For example, research could focus on positive-additive and negative-additive relations, coherence relations based on pragmatic versus semantic sources, or coherence relations with basic versus non-basic order in German.

Although the response latency data are difficult to interpret, the response accuracy data in our study suggest that the cumulative cognitive complexity approach can be applied to German primary school children and adults. We demonstrated that the processing of negative-causal coherence relations is cognitively more demanding than the processing of positive-causal coherence relations in German children from Grades 1 to 4 and adults in reading and listening comprehension. We also showed that children’s comprehension of negative-causal coherence relations is still developing throughout the course of primary school. These findings have important implications for classroom instruction and the composition of textbooks in primary school. Considering the crucial role the establishment of local coherence...
plays in written and spoken text comprehension (Graesser et al., 1997; Van Dijk & Kintsch, 1983), the finding that primary school children even at Grades 3 and 4 exhibit great difficulties in processing negative-causal coherence relations is of high practical relevance. To ensure successful text comprehension as early as possible, a reasonable and promising, if not necessary, strategy should be to include systematic trainings concerning the establishment of coherence relations (e.g., in the form of strategy exercises) into primary school curricula.

**Acknowledgments.** The research presented in this article was supported by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF, grants 01 GJ 0985, 01 GJ 0986, 01 GJ 1206A, 01 GJ 1206B). We would like to thank David Nitz for programming assistance and the numerous student research assistants who were involved in collecting data for this project. Researchers who are interested in the material used in the visual and auditory semantic verification tasks are invited to send an e-mail to the first or the second author.
Chapter V

Study 4

Construct Validity of a Process-oriented Test Assessing Skills of Syntactic Integration in German Primary School Children

A version of this chapter is submitted as:

Construct Validity of a Process-oriented Test Assessing Skills of Syntactic Integration in German Primary School Children

Julia Knoepke, Tobias Richter, Maj-Britt Isberner, Johannes Naumann, & Yvonne Neeb

Abstract. Reading comprehension is based on the efficient accomplishment of several cognitive processes at the word, sentence, and text level. To the extent that each of these processes contributes to reading comprehension, it can cause reading difficulties if it is deficient. To identify individual sources of reading difficulties, tools are required that allow for a reliable and valid assessment of individual differences in specific cognitive processes of reading comprehension. The present study demonstrates the usefulness of this process-oriented approach to assessing reading comprehension skills using the example of a novel test for assessing syntactic integration skills in German primary school children. The test comprises a grammaticality-judgment task that contains items with carefully varied features which are known to facilitate or impede syntactic integration processes. By means of explanatory item-response models, we demonstrate that empirical item difficulties vary as a function of experimentally varied item features, indicating that the test indeed assesses skills of syntactic integration. Moreover, the test measures individual differences in syntactic integration across the whole range of skill levels and is sensitive to developmental changes. We conclude that the test is a valid tool to assess individual differences and to detect deficits in this component process of reading comprehension.

Keywords: Explanatory item-response model, grammaticality judgment task, reading comprehension, syntactic integration

Introduction

From a cognitive perspective, reading comprehension may be regarded as the efficient mastery of cognitive component processes at the word, sentence, and text level (Knoepke & Richter, in press; Perfetti et al., 2005). In addition to the fact that the component processes of reading interact in various ways, each of them makes a unique contribution to reading
comprehension. Accordingly, all component processes are potential causes for reading
difficulties when they are deficient (Vellutino et al., 2004). Therefore, the construction of
successful and target-oriented reading interventions and remediation programs requires
diagnostic instruments which allow for a reliable and valid assessment of individual
differences in cognitive component processes of reading comprehension.

However, diagnostic instruments assessing individual differences in cognitive
component processes of reading comprehension in an adequate way are still scarce. For
example, most of the available tests assessing reading comprehension skills in German (e.g.,
Hamburger Leseprobe, May & Arntzen, 2000; HAMLET 3-4, Lehmann et al., 1997) and
English (e.g., WRAT 4, Wilkinson & Robertson, 2006; WRMT-III, Woodcock, 2011) focus
either on skills of visual word recognition or on rather global measures of reading
comprehension, thereby neglecting the complex structure of cognitive component processes
of reading comprehension at the word, sentence, and text level. Moreover, the majority of
tasks used in these tests involve additional cognitive skills unrelated to reading, such as
picture recognition or the memorization of response alternatives in multiple-choice tasks (e.g.,
FLVT 5-6, Souvignier et al., 2008; GSRT, Wiederholt & Blalock, 2000). Therefore, measures
of reading ability might be contorted at least to a certain degree by individual differences in
reading-unrelated cognitive skills. Finally, measures of reading comprehension often rely on
accuracy, i.e. the number of successfully solved test items, as the sole diagnostic criterion.
However, over and above accuracy, the time it takes to solve reading-related tasks may
provide important information about the efficiency of cognitive processes involved in reading
comprehension. Despite the fact that time-based measures are a common indicator of
cognitive effort and automaticity of cognitive processes in experimental studies of reading in
L1 and L2 (e.g., Hulstijn, van Gelderen, & Schoonen, 2009; Lim & Godfroid, 2015; Trapman,
van Gelderen, van Steensel, van Schooten, & Hulstijn, 2014), such measures are very rarely
used in standardized tests for the assessment of reading comprehension skills (e.g., ELFE 1-6,
Lenhard & Schneider, 2006), with the exception of speed tests such as one minute of reading
(e.g., Deno, 1985; Deno et al., 2001; SLRT-II, Moll & Landerl, 2010; SLS 2–9, Wimmer &
Mayringer, 2014).

The aim of the present study was to present a process-oriented alternative to product-
oriented tests. This approach is characterized by selectively assessing specific component
processes of reading comprehension at the word, sentence, and text level, by well-defined
reading tasks, theoretically based test items with carefully varied item features, and by using
accuracy as well as response latencies as diagnostic criteria. In the following, we illustrate the
usefulness of this approach using the example of a novel test for assessing syntactic integration skills as one of the cognitive component skills of reading comprehension in German primary school children. The test is novel insofar as it combines carefully designed and theoretically based test items (with item features that—according to psycholinguistic theory and research—are known to facilitate or impede syntactic integration processes) with assessing both accuracy and response latencies as measures of the efficiency of syntactic integration skills. In the following, we will first discuss problems of existing diagnostic approaches. Then, we describe the rationale behind the process-oriented test, a visual grammaticality judgment task for 3rd and 4th graders. The visual grammaticality judgment task is complemented by a strictly parallel auditory grammaticality judgment task for 1st to 4th graders, which allows examining commonalities and differences between syntactic skills in reading and listening comprehension. The core of the paper are results from explanatory item response models which provide empirical evidence of the construct validity of the novel test by showing that item features that should affect the ease of syntactic integration processes predict the empirical difficulties of the test items (Wilson & De Boeck, 2004).

**Reading Comprehension is Based on Efficient Cognitive Processes**

Successful comprehension of written texts is based on efficient cognitive processes at the word, sentence, and text level (Perfetti et al., 2005). On the word level, the reader has to identify written word forms and assign meanings to individual words. Readers accomplish this task indirectly by *phonological recoding* of written words and directly by comparing written word forms to orthographical representations in the mental lexicon (*orthographical decoding*) (Coltheart et al., 2001). In addition, word meanings have to be retrieved from the mental lexicon (*access to word meanings*). On the sentence level, readers need to integrate word forms and their meanings into a coherent sentence meaning (*semantic integration*) in consideration of the sentence’s grammatical structure (*syntactic integration*) (Müller & Richter, 2014; Pickering & van Gompel, 2006; Pylkkänen & McElree, 2006). Finally, on the text and discourse level, readers need to construct and continuously update a situation model (mental model) of the text content by *establishing local and global coherence relations* between adjacent and distant statements and by integrating text information with prior knowledge (Johnson-Laird, 1981; Van Dijk & Kintsch, 1983). Each of these cognitive processes makes a unique contribution to reading comprehension. Consequently, individual differences in one or more of these cognitive processes result in individual differences in reading comprehension (Knoepke & Richter, in press). In order to create successful and
adaptive intervention and remediation programs, it is important to determine the exact source and severity of individual reading deficits.

To accomplish this, psychological tests are required that assess individual differences in each of these cognitive component processes selectively. However, most available tests assessing reading comprehension skills are subject to several restrictions. We will illustrate these restrictions with examples of German tests, but similar restrictions apply to tests in English as well.

First, most of the available tests either focus on visual word recognition skills (e.g., SLRT-II; WLLP-R, Schneider et al., 2011) or assess reading comprehension skills in a product-oriented fashion (e.g., FLVT 5-6; Hamburger Leseprobe; LGVT 6-12, Schneider et al., 2007; SLS 2–9; VSL, Walter, 2013). This means that reading comprehension is measured after a text has been read, implying that these tests assess aspects of comprehension outcomes, i.e. aspects of the resulting mental representation, rather than the quality of comprehension processes going on during reading. The diagnostic merits of product-oriented tests notwithstanding, it is clear that these tests provide no straightforward way to identify individual sources of poor reading abilities and, consequently, they are no adequate basis for the construction of individual target-oriented intervention programs.

A second restriction concerns the use of tasks which involve cognitive skills that are not specific to reading comprehension, such as finding pictures that match target words (e.g., ELFE 1-6; HAMLET 3-4; WLLP-R,) or a correct answer among several alternatives in a multiple-choice task (e.g., ELFE 1-6; FLVT 5-6; LESEN 6-7, Bäuerlein et al., 2012a; LESEN 8-9, Bäuerlein et al., 2012b; VSL). For these tasks, one cannot rule out the possibility that measures of reading skills are contorted at least to a certain degree by individual differences in reading-unrelated cognitive skills (such as deficits in picture comprehension or memory).

Finally, reading comprehension skills are often operationalized as the number of successfully solved test items (e.g., HAMLET 3-4; Hamburger Leseprobe; VSL). While this approach may be informative with regard to the quality of the mental representations constructed during reading, it does not tell us anything about the efficiency of the cognitive processes involved. By an efficient cognitive process, we mean a process that is not only reliable, that is, likely to generate a correct outcome, but also well routinized, implying that the process makes low demands on cognitive resources. This component of efficiency is captured in the time it takes to carry out reading-related tasks. Despite the fact that time-based measures are a common indicator of automaticity of cognitive processes in experimental studies on reading in L1 and L2 and on L2 acquisition (e.g., Hulstijn et al., 2009; Lim &
Godfroid, 2015; Trapman et al., 2014), at present, time measures are very rarely used in standardized assessments of reading comprehension skills (e.g., ELFE 1-6). Only speed tests of reading fluency incorporate a component of processing time as an inherent part of reading assessment (e.g., IEL-1, Diehl & Hartke, 2012; LDL, Walter, 2010; SLRT-II; one minute of reading, Deno, 1985; Deno et al., 2001). However, these tests usually focus on the fluency of single-word reading and do not assess individual differences in specific cognitive component skills of reading comprehension.

Assessing Syntactic Integration Skills in German

Syntactic integration skills are a case in point to illustrate the usefulness of a process-based approach to reading comprehension skills. Several studies suggest a strong relationship between syntactic integration skills and text comprehension. Byrne (1981) found that children with poor reading skills had significantly more difficulties than skilled readers to find pictures matching syntactically complex sentences when the task was not facilitated by the semantic content of the sentence. Moreover, poor readers were found to be less able than skilled readers to correct grammatical violations in spoken sentences or to supply a missing word (Tunmer et al., 1987), they were less able than skilled readers to restructure the words in a scrambled sentence (Nation & Snowling, 2000), and they performed inferior to skilled readers on Bishop’s (1983) test for the reception of grammar (TROG; Stothard & Hulme, 1992). In a longitudinal study, Casalis and Louis-Alexandre (2000) found that morpho-syntactic skills in French kindergarten children were predictive of their sentence comprehension skills at the end of Grade 2. In line with these findings, Plaza and Cohen (2003) found that French children’s performance in a grammaticality judgment and error correction task predicted their reading and spelling skills at the end of Grade 1 and that their grammatical skills accounted for unique variance in their reading and spelling skills besides phonological awareness, naming speed, and auditory memory (for a critical discussion on the relationship between skills of syntactic integration and reading comprehension and on possible mediators and moderators, see Oakhill & Cain, 1997).

In sum, these findings indicate that individual differences in syntactic integration account for a unique portion of variance in reading comprehension skills and imply that deficient skills of syntactic integration can impede reading comprehension. Consequently, to identify readers whose poor reading abilities are due to deficient skills of syntactic integration, a psychological test is required that reliably measures individual differences in syntactic integration skills.
In German, several tests assess dimensions of language development in children and many of them include at least one subtest assessing syntactic skills. However, most of these standardized diagnostic tests are limited in certain ways. First, the majority of these tests focus on language production skills rather than on language comprehension or, at least, they involve language production as part of the task, preventing selective assessment of comprehension skills (e.g., MSS, Holler-Zittlau, Dux, & Berger, 2003; P-ITPA, Esser, Wyschkon, Ballaschk, & Hänsch, 2010; PET, Angermaier, 1977; SET 5-10, Petermann, Metz, & Fröhlich, 2010; SVV, Grimm, 2003). Second, most tests assessing syntactic skills in German children involve cognitive skills that are not specific to language processing such as picture identification, the memorization of multiple response alternatives, acting out sentence contents etc. (e.g., ETS 4-8, Angermaier, 2007; HSET, Grimm & Schöler, 2001; KISTE, Häuser, Kasielke, & Scheidereiter, 1994; MSVK, Elben & Lohaus, 2001; SETK 3-5, Grimm, Aktas, & Frevert, 2001; TROG-D, Fox, 2011). Third, to our knowledge, none of these tests includes measures of processing latencies (e.g., ADST, Steinert, 2011; ETS 4-8; P-ITPA; SET 5-10; SETK 3-5; TROG-D) that would be needed to assess the efficiency of syntactic integration processes (see Jeon & Yamashita, 2014 for further discussion). Finally, as far as we know, there is no German test that assesses syntactic skills in reading and listening comprehension selectively and in a methodologically stringent way, that is, by means of strictly parallel tasks and materials in both modalities. Higher-order cognitive component processes of reading and listening comprehension such as syntactic and semantic integration are often regarded to be strongly overlapping after written word forms have been decoded, a theoretical position put forward, for example, by the simple view of reading (Gough & Tunmer, 1986). The simple view of reading states that after written words have been decoded, processing written language is basically the same as processing spoken language, i.e. cognitive processes of language comprehension beyond the word level (such as, for example, syntactic integration) should be highly comparable in both modalities. However, to our knowledge, this assumption has never been tested for specific component processes of reading such as syntactic integration. This lack of evidence may be due to the fact that there are no diagnostic instruments assessing the efficiency of specific comprehension processes in both modalities adequately and selectively.

The aim of the present study was to test the construct validity of a novel test, which assesses syntactic integration skills in listening and reading comprehension in beginning readers in a methodologically stringent and process-oriented way, by way of explanatory item response models.
Assessing Syntactic Skills: Principles of Test Construction

To assess skills of syntactic integration in primary school children, we constructed a visual grammaticality judgment task for children at Grades 3 and 4 and an auditory grammaticality judgment task, which can be used with children from Grade 1 to 4. The visual grammaticality judgment task is part of a more comprehensive German-language test battery that consists of six subtests, each of them assessing a cognitive component process of reading comprehension at the word, sentence, or text level selectively (ProDi-L: Prozessbezogene Diagnostik des Leseverstehens bei Grundschulkindern [Process-based assessment of reading skills in primary school children], Richter et al., in press). An earlier study by Richter et al. (2012) revealed that individual differences in syntactic integration skills, assessed via the grammaticality judgment task of ProDi-L, are related to individual differences in reading comprehension, assessed via the standardized text comprehension subtest of ELFE 1-6. In addition, there is a strictly parallel German-language test battery for measuring the corresponding component processes of listening comprehension (ProDi-H: Prozessbezogene Diagnostik von Hörverstehensfähigkeiten bei Grundschulkindern [Process-based assessment of listening skills in primary school children]). In the grammaticality judgment tasks, children are presented with written (ProDi-L) and spoken (ProDi-H) grammatical sentences such as (5.1) and ungrammatical sentences such as (5.2):

\[\text{(5.1) } \text{Die Arbeiter bauen das Haus} / \text{The workmen build the house}\]

\[\text{(5.2) } *\text{Die Hexe auf einem Besen reitet} / *\text{The witch on a broom rides}\]

By pressing one of two response keys they indicate whether the spoken and written sentences are grammatically correct or not. This kind of task requires and, therefore, assesses cognitive skills of syntactic integration. Moreover, it does not involve any other skill that is not specific to language processing such as picture recognition or the memorization of alternative multiple-choice-responses. Both response accuracy, which is indicative of the reliability of cognitive processes, and response latency, which is indicative of their degree of routinization, are recorded.

Ability tests should include items of varying difficulty that allow selective measurements on different levels of ability. A straightforward and theoretically sound way to create items of varying difficulty for a test of syntactic integration skills is to vary specifically those item features that—according to psycholinguistic theory and research—may be expected to facilitate or impede syntactic integration processes. Subjects with good syntactic
integration skills should be able to solve overall more and, in particular, more difficult items, whereas subjects with poor syntactic integration skills are expected to solve fewer items and, in particular, the easier ones. If the items of the grammaticality judgment tasks are well constructed and indeed assess skills of syntactic integration, these item features should also have a measurable effect on empirical item difficulties (De Boeck & Wilson, 2004). Thus, empirical item difficulties should be predictable from the experimentally varied item features.

One of the features varied in the present test to generate items of different difficulty is syntactic complexity, which has been found to be associated with difficulties in sentence and text comprehension. For example, Graesser, Hoffmann, and Clark (1980) found that increasing syntactic complexity increased reading times of sentences. Ferreira, Henderson, Anes, Weeks, and McFarlane (1996) found longer processing times for words in a syntactically demanding context (i.e., in a garden-path sentence) than in a syntactically less demanding context. Marton, Schwartz, and Braun (2005) found that children’s response accuracy to comprehension questions decreased with increasing syntactic complexity (see also Marton & Schwartz, 2003 and Marton, Schwartz, Farkas, & Katsnelson, 2006, who found decreased performance accuracy with increasing morphological complexity in Hungarian children). Accordingly, error rates and response latencies are expected to increase with syntactic complexity of the test items.

Another feature relevant for the difficulty of grammaticality judgments is the grammaticality of sentences itself. Flores d’Arcais (1982, Exp. 4) found longer fixation durations on syntactically erroneous words in visually presented ungrammatical sentences as compared to grammatically appropriate words in grammatically well-formed sentences. In accordance with these findings, Baum (1989; 1991) found in a word-detection paradigm that target words were detected more slowly in ungrammatical as compared to grammatical spoken sentences. Finally, Friederici, Hahne, and Saddy (2002) found participants to make more errors in a grammaticality judgment task when judging ungrammatical as compared to grammatical sentences. On this account, we expected ungrammatical items to be more difficult than grammatical items in terms of longer response latencies and lower response accuracy for ungrammatical items.

Finally, the type of grammatical violation was varied systematically to create items of varying difficulty. There are several grammatical markers or cues (Bates & MacWhinney, 1987, 1989) such as word order or case marking that help a reader or listener to comprehend a sentence. However, recent findings suggest that German children have not yet fully acquired all of them equally well at the entry of primary school. Dittmar, Abbot-Smith, Lieven, and
Tomasello (2008) and Schipke, Knoll, Friederici, and Oberecker (2012) found that 3 to 5-year-olds and 6 to 7-year-olds exhibit difficulties in using case-marking information to assign thematic roles. Instead, they rely on word-order information even if case marking is the more reliable cue. These findings indicate that young German children are less sensitive to case marking than to word-order information. Against this background, we expected detecting case-marking violations to be more difficult for primary school children than detecting word-order violations.

All items of the visual and the auditory grammaticality judgment task were systematically varied with respect to the three aforementioned features: syntactic complexity, grammaticality, and (for the ungrammatical sentences) type of violation. The ungrammatical sentences contained either a word-order violation (5.2) (here repeated as ((5.3)), a case-marking violation (5.4), or a violation of the verb’s tense form (5.5):

(5.3) *Die Hexe auf einem Besen reitet / *The witch on a broom rides
(5.4) *Die Schafe fressen dem-DAT Gras / *The sheep eat the-DAT grass
(5.5) *Lisa hat einen Brief schrieb / *Lisa has wrote a letter

Usually, in German main clauses, the perfect tense form requires the finite auxiliary (hat/has) to occur in V2-position and an infinite past participle form (ge-schrieb-en/written) in sentence final position. Furthermore, in German main clauses, finite verb forms always occur in V2-position. The violations of tense form in our study consist of a finite simple past tense verb form (e.g., schrieb) in sentence final position. Thus, they actually contain two types of violation, a violation of word order (the finite verb occurs in sentence final position) and a morphological violation (the verb form in sentence final position is incorrectly conjugated). This might enable even faster and more accurate recognition of tense-form violations as compared to violations of word order or case marking.

We would like to point out that this test alone is not meant to be sufficient for a comprehensive diagnosis of individual sources of reading difficulties. A comprehensive diagnosis requires a test battery that assesses individual differences in all cognitive component processes of reading comprehension at the word, sentence, and text level selectively. Such a test battery is provided by ProDi-L, with the grammaticality judgment task presented in this paper being part of this comprehensive process-oriented reading test.
Construct Validity of the Grammaticality Judgment Task

In the present study, empirical item difficulties were estimated using a one-parameter logistic model (1PL or Rasch model) for the accuracy data and a Rasch-analogue model for the (log-)transformed response latencies. In a second step, item difficulties were modeled as a function of theoretically-based item features by estimating explanatory item-response models, i.e. a logistic multilevel regression model for the accuracy and a linear multilevel regression model for the (log-transformed) response latencies (e.g., Hartig, Frey, Nold, & Klieme, 2012). The predicted item difficulties from the item-response models were then correlated with the empirical item difficulties to obtain a test of construct validity.

The Rasch model for logit-transformed response accuracy (i.e., log-transformed odds ratios for correct responses) as dependent variable with items as dummy-coded fixed effects and subjects as random effect was as follows:

\[
\text{logit}(\text{accuracy}_{ij}) = b_{0j} + b_{1j} X_{1ij} + b_{2j} X_{2ij} + \ldots + b_{(k-1)j} X_{(k-1)ij} + r_{ij}.
\]

\[
b_{0j} = g_{00} + u_{0j} \quad \text{Random coefficient, } u_{0j}: \text{person parameter}
\]
\[
b_{1j} = g_{10}
\]
\[
\ldots
\]
\[
b_{(k-1)j} = g_{(k-1)0} \quad \text{Fixed coefficients: item parameters}
\]

The one item that is not represented by a dummy-coded fixed effect is called the reference item. The empirical item difficulties can be computed from this model as the sum of the intercept of the model $g_{00}$ and the coefficients $b_{(k-1)j}$ for each specific item $X_{(k-1)ij}$ (the item difficulty of the reference item is captured by the intercept itself). An analogous model was estimated for log-transformed response latencies as dependent variable (Van Breukelen, 2005).

In a second step, a Logistic Linear Test Model (LLTM, Fischer, 1974) with the theoretically derived item features as fixed item-level predictors and a random effect of subjects was estimated:

\[
\text{logit}(\text{accuracy}_{ij}) = b_{0j} + b_{1j} (\text{feature}_{1})_{ij} + b_{2j} (\text{feature}_{2})_{ij} + \ldots + b_{qj} (\text{feature}_{q})_{ij} + r_{ij}.
\]

\[
b_{0j} = g_{00} + u_{0j} \quad \text{Random coefficient, } u_{0j}: \text{person parameter}
\]
\[
b_{1j} = g_{10}
\]
\[
\ldots
\]
\[
b_{(k-1)j} = g_{(k-1)0} \quad \text{Fixed coefficients: effects of item features}
\]
Again, an analogous linear multilevel model was estimated for the (log-transformed) response latencies as dependent variable (Van Breukelen, 2005). Based on the intercept $g_{00}$ and the regression coefficients $b_{qj}(\text{feature}_{ij})$ of the LLTM and the LLT-analogous model, item difficulties predicted from their item feature characteristics were computed. A substantial correlation between the empirical item difficulties from the Rasch-model (or the Rasch-analogue model) and the LLTM (or the LLT-analogous model) may be regarded as evidence for the construct validity of the test.

Finally, in terms of criterion validity, a test assessing skills of syntactic integration should also detect developmental changes and learning progress during the primary school years. Thus, response accuracy was expected to increase and response latency was expected to decrease from Grade 3 to 4 (visual grammaticality judgment task) and from Grade 1 to 4 (auditory grammaticality judgment task). To test this hypothesis, grade level was included as another fixed item-level predictor into the LLTM for response accuracy and the LLT-analogous model for the log-transformed response latencies.

Method

Participants

Participants were 1380 primary school children (678 boys and 658 girls, for 44 children gender information was missing) recruited from 26 schools (95 classes) in Cologne, Kassel, and Frankfurt am Main (Germany). One-thousand-one-hundred-and-eighteen children (548 boys, 531 girls, for 39 children gender information was missing) from Grades 1 to 4 (Grade 1: $n=439$; Grade 2: $n=237$; Grade 3: $n=232$; Grade 4: $n=209$; for one child grade level information was missing) completed the auditory version of the task, and 691 children (330 boy, 354 girls, for 7 children gender information was missing) from Grades 3 and 4 (Grade 3: $n=332$; Grade 4: $n=359$) completed the visual version of the task. Of these children, 429 children in Grades 3 and 4 participated in both the auditory and the visual version of the grammaticality judgment task. The children’s age ranged from 5;5 to 12;4 years (Grade 1: $M=7;6$; $SD=0;6$; Grade 2: $M=8;5$; $SD=0;6$; Grade 3: $M=9;7$; $SD=0;7$; Grade 4: $M=10;6$; $SD=0;5$) in the auditory version of the task and from 6;9 to 12;4 (Grade 3: $M=9;5$; $SD=0;7$; Grade 4: $M=10;5$; $SD=0;5$) in the visual version of the task. Socio-demographic data were collected via a parent questionnaire and were supplemented by information from a teacher questionnaire when information from the parent questionnaire was missing. Children only participated in the study when parents had provided written consent.
Materials

The grammaticality judgment task contained 38 written sentences in the visual version and 38 spoken sentences in the auditory version. Two additional practice sentences preceded each version and served as ice-breaker items as well. They were excluded from the analysis. Half of the test sentences in each version (19 sentences) were grammatically well-formed German sentences (5.1), whereas the other half contained a grammatical violation. Of these 19 ungrammatical sentences, 10 sentences contained violations of word order (5.3), 4 sentences contained violations of case marking (5.4), and 5 sentences contained violations of the verb-tense form (5.5). Written and spoken sentences were comparable with respect to length (number of characters, spoken sentences: $M=37.97; SD=12.01; \text{Min}=17; \text{Max}=59$; written sentences: $M=37.82; SD=11.60; \text{Min}=21; \text{Max}=56$), syntactic complexity (number of syntactic phrases, spoken sentences: $M=4.16; SD=1.37; \text{Min}=2; \text{Max}=7$; written sentences: $M=4.21; SD=1.42; \text{Min}=2; \text{Max}=7$), and number and types of violations. For ungrammatical sentences, the relative position of the violation was also held constant in the parallel versions. The position of the violation was determined as the point from which a grammatical completion was no longer possible (position of violation divided by number of words, spoken sentences: $M=0.89; SD=0.22$; written sentences: $M=0.92; SD=0.15$; a relative position of 1 indicates a violation at the last word of a sentence). Half of the spoken sentences were recorded by a male and the other half by a female speaker.

Procedure

The grammaticality judgment task was part of a cross-sectional study investigating processes of listening and reading comprehension with various measures on the word, sentence, and text level (Richter et al., 2012; Richter et al., 2013). The children were tested together in classrooms of the participating schools. The grammaticality judgment task was presented on notebook computers, embedded in a story of an extraterrestrial named Reli who asked the children for their help to learn the earthinglings’ language by telling him when he did something wrong. Reli introduced the tasks in short animated video clips and walked the children through the tasks. In the auditory version of the grammaticality judgment task, children listened to the practice and the test sentences via headphones. In the visual version of the task, the sentences were presented in the center of the screen one at a time (font: Verdana, visual angle: 1.5 degrees). Each child received all 38 test sentences of a version in randomized order. For each sentence, children were asked to judge whether it was correct or
not by pressing one of two response buttons (green button on the keyboard for *yes, the sentence is correct* or a red button for *no, the sentence is not correct*). Prior to the presentation of test items, the two practice sentences were presented, for which children received feedback from Reli. When they gave an incorrect response, the practice sentences were repeated. Log-transformed response latencies (measured from stimulus onset to the press of a response button) and response accuracy were recorded as dependent variables.

**Results**

Responses that were unusually slow or fast given the typical item- or person-specific response latencies (3 $SD$ or more below the item-specific mean and 2 $SD$ or more below or above the person-specific mean after standardizing each item by its item-specific mean) were excluded from the analyses because these responses, most likely, came about irregularly (e.g., the child accidentally pressed the response button too early, or he or she was distracted by something in the classroom). Due to this procedure there was a loss of 5.7% of the children’s data in the visual version of the task and 7.7% in the auditory version. Descriptive statistics for the visual and the auditory grammaticality judgment tasks are reported in Table 5.1.

Empirical item difficulties were estimated using Rasch models for the accuracy data and Rasch-analogous models for log-transformed response latencies (Van Breukelen, 2005) with items nested within subjects, fixed effects for items (items included as dummy-coded predictors), and random effects for subjects. Effects of experimentally varied item features on response accuracy and latency were estimated using Linear Logistic Test Models (LLTMs, Fischer, 1974) with a logit link function (Dixon, 2008) for response accuracy and LLT-analogous models for log-transformed response latency with items nested within subjects, experimentally varied item features as fixed effects (syntactic complexity as grand-mean centered predictor, violation of word order, tense form, and case marking as dummy-coded predictors), and subjects as random effects. All models were estimated and tested with the software packages *lme4* (Bates et al., 2014) and *lmerTest* for R (Kuznetsova et al., 2014). Parameters were estimated with Restricted Maximum Likelihood (REML). All significance tests were based on a type-I error probability of .05. Separate models were estimated for the visual and the auditory grammaticality judgment task and for accuracy and response latency as dependent variables.
Table 5.1:

Descriptive Statistics for Response Latency (Log-transformed) and Accuracy as Dependent Variables in the Auditory and Visual Grammaticality Judgment Task

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Grade 1</th>
<th>Grade 2</th>
<th>Grade 3</th>
<th>Grade 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M (SD)</td>
<td>n</td>
<td>M (SD)</td>
<td>n</td>
</tr>
<tr>
<td>Auditory grammaticality task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response Accuracy(a)</td>
<td>39207</td>
<td>0.821 (0.38)</td>
<td>15006</td>
<td>0.752 (0.43)</td>
<td>8420</td>
</tr>
<tr>
<td>Response Latency(b)</td>
<td>39207</td>
<td>8.038 (0.36)</td>
<td>15006</td>
<td>8.082 (0.41)</td>
<td>8420</td>
</tr>
<tr>
<td>Visual grammaticality task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response Accuracy(a)</td>
<td>24764</td>
<td>0.883 (0.32)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Response Latency(b)</td>
<td>24764</td>
<td>8.204 (0.57)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. N/n = number of data points (items X participants). \(a\)relative frequency, \(b\)log-transformed. For one child (36 observations), Grade level information was missing.

Visual Grammaticality Judgment Task

Empirical item difficulties (depicted as the log-transformed odds ratios for solving a specific item across participants) and person abilities (log-odds for providing a correct response across items) estimated with the Rasch model for logit-transformed response accuracy as dependent variable are depicted in Figure 5.1a. Both distributions were unimodal. The distribution of person abilities was negatively skewed. All item difficulties were higher than 0.0 (probability of 50% to solve an item), indicating that the items were relatively easy for 3rd and 4th graders in terms of response accuracy (values with a positive sign indicate a probability higher than 50% to solve a specific item, whereas values with a negative sign indicate a probability less than 50% to solve a specific item). Empirical item difficulties (log-transformed latencies for providing a response to a specific item across participants) and person abilities (log-transformed latencies for providing a response across items) estimated with the Rasch-analogous model for log-transformed response latency as dependent variable are depicted in Figure 5.1b. Again, both distributions were unimodal and symmetrical. For both response accuracy and latency, the distributions of the item difficulties overlap substantially with the distributions of person abilities, suggesting that the set of items
measures individual differences across the whole range of person abilities. However, items
differentiated individual skills of syntactic integration best for children with average
(response latency) or average to highly efficient syntactic integration processes (accuracy).

The parameter estimates for the fixed and random effects of the LLTM and the LLT-
analogous model are provided in Table 5.2 (Model 1). All main effects reached significance.
As expected, children’s response accuracy decreased ($\beta=-0.04$; $z=-3.03$; $p<.05$) with
increasing syntactic complexity. Their responses were significantly less accurate for sentences
containing a violation of word order ($\beta=-0.29$; $z=-5.83$; $p<.05$), tense form ($\beta=-0.25$; $z=-3.91$;
$p<.05$), or case marking ($\beta=-0.95$; $z=-15.88$; $p<.05$) as compared to grammatically correct
sentences. This effect was strongest for sentences containing violations of case marking and
weakest for violations of tense form. The estimates from the model for log-transformed
response latency indicate that children needed significantly more time ($\beta=0.06$; $t(686)=29.37$;
$p<.05$) to respond to syntactically complex sentences as compared to syntactically less
complex sentences. Furthermore, they needed more time to respond to sentences containing a
violation of word order ($\beta=0.05$; $t(686)=7.03$; $p<.05$), tense form ($\beta=0.09$; $t(686)=10.55$;
$p<.05$), or case marking ($\beta=0.07$; $t(686)=7.03$; $p<.05$) as compared to grammatically correct
sentences. This effect was strongest for sentences containing violations of tense form and for
violations of word order.

Based on the specific feature characteristics of each item and the regression
coefficients for these features from the LLTM and the LLT-analogous model, predicted item
difficulties for response accuracy and response latency were computed. The predicted item
difficulties correlated strongly and positively with the empirical item difficulties. For response
accuracy, 45% of the variance in the empirical item difficulties was explained by the
predicted item difficulties (Figure 5.2a). For response latency, 38% of the variance in the
empirical item difficulties was explained by the predicted item difficulties (Figure 5.2b).
Thus, the systematic variation of item features which are expected to facilitate or increase the
difficulty of syntactic integration processes resulted in items of varying difficulty. These
findings suggest that the items of the visual grammaticality judgment task indeed assess skills
of syntactic integration.

Finally, a test assessing skills of syntactic integration should detect developmental
changes and learning progress during primary school years. To test this prediction, grade level
was included as further dummy-coded fixed effect into the LLTM for response accuracy and
the LLT-analogous model for response latencies. The regression coefficients are presented in
Table 5.2 (Model 2).
Figure 5.1. Item difficulties (left) and person abilities (right) for logit-transformed response accuracy (top) and log-transformed response latency (bottom) in the visual grammaticality judgment task in Grades 3 and 4.

As expected, children at Grade 4 provided significantly faster ($\beta=-0.20; t(685)=-7.68; p<.05$) and overall more accurate responses ($\beta=0.20; z=2.77; p<.05$) than children at Grade 3. This finding indicates that the visual grammaticality judgment task is sensitive to developmental changes in cognitive processes of syntactic integration.
Figure 5.2. Variance in empirically observed item difficulties explained by item difficulties predicted by item characteristics for accuracy (top) and latency (bottom) in the visual grammaticality judgment task in Grades 3 and 4.

Auditory Grammaticality Judgment Task

Empirical item difficulties (log-odds for solving a specific item across participants) and person abilities (log-odds for providing a correct response across items) estimated with the Rasch model for logit-transformed response accuracy as dependent variable are depicted in Figure 5.3a. Both distributions were unimodal. The distribution of item difficulties exhibited a slightly negative skew. Again, most item difficulties were higher than 0.0, indicating that the items were relatively easy to solve for students from Grades 1 to 4.
Table 5.2:

Fixed Effects and Variance Components in the LLTMs for Response Accuracy and LLT-analogous Models for Response Latency in the Visual Grammaticality Judgment Task

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$ (SE)</td>
<td>$\beta$ (SE)</td>
<td>$\beta$ (SE)</td>
<td>$\beta$ (SE)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>8.162 (0.01)*</td>
<td>8.265 (0.02)*</td>
<td>2.479 (0.05)*</td>
<td>2.373 (0.06)*</td>
</tr>
<tr>
<td>Number of Syntactic Phrases</td>
<td>0.061 (0.00)*</td>
<td>0.061 (0.00)*</td>
<td>-0.044 (0.01)*</td>
<td>-0.044 (0.01)*</td>
</tr>
<tr>
<td>Violation of Word Order</td>
<td>0.049 (0.01)*</td>
<td>0.049 (0.01)*</td>
<td>-0.292 (0.05)*</td>
<td>-0.292 (0.05)*</td>
</tr>
<tr>
<td>Violation of Tense Form</td>
<td>0.094 (0.01)*</td>
<td>0.094 (0.01)*</td>
<td>-0.249 (0.06)*</td>
<td>-0.249 (0.06)*</td>
</tr>
<tr>
<td>Violation of Case Marking</td>
<td>0.071 (0.01)*</td>
<td>0.071 (0.01)*</td>
<td>-0.949 (0.06)*</td>
<td>-0.949 (0.06)*</td>
</tr>
<tr>
<td>Grade Level</td>
<td>-0.198 (0.03)*</td>
<td></td>
<td>0.205 (0.07)*</td>
<td></td>
</tr>
</tbody>
</table>

Variance Components

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>0.119 (0.34)</td>
<td>0.109 (0.33)</td>
<td>0.623 (0.79)</td>
<td>0.615 (0.78)</td>
</tr>
</tbody>
</table>

*Note. Number of syntactic phrases (grand-mean centered). Violation of Word Order (dummy coded): Sentences containing a violation of word order (=1) vs. all other sentence types (=0). Violation of Tense Form (dummy coded): Sentences containing a violation of tense form (=1) vs. all other sentence types (=0). Violation of Case Marking (dummy coded): Sentences containing a violation of case marking (=1) vs. all other sentence types (=0). Grade Level (dummy coded): Grade 4 (=1) vs. Grade 3 (=0). *p < .05 (two-tailed).

Empirical item difficulties (log-transformed latencies for providing a response to a specific item across participants) and person abilities (log-transformed latencies for providing a response across items) estimated with the Rasch-analogous model for log-transformed response latency as dependent variable are depicted in Figure 5.3b. The distributions of item difficulties and person abilities were unimodal and symmetrical. For both response accuracy and latency, the distribution of item difficulties overlapped substantially with the distribution of person abilities, indicating that the items are suitable to reveal individual differences in subjects’ skills of syntactic integration. Again, the items differentiated individual skills of syntactic integration best for children with average or efficient syntactic integration processes.

The parameter estimates for the fixed and random effects of the LLTM and the LLT-analogous model are provided in Table 5.3 (Model 1). In the LLTM with logit-transformed response accuracy as dependent variable, the main effects for violation of word order (β=--
Study 4

0.73; \( z = -19.90; p < .05 \), tense form (\( \beta = -0.43; z = -8.89; p < .05 \)), and case marking (\( \beta = -3.06; z = -67.98; p < .05 \)) reached significance. Children responded less accurately to sentences containing one of the three types of violation as compared to grammatically correct sentences. This effect was strongest for violations of case marking and weakest for violations of tense form. Somewhat unexpectedly, there was no significant main effect for syntactic complexity. The LLT-analogous model for log-transformed response latency as dependent variable revealed significant main effects for syntactic complexity (\( \beta = 0.11; t(1111) = 100.46; p < .05 \)), violation of tense form (\( \beta = 0.07; t(1111) = 15.78; p < .05 \)), and violation of case marking (\( \beta = 0.15; t(1111) = 30.10; p < .05 \)). Children’s response latency increased with increasing syntactic complexity and they needed more time to respond to sentences containing violations of tense form or case marking as compared to grammatically correct sentences. This effect was strongest for violations of case marking. There was no significant main effect for violation of word order.

Again, we predicted item difficulties from the LLTM and the LLT-analogous model for response accuracy and latency. The predicted item difficulties correlated substantially and positively with the empirical item difficulties. For response accuracy, 73% of the variance in the empirical item difficulties was explained by the predicted item difficulties (Figure 5.4a). For response latency, 54% of the variance in the empirical item difficulties was explained by the predicted item difficulties (Figure 5.4b). Overall, the systematic variation of item features resulted in items of varying difficulty. Thus, the items of the auditory grammaticality judgment task seem to be suitable to assess skills of syntactic integration.

Finally, grade level was included into the LLTM for response accuracy and the LLT-analogous model for response latency in the form of three dummy-coded fixed effects for Grades 2, 3, and 4 (Table 5.3, Model 2) with Grade 1 as the reference category. There were no significant differences between Grade 2 and Grade 1. However, children provided faster (\( \beta = -0.04; t(1108) = -2.84; p < .05 \)) and overall more correct responses (\( \beta = 0.85; z = 11.36; p < .05 \)) in Grade 3 than in Grade 1. Moreover, they responded even faster (\( \beta = -0.10; t(1108) = -6.02; p < .05 \)) and with even higher accuracy (\( \beta = 1.39; z = 17.10; p < .05 \)) in Grade 4 as compared to Grade 1. These results suggest that the auditory grammaticality judgment task is suitable to detect developmental changes in cognitive processes of syntactic integration during primary school years (Figures 5.5a and b).
Figure 5.3. Item difficulties (left) and person abilities (right) for logit-transformed response accuracy (top) and log-transformed response latency (bottom) in the auditory grammaticality judgment task in Grades 1 to 4.
Figure 5.4. Variance in empirically observed item difficulties explained by item difficulties predicted by item characteristics for accuracy (top) and latency (bottom) in the auditory grammaticality judgment task in Grades 1 to 4.
a) **Auditory Grammaticality Judgment Task (Response Accuracy)**  
**Grade Level**

![Graph showing model-based estimated probability of correct responses with standard error for different grade levels.](image)

b) **Auditory Grammaticality Judgment Task (Response Latency)**  
**Grade Level**

![Graph showing model-based estimated response latency with standard error for different grade levels.](image)

*Figure 5.5.* Model-based estimated probability of correct responses with standard error (top) and model-based estimated response latency with standard error (bottom) in the auditory grammaticality judgment task by grade level.
Table 5.3:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Response Latency</th>
<th></th>
<th>Response Accuracy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Intercept</td>
<td>8.002 (0.01)*</td>
<td>8.033 (0.01)*</td>
<td>2.494 (0.04)*</td>
<td>1.902 (0.05)*</td>
</tr>
<tr>
<td>Number of Syntactic Phrases</td>
<td>0.107 (0.00)*</td>
<td>0.107 (0.00)*</td>
<td>-0.004 (0.01)</td>
<td>-0.003 (0.01)</td>
</tr>
<tr>
<td>Violation of Word Order</td>
<td>0.003 (0.00)</td>
<td>0.003 (0.00)</td>
<td>-0.734 (0.04)*</td>
<td>-0.732 (0.04)*</td>
</tr>
<tr>
<td>Violation of Tense Form</td>
<td>0.069 (0.00)*</td>
<td>0.069 (0.00)*</td>
<td>-0.430 (0.05)*</td>
<td>-0.431 (0.05)*</td>
</tr>
<tr>
<td>Violation of Case Marking</td>
<td>0.145 (0.00)*</td>
<td>0.145 (0.00)*</td>
<td>-3.061 (0.05)*</td>
<td>-3.063 (0.05)*</td>
</tr>
<tr>
<td>Grade 2</td>
<td>-0.016 (0.02)</td>
<td></td>
<td>0.733 (0.07)*</td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>-0.044 (0.02)*</td>
<td></td>
<td>0.851 (0.07)*</td>
<td></td>
</tr>
<tr>
<td>Grade 4</td>
<td>-0.097 (0.02)*</td>
<td></td>
<td>1.391 (0.08)*</td>
<td></td>
</tr>
</tbody>
</table>

Variance Components

| Subjects                   | 0.036 (0.19)      | 0.034 (0.19)  | 0.834 (0.91)      | 0.571 (0.76)  |

Note: Number of syntactic phrases (grand-mean centered). Violation of Word Order (dummy coded): Sentences containing a violation of word order (=1) vs. all other sentence types (=0). Violation of Tense Form (dummy coded): Sentences containing a violation of tense form (=1) vs. all other sentence types (=0). Violation of Case Marking (dummy coded): Sentences containing a violation of case marking (=1) vs. all other sentence types (=0). Grade 2 (dummy coded): Grade 2 (=1) vs. Grade 1, 3, and 4 (=0). Grade 3 (dummy coded): Grade 3 (=1) vs. Grade 1, 2, and 4 (=0). Grade 4 (dummy coded): Grade 4 (=1) vs. Grade 1, 2, and 3 (=0).
*p < .05 (two-tailed).

Discussion

The aim of the present study was to demonstrate the usefulness of a process-oriented approach to assessing individual differences in cognitive component skills of reading comprehension. To this end, we used the example of a novel test for assessing syntactic integration skills in German primary school children. This test stands out due to its careful test design and stringent methodology: (1) It assesses individual differences in skills of syntactic integration in a process-oriented fashion, i.e. separately from other reading-related cognitive component skills. (2) It does not involve reading-unrelated cognitive skills such as picture
recognition or the memorization of multiple response alternatives. (3) It assesses both the reliability and automaticity of syntactic integration processes by recording response accuracy, being indicative of the reliability of a cognitive process, and response latency, being indicative of its automaticity. (4) Finally, individual differences in syntactic integration skills are assessed in reading as well as listening comprehension in a methodologically stringent way, i.e. by means of parallel tasks and materials in both modalities. The construct validity of this test was examined by way of explanatory item response models.

For both the visual and the auditory grammaticality judgment task, items measured individual differences across the whole range of person abilities. For response accuracy and latency, the distribution of item difficulties overlapped substantially with the distribution of person abilities. Items differentiated individual skills of syntactic integration best for children with average or efficient syntactic integration processes in both versions of the test.

Furthermore, we demonstrated that empirical item difficulties varied as a function of experimentally varied item features which are known to facilitate or hinder processes of syntactic integration. These findings suggest that the items of both tasks indeed assess skills of syntactic integration. In the following, we summarize and discuss how empirical item difficulties varied as a function of experimentally varied item features, namely syntactic complexity, grammaticality, and type of syntactic violation.

As expected, children’s response latencies increased and accuracy decreased with increasing syntactic complexity. The only exception was found in the accuracy data of the auditory version. A possible explanation might be that even the most complex syntactic sentences in the test are not complex enough to show individual differences in response accuracy in the auditory processing modality. For primary school children, spoken language processing requires less cognitive processing resources as compared to reading comprehension. The available resources might be used to deal sufficiently with more complex sentences, resulting in highly accurate responses for spoken grammatical sentences of all complexity levels (with error rates of less than 10%). Yet, the impact of syntactic complexity on spoken sentence processing is reflected in response latencies, indicating that the test is still suitable to measure individual differences in the automaticity of syntactic integration processes.

As to grammaticality, children needed more time and made more errors when responding to ungrammatical sentences—indeed of violation type—as compared to grammatically correct sentences. As expected, children responded least accurately to sentences containing a case-marking violation and most accurately to sentences containing a
violation of tense form, with sentences containing a violation of word order in between them in the visual as well as in the auditory grammaticality judgment task. Overall, analogous results were found for response latencies. However, there were two unexpected findings: First, children did not need more time to respond to sentences containing a violation of word order as compared to grammatically correct sentences in the auditory grammaticality judgment task. A likely explanation might be that, as discussed earlier, German children at the entry of primary school are already highly sensitive to word-order information as a cue to spoken sentence comprehension (Dittmar et al., 2008; Schipke et al., 2012). Therefore, they might have little problems detecting word-order violations when sentences are presented auditorily. This explanation is supported by the finding that the overall response accuracy to spoken sentences containing word-order violations is comparatively high (>85%), suggesting that the missing significant main effect is not simply due to a speed-accuracy trade-off. We assume that processing of spoken word order violations is comparatively easy (as indicated by short response latencies), because word order violations occur more often in spoken than in written language comprehension. The second unexpected finding was that, in the visual version of the task, children’s response latencies to sentences containing a violation of tense form were as high as their response latencies to sentences containing a violation of case marking. Possibly, children encountering the incorrect verb tense form at the end of the written sentence regressed to the finite auxiliary that appeared earlier in the sentence to recheck whether the auxiliary and the sentence final verb form matched or not. Such a regression to the auxiliary might lead to longer response latencies for violations of tense form and is possible only in the visual grammaticality judgment task.

Finally, when empirical and predicted item difficulties were correlated, there were a few items deviating considerably from the estimated regression line. In the visual grammaticality judgment task children responded to (5.6) and (5.7) faster than would have been predicted by the item features, i.e. the items were actually somewhat easier than would have been expected.

(5.6) Ein Vogel ist zum Nest geflogen / A bird has flown to a nest
(5.7) Das Eis ist in der Sonne geschmolzen / The ice has melted in the sun

Moreover, sentence (5.6) also elicited more accurate responses than would have been expected from its item feature characteristics. The most likely explanation is that there are some other item features, such as the verb’s valence or lexical effects, which were not varied
systematically but nevertheless determine item difficulties to some degree. In contrast, sentence (5.8) of the auditory grammaticality judgment task turned out to be more difficult with respect to response latencies than one would expect based on its item feature characteristics.

(5.8) *Der Fuchs hat die Gans gestohlen* / The fox has stolen the goose

A possible explanation is that sentence (5.8) triggers the association of a German nursery song called *Fuchs du hast die Gans gestohlen* and that the activation of this association slows down response latencies for this item despite its feature characteristics. However, despite these few exceptions, the predicted item difficulties correlated overall strongly and positively with the empirical item difficulties, suggesting that the items of the visual and the auditory grammaticality judgment task indeed assess skills of syntactic integration.

Last but not least, a test assessing skills of syntactic integration should detect developmental changes and learning progress during primary school years. Our findings indicate that both the visual and the auditory grammaticality judgment task are sensitive to developmental changes in cognitive processes of syntactic integration.

Overall, the findings of the present study provide evidence in favor of the construct validity of the grammaticality judgment task, that is, the grammaticality judgment task together with the systematically constructed item set provides a valid instrument to assess individual differences in skills of syntactic integration in German primary school children from Grade 3 to 4 (visual version of the task) and Grade 1 to 4 (auditory version of the task). Whereas response accuracy in the grammaticality judgment task is indicative of the reliability of syntactic integration processes, response latencies are indicative of the degree of automatization. Richter et al. (2012) demonstrated that both measures in the grammaticality judgment task are significantly related to reading comprehension. The authors found that primary school children’s performance on the standardized text comprehension subtest of ELFE 1-6 (Lenhard & Schneider, 2006) was positively associated with response accuracy and negatively associated with log-transformed reading latencies in the grammaticality judgment task of ProDi-L. Moreover, the positive relationship between response accuracy in the grammaticality judgment task and reading comprehension was even more strongly pronounced when processes of syntactic integration were highly automatized as indicated by a significant interaction of response accuracy and response latency. These findings suggest that inaccurate and slow responses in our grammaticality judgment task are indicative of
inefficient syntactic integration processes, which, in turn, may be one of the causes for individual reading difficulties. If deficient syntactic integration processes contribute to poor reading comprehension in individual readers, it should be possible to use this test to detect the deficit. Consequently, it serves as a basis for the construction of interventions and remediation programs that are tailored to the individual needs of poor readers with deficient processes of syntactic integration.

The present study emphasized the necessity for process-oriented tests of assessing reading comprehension skills to identify individual deficits in cognitive component skills of reading comprehension, which can cause reading difficulties. By means of explanatory item-response models, we demonstrated that process-oriented tests (such as ProDi-L) involving well-defined reading tasks (such as the grammaticality judgment task presented in the present study), theoretically-based test items with carefully varied item features, and accuracy as well as response latency as diagnostic measures, are valid tools to achieve that goal.

Acknowledgments. The data presented in this article were collected as part of the project Process-based assessment of reading and listening skills in primary school children, which was funded by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF, grants 01 GJ 0985, 01 GJ 0986, 01 GJ 1206A, and 01 GJ 1206B). We would like to thank David Nitz for programming assistance and several student assistants for their help in constructing stimulus materials and collecting data. Researchers who are interested in the material used in the visual and auditory grammaticality judgment tasks are invited to send an e-mail to the first or the second author.
Chapter VI

Sources of Poor Reading Comprehension
Reading Comprehension: Individual Differences, Disorders, and Underlying Cognitive Processes

A version of this chapter is published as:
Reading Comprehension: Individual Differences, Disorders, and Underlying Cognitive Processes

Julia Knoepke & Tobias Richter

Abstract. Poor readers are classified as dyslexic readers when they show poor below age-average reading comprehension in the absence of general cognitive deficits. However, a diagnosis of dyslexia based on this definition bears no information about the cause of the individual reading deficit or the kind and extent of required intervention. Identifying the specific cause and severity of reading comprehension problems is essential to create adequate and target-oriented intervention programs for poor readers. This chapter provides an overview of the cognitive processes underlying reading comprehension and discusses how reading disorders can be characterized in terms of deficits in phonological recoding, orthographical decoding, access to word meanings, syntactic and semantic integration, and establishing local and global coherence. We conclude that ‘the’ dyslexic reader defined by one specific cognitive deficit is a misconception. Instead, the sources and symptoms of reading disability are multifaceted and heterogeneous, and the individual pattern of deficits needs to be considered when planning remediation and intervention programs.

Keywords: Aptitude-achievement discrepancy, dyslexia, reading comprehension, reading difficulty, text comprehension

Introduction

Reading comprehension is one of the preconditions for a successful educational development. Therefore, one of the most important goals of the educational system is the early identification of poor readers and the development of individual intervention and remediation programs to help them overcome their reading difficulties. But, under what conditions is a reader considered to be a poor reader? Usually, poor readers are diagnosed with specific reading disability (developmental dyslexia) when they show below age-average reading comprehension in the absence of any other cognitive deficit and adverse
Sources of Poor Reading Comprehension

environmental factors (American Psychiatric Association, 2013; World Health Organization, 2010). Thus, only readers performing substantially worse on standardized reading tests than expected levels based on their general level of cognitive functioning are considered to be dyslexic. To date, this discrepancy model of dyslexia is widely used by educators and researchers to identify poor readers and assign them to specific training and remediation programs.

Despite the widespread use of the discrepancy model, diagnosing a reader as dyslexic provides no information about individual underlying causes of poor reading comprehension nor the kind and the extent of required intervention. Even worse, the operational definition of a separate category of dyslectic readers according to the discrepancy model requires the use of cut-off values that, besides lacking a substantial rationale, exclude poor readers from intervention programs who show a broader range of cognitive disabilities.

In this chapter, we will focus on reading-specific cognitive processes as sources for reading difficulties, excluding such possible sources as working memory, general knowledge, visual, attentional, or neurological deficits (for a review on potential causes of dyslexia that are not specific to reading see Vellutino et al., 2004; Vidyasagar & Pammer, 2010). We will first discuss the traditional definition of dyslexia based on the discrepancy model and its problems. We argue that a more fruitful approach to characterize poor readers and their individual needs for reading intervention would be to examine reading comprehension deficits in a manner that is consistent with the cognitive processes that constitute reading comprehension rather than to simply diagnose a reader as dyslexic or not. Thus, our goal is to provide an overview of the cognitive processes underlying reading comprehension at the word, sentence, and text level and delineate why and how deficits in these processes can contribute to a low level of reading comprehension. We emphasize that identifying the specific origin of reading difficulties is essential to being able to assign poor readers to an appropriate intervention program.

Diagnostic Criteria of Dyslexia and their Problems

Estimates of developmental dyslexia prevalence range from 10 to 15%, depending on the exact operational definition (Vellutino et al., 2004). These numbers render dyslexia one of the most prevalent learning disorders. According to the International Classification of Diseases (ICD-10), dyslexic readers manifest “a specific and significant impairment in the development of reading skills that is not solely accounted for by mental age, visual acuity
problems, or inadequate schooling” (F81.0, World Health Organization, 2010). The term dyslexia is nonexistent in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5, APA, 2013). Instead, the manual contains a similar definition of Specific Learning Disorder with reading difficulties as the further specification. Thus, dyslexic readers exhibit severe difficulties in the acquisition of basic reading and spelling skills in the absence of a general learning deficit (Rack et al., 1992; Vellutino et al., 2004). Schools, remediation programs, and researchers following this definition rely primarily on two skill criteria to identify dyslexic readers: Reading skills that are significantly worse than would be expected based on (1) a reader’s chronological age and (2) a reader’s cognitive abilities or mental age (often operationalized by measures of intelligence). Thus, readers with at least an average IQ (≥ 80-90; Siegel, 1988) who perform unexpectedly poor on reading tasks as compared to their peer’s performance are considered dyslexic. These readers are usually distinguished from another group of poor readers called general backward readers (Rutter & Yule, 1975) or garden-variety poor readers (Stanovich, 1988) who also fail to acquire age appropriate reading skills, but in contrast to dyslexic readers, they are additionally characterized by a broader range of cognitive deficits accompanied by a low IQ (≤80). This is also known as the aptitude-achievement discrepancy, which indicates that a dyslexics’ ability to read (achievement) diverges from their expected levels based on their intellectual capacity (aptitude). The discrepancy model is largely based on the work by Yule et al. (1974) who found considerably more poor readers (a “hump”) at the lower end of the reading skill distribution of readers than would have been statistically expected assuming a normal distribution. Yule et al. and Rutter and Yule (1975) assume that a subgroup of the poor readers must be qualitatively different from the normally developing poor readers because of a specific reading deficit. Furthermore, Yule (1973) claimed that the future prospects concerning reading development are significantly worse for dyslexic than for backward readers and conclude that the distinction between these two groups of readers is both meaningful and beneficial for remediation.

These conclusions, however, have been extensively challenged in recent years. Two major arguments against the usefulness of the dyslexia definition in identifying and characterizing poor readers have been advanced. The first and most important objection is that dyslexia, defined as unexpectedly low reading achievement despite normally developing cognitive skills, lacks diagnostic value with respect to the kind of underlying deficit and required intervention. One specific unitary deficit in poor readers is a misconception. Instead, the sources of individual reading deficits are multifaceted and heterogeneous. Several
reading-related cognitive component skills may be impaired in poor readers to different
degrees. Therefore, each deficit requires a specific intervention that addresses the specific
impaired reading-related process and its degree of severity (Coltheart & Jackson, 1998).

The second objection concerns the assumption of a discrete group of dyslexic readers
that differ qualitatively from a group of general backward readers. Stanovich (2005), one of
the most emphatic opponents of the discrepancy criterion, maintained that the literature lacks
evidence showing that dyslexic and general backward readers process reading-related
information in a different manner. Siegel (1988) and Stanovich and Siegel (1994) measured
the performance of poor readers with high IQ scores on several tasks that tapped cognitive
reading-specific skills and compared it with the performance of readers with lower IQ scores.
The children in both studies were presented with a battery of reading-related tasks, for
example, word and non-word reading, spelling, phonological recoding, grammatical closure,
and sentence repetition, and they were also presented with tasks assessing skills that are less
specific to reading such as working memory capacity. Both studies consistently indicated that
the distinction of good vs. poor readers strongly predicted children’s performance on reading-
specific tasks, whereas IQ scores (high vs. low) did not (see also Vellutino, Scanlon, & Lyon,
2000 and a meta-analysis by Stuebing et al., 2002). Hence, several authors (e.g., Shaywitz et
al., 1992; Stanovich, 1988) have argued that dyslexic readers represent the lower end of a
continuous distribution of readers rather than a discrete category. If readers vary gradually on
a continuum of reading ability and if this reading ability is independent of IQ, then setting an
arbitrary IQ-based cut-off between dyslexic and general backward readers has no basis.
However, despite these findings, the distinction between dyslexic and general backward
readers on the basis of IQ scores is still widely used. Consequently, children classified as poor
but not dyslexic readers are often excluded from research and interventional programs based
on an arbitrary cut-off criterion (Catts et al. 2003; Shaywitz et al., 1992; Stanovich, 1988).
Instead, as Siegel emphasizes, identify the impaired component processes of reading and the
particular form and extent of the deficit would be far more helpful, including determining the
appropriate strategy for enhancing the deficient processes (see also Catts et al., 2003).

In view of these findings, we argue for a cognitive-psychological approach to reading
comprehension difficulties. Rather than defining a group of dyslexics on the basis of
questionable criteria, a more fruitful approach would be to examine the reading difficulties in
terms of the underlying cognitive processes and to determine the extent to which component
processes are impaired and the appropriate strategy for improving the mastery of these
processes. The aim of the following sections is to provide an overview of the possible sources
What Causes Poor Reading Comprehension?

Cognitive-psychological research on reading comprehension has identified a number of cognitive processes at the word, sentence, and text level that contribute to reading comprehension (Müller & Richter, 2014; Perfetti, 2001; Richter & Christmann, 2009). First, readers must recognize the written word forms of a text. According to dual-route models of visual word recognition, readers accomplish this task via two different routes (Coltheart et al., 2001). To be able to recognize unknown or infrequent word forms, readers use a non-lexical, rule-based phonological route by translating the word letter-by-letter into a phonemic representation (phonological recoding). The phonemic representation is subsequently mapped on to an entry in the mental lexicon. When processing familiar and highly frequent word forms, readers use an orthographic or lexical route by which word forms are recognized holistically and mapped directly on to an entry in the mental lexicon (for evidence supporting the dual route cascaded model of visual word recognition—DRC, see e.g. Paap & Noel, 1991; Ziegler et al., 2000). After successfully recognizing a word form, readers need to retrieve its meaning from the mental lexicon. At the sentence level, they must integrate the word forms syntactically and semantically. Finally, in text and discourse comprehension, several sentences need to be integrated into a coherent mental model of the text by establishing local and global coherence relations between adjacent and distant sentences (McNamara & Magliano, 2009; Van Dijk & Kintsch, 1983). This multi-level structure of component skills implies that reading comprehension succeeds to the extent that readers master all of the cognitive processes involved in reading efficiently. Individual differences in these processes are potential sources of individual differences in reading comprehension skills. Hence, deficits in the mastery of these processes potentially cause specific types of reading difficulties.

Individual Differences at the Word Level

The majority of studies investigating possible causes of poor reading comprehension have focused on word-level processes. This method seems to be a reasonable starting point, because the ability to recognize written word forms is clearly crucial for reading
Sources of Poor Reading Comprehension

The importance of word-level processes for individual differences in reading comprehension is expressed very clearly in the simple view of reading (SVR, Gough & Tunmer, 1986; Hoover & Gough, 1990), which assumes reading comprehension \( (R) \) to be the product of two types of cognitive abilities, the general ability to comprehend language \( (C) \) and the ability to decode written word forms \( (D) \):

\[
R = D \times C
\]

The multiplicative combination of \( D \) and \( C \) implies that good decoding skills and good general comprehension skills are each necessary but not sufficient to bring about good reading comprehension. Instead, reading comprehension is impaired when only one of the two abilities is low. According to the simple view of reading, decoding is the only process that distinguishes reading from listening comprehension. Consequently, visual word recognition is a prominent candidate when looking for possible sources of reading difficulties.

Another general theoretical approach that emphasizes the role of word recognition processes in reading comprehension is Perfetti’s (1985) verbal efficiency hypothesis, which states that efficient word recognition constitutes the fundament of successful reading comprehension. The underlying idea is that efficient (i.e., rapid and reliable) word-recognition processes save cognitive resources, which are then available for higher cognitive processing, such as sentence and text level processing. The verbal efficiency hypothesis was further refined into the lexical quality hypothesis by Perfetti and Hart (2001, 2002; see also Perfetti, 2010), which emphasizes that the quality of the representations of word forms, including the stability and interconnectedness of their constituents (phonological, orthographic, morphological, and semantic components), is the basis for good reading comprehension.

Individual Differences in Phonological Recoding

Most explanatory approaches of dyslexia and of poor reading comprehension in beginning readers agree that a likely source of reading disability is a deficit in phonological recoding. This idea is appealing from a developmental point of view. Phonological recoding skills are the key to the acquisition of reading skills, because word forms are still unknown to beginning readers and need to be recoded letter-by-letter (Coltheart et al., 2001; see also the developmental model by Frith, 1985). As a consequence, deficient phonological recoding hinders the child to read the majority of written word forms and impairs all further stages of reading development. Deficits in phonological recoding may be caused by deficits in general
Sources of Poor Reading Comprehension

Stanovich (1988) and Stanovich and Siegel (1994) compared the performance of poor and skilled readers on several tasks accessing phonological skills in written and auditory modality. They found that all poor readers, in contrast to skilled readers, exhibited severe problems with tasks, such as regular and exception word naming, non-word naming, and rhyme production. Based on his findings, Stanovich (1988) created the phonological-core variable-differences model, which states that poor readers primarily suffer from a deficit in phonological processing skills that prevents them from the acquisition of age-appropriate reading abilities. Evidence in favor of this assumption comes from various sources. For example, Snowling (1980) found that poor dyslexic readers, in contrast to skilled readers, had difficulties recognizing an auditorily presented word in its written form and exhibited the same difficulties in the reverse order. Because this task required grapheme-phoneme-conversion in both directions, Snowling concluded that the poor readers had difficulties in mapping sounds on letters and letters on sounds. Griffith and Snowling (2001) investigated whether the phonological deficit of poor readers is due to deficient phonological representations or to a deficit in retrieving the phonological information. They found that 11- to 12-year-old poor readers with the diagnosis of dyslexia performed worse than good readers of the same age in rapid-naming and non-word reading tasks that required the retrieval of phonological information. However, in an auditory word-gating task, no differences were found between good and poor readers on the amount of phonetic input they needed to identify a spoken word. The authors conclude that the deficit of poor readers is due to deficient retrieval processes rather than deficient phonological representations. This interpretation was further supported by more recent studies by Ramus et al. (2013) and Dickie et al. (2013). Their results indicate that phonological deficits in poor readers are not due to deficient phonological representations but rather to poor skills in assessing or manipulating them. However, using a similar word-gating paradigm to the one used by Griffith and Snowling, Boada and Pennington (2006) found evidence for deficient implicit phonological representations in poor readers rather than deficient phonological retrieval processes. In contrast to the findings by Griffith and Snowling, the poor readers in the study by Boada and Pennington needed more phonetic input to correctly recognize the first letter in a word than the chronological age-control group and more phonetic input to correctly recognize the whole word than the chronological age-control group and the reading age-control group. The authors concluded that poor readers have more “immature phonological representations” (2006, p. 177) than their age and reading peers.
A number of longitudinal and training studies provided evidence to support the assumption of a causal relationship between phonological deficits and poor reading abilities (Rack et al., 1992; Vellutino et al., 2004). These studies demonstrated that children’s phonological skills in kindergarten predict reading comprehension in primary school (e.g., Bradley & Bryant, 1983; Scanlon & Vellutino, 1996). Moreover, interventions strengthening the phonological awareness in kindergarten and at the beginning of primary school were shown to have a positive impact on later reading comprehension skills (e.g., Bradley & Bryant, 1983; for a meta-analysis, see Bus & van IJzendoorn, 1999). Some studies suggest that phonological deficits persist even in adults with childhood diagnosis of dyslexia (e.g., Wilson & Lesaux, 2001; Ransby & Swanson, 2003). However, Castles and Coltheart (2004) emphasized that extant studies providing evidence in favor of a causal relationship between phonological skills and reading skills should be interpreted with caution. They criticized that most of these studies merely show a correlational relationship rather than a causal one and are circular in their argumentation. They also claimed that most longitudinal and training studies fail to meet the necessary criteria to unequivocally ascribe success in reading acquisition to good phonological awareness skills or to phonological awareness trainings. They stated that in terms of a causal relationship, for example, phonological awareness trainings should improve reading skills specifically, i.e. “only reading-related skills” (2004, p. 76) should benefit from the training. Another claim is that there must be no letter-sound knowledge at all prior to phonological awareness training to not confound training effects with “implicit reinforcement of pre-existing reading skills” (2004, p. 99). Given that most studies fail to meet these and other critical criteria, Castles and Coltheart concluded that the causal relationship between phonological skills and reading performance still needs to be replicated in future research. However, Hulme et al. (2005) criticized Castles and Coltheart’s (2004) “conception of causation [as] overly narrow” (2005, p. 360). They argued that effects of phonological skills on reading development might be moderated or mediated by other reading-related skills such as letter-sound knowledge, but these influences do not preclude the importance of phonological skills in reading acquisition and development.

Remarkably, the close relationships of phonological deficits and poor reading comprehension are cross-linguistically evident in poor readers of languages other than English. Wimmer (1996) and Ziegler et al. (2003) found that 9- to 13-year-old dyslexic readers in German completed non-word reading tasks as slowly as English dyslexic readers (and more slowly compared to word reading tasks). However, German dyslexic readers performed with notably higher accuracy on non-word reading tasks compared to English
dyslexic readers of the same age. The authors attribute the higher accuracy of German
dyslexic readers to the transparent orthography of German. The grapheme-phoneme-
conversion rules are highly consistent in the German language. Thus, phonological recoding
is much easier in German compared to languages with an opaque orthography such as English
and is therefore acquired earlier (Wimmer & Goswami, 1994). As a result, even dyslexic
readers in German have little difficulties reading non-words accurately, but they lack the
necessary automaticity to read non-words with little cognitive effort as indicated by long
reading times. Wimmer concluded that the deficit underlying poor reading performance is a
phonological deficit in both languages (see also Mayringer & Wimmer, 2000), but this deficit
is somewhat differently expressed in German than in English poor readers.

A possible objection concerning the generalizability of previous findings is that many
investigations concentrated on beginning readers. Beginning readers are bound to rely
primarily on the non-lexical phonological recoding route when recognizing words, because
most written word forms are unknown for them. Hence, for beginning readers, most of the
variance in reading comprehension skills is not surprisingly explained by phonological
recoding skills. However, more experienced readers increasingly make use of the more
efficient (lexical) route of orthographical decoding, depending on the size, quality, and
accessibility of their sight vocabulary (Frith, 1985). Thus, orthographical decoding skills
during the primary school years become an increasingly important source of individual
differences in reading comprehension (although phonological recoding skills remain a strong
and unique predictor even in Grade 4, Knoepke, Richter, Isberner, Naumann, & Neeb, 2014).

**Individual Differences in Orthographical Decoding**

Several studies suggest that a deficit in orthographical decoding, also called surface
dyslexia, can cause severe reading comprehension problems as well. Castles and Coltheart
(1993) disentangled both types of word recognition deficits using non-word and exception-
word reading. Because phonological recoding skills are required for non-word reading and
orthographical decoding skills are required for exception-word reading, poor readers with a
phonological deficit should exhibit difficulties reading non-words but less difficulties reading
exception words. In contrast, poor readers with a deficient orthographical decoding route
should exhibit difficulties reading exception words but fewer difficulties reading non-words.
This pattern of double dissociation was obtained in two experiments. In the first experiment,
Castles and Coltheart (Exp. 1) investigated 8- to 14-year-old dyslexic readers’ performance
on non-word and exception-word reading tasks and found that 85% of the dyslexic readers
showed the expected double dissociation. Either their non-word reading skills were significantly poorer than would be expected based on their exception-word reading performance (55%) or their exception-word reading performance was significantly poorer than would be expected based on their non-word reading performance (30%). Thirty four percent of the dyslexic readers even performed poorly on just one of the tasks, whereas they exhibited no difficulties at all with the other task. In their second experiment (Exp. 2), Castles and Coltheart found that readers performing poorly on exception-word reading had no problems comprehending spoken exception words, ruling out an alternative explanation in terms of general language deficits (similar results were obtained by Manis et al., 1996).

Some evidence exists showing that the prevalence of the two types of deficits depends on language-specific differences. As noted earlier, several studies suggested that dyslexic readers’ phonological recoding is slow but reliable in transparent orthographies such as German, in contrast to opaque orthographies such as English (e.g., Mayringer & Wimmer, 2000; Wimmer, 1996; Ziegler et al., 2003). Complementing these findings, more recent studies indicated that dyslexic readers’ orthographical decoding route is more likely to be deficient in transparent orthographies (e.g., Martens & de Jong, 2006; Zoccolotti et al., 2005). The word-length effect has been used to investigate this deficit. When recognizing words via the non-lexical, phonological recoding route, i.e. by means of grapheme-to-phoneme-conversion, the length of written-word forms is positively related to the time it takes to recognize the word. However, when words are recognized via the orthographical decoding route, whole-word forms are directly mapped on to their respective lexical entries, and word length has no impact on word recognition times. In a word-naming study based on this logic, Zoccolotti et al. (2005) found that skilled Italian readers’ sensitivity to word length decreased from Grade 1 to Grade 2, suggesting a shift from phonological recoding to orthographical decoding. In contrast, dyslexic third graders were as sensitive to word length during word and non-word naming as first graders indicating that they still primarily relied on phonological recoding. Similar results were obtained in Dutch by Martens and de Jong (2006) and in German by Ziegler et al. (2003; for additional evidence suggesting a strong relationship between orthographical decoding skills and text comprehension in German primary school children, see Knoepke et al., 2014).

These findings clearly indicate that conceptualizing dyslexia as a purely phonological deficit fails to explain the variety of poor readers. At least two types of word recognition deficits exist, a more phonologically-based and a more orthographically-based deficit that can underlie reading comprehension problems (e.g., Castles & Coltheart, 1993; Manis et al.,
Sources of Poor Reading Comprehension

This distinction has implications for remediation and intervention programs. The assumption that phonological or grapheme-phoneme-conversion trainings suggested by phonological-core deficit models of dyslexia would work equally well for all poor readers is unreasonable. Instead, testing poor readers on a broader range of word-recognition skills is essential to determine their specific training needs.

Dual-route models make important contributions to the description and explanation of visual word recognition processes, their acquisition and development, various types of word recognition deficits, and language-specific differences with respect to opacity and transparency, but other theoretical approaches of visual word recognition also exist that reject the idea of two functionally distinct routes. Instead they model word recognition in a single information-processing network as in, for example, the parallel-distributed-processing (PDP) model (Seidenberg & McClelland, 1989) or the connectionist triangle model (Plaut et al., 1996). These models explain and predict the various types of deficits in visual word recognition by impaired distributed representations or “computational resource limitations” (e.g., Manis et al., 1996, p. 189), by impaired network pathways (e.g., Plaut, 1999), or by impairment of neurological areas involved in the network responsible for reading (e.g., Woollams, 2014). In many cases, these models make similar predictions as dual route models. Thus, deciding among these different approaches is difficult based on the available evidence.

Dual-route models of visual word recognition have been designed to explain reading acquisition, development, and disorders in Indo-European languages with alphabetic scripts such as German, English, and Spanish. Consequently, this approach is probably not suitable to fully explain word recognition processes and thus the relationships between visual word recognition and reading comprehension skills in languages with non-alphabetic scripts such as Chinese and Japanese (for a more detailed discussion on universal principles of visual word recognition, see e.g. Frost, 2012).

Individual Differences in the Quality of and Access to Meaning Representations

The retrieval of word meanings is an additional word-level source of reading comprehension problems. The retrieval of word meanings is the basis of text comprehension, suggesting that individual differences in the mastery of this process are a proximal predictor of reading comprehension problems (Richter et al., 2013). According to Perfetti and Hart’s lexical quality hypothesis (2001, 2002; Perfetti, 2007), lexical representations comprise not only formal properties of words (such as the word’s phonology or orthography) but also meaning representations. Moreover, the overall quality of a lexical representation depends on
the qualities of these components and their interconnectedness. If one of them is not (fully) specified, the lexical representation is lower in quality. A substantial amount of low-quality lexical representations will hamper reading comprehension (Perfetti & Hart, 2001, 2002).

In a study with adult readers of varying reading comprehension skills, Perfetti and Hart (2001) demonstrated that the skilled and poor readers differed in the quality of their meaning representations. The participants were presented with written word pairs such as king – royalty (2001, p. 76) and were required to decide whether the words were semantically related. The word pairs appeared word-by-word with differing inter-stimulus intervals and contained either a homophone, such as night (homophonic meaning: knight) in night – royalty (2001, p. 76), or no homophone. The authors expected skilled readers to make faster decisions and to show an earlier interference effect for homophones compared to poor readers. They reasoned that skilled readers have faster access to word meanings because of their superior meaning representations. In line with this assumption, they observed faster decision times and earlier interference effects in the presence of homophones for skilled compared to poor readers.

In a cross-sectional study with primary school children from Grade 1 to 4, Richter et al. (2013) directly tested the assumption that the quality of meaning representations is a proximal predictor of reading comprehension at the text level. The children were presented with tasks accessing the quality of their phonological representations (phonological comparison task), their orthographical representations (lexical decision task), and their meaning representation (semantic verification task), as well as their reading comprehension skills at the text level (ELFE 1-6, Lenhard & Schneider, 2006). The results indicate that the overall quality of the children’s lexical representations and the efficiency of access to these representations explained a substantial amount of variance in their reading skills. Moreover, the effect of the quality of phonological and orthographical representations on reading comprehension was found to be mediated by the quality of meaning representations. Notably, individual differences in the quality of meaning representations accounted for a substantial amount of variance in reading comprehension that could not be explained by variance in word recognition skills (Richter et al., 2013). A study by Nation and Snowling (1998) suggests a similar conclusion by showing that semantic deficits can explain word recognition and reading comprehension problems in poor readers with normal phonological recoding skills.

Nation and Snowling (1999) used a priming paradigm to demonstrate qualitative differences in the abstract semantic knowledge of children classified as good vs. poor readers. In a priming experiment, the good readers made faster lexical decisions on target words (e.g.,
Sources of Poor Reading Comprehension

When a prime of the same category (dog, 1999, p. B1) preceded the target word than when they were preceded by an unrelated word. However, poor readers’ responses were primed by preceding category members only when prime and target words were highly associated. In contrast, both good and poor readers showed comparable priming effects when prime and target words were functionally related (e.g., shampoo – hair, 1999, p. B1). The authors assumed that the poor readers primarily possessed an event-based semantic word knowledge, whereas the better readers had already built abstract semantic representations.

In sum, a number of studies using different methods and focusing on different age groups indicate that a low quality and accessibility of word-meaning representations can cause reading comprehension problems in addition to the deteriorating effects of deficits in phonological recoding and orthographic decoding.

Individual Differences Beyond the Word Level

The explanatory approaches of poor reading comprehension skills discussed in the previous sections attribute poor reading abilities primarily to word-level skills. Word-level processes are clearly a major source of reading comprehension difficulties, but the existence of readers who show poor reading comprehension despite adequate word reading skills suggests that cognitive processes must be considered in addition to the word-level to better understand reading comprehension difficulties (e.g., Cain et al., 2001; Nation & Snowling, 1998, Exp. 2, 1999; Stothard & Hulme, 1992). Several studies have demonstrated, in accordance with the simple view of reading, that a substantial amount of variance in reading comprehension can be explained by individual differences in general language (listening) comprehension (Catts et al., 2003; Johnston & Kirby, 2006; Joshi & Aaron, 2000; Kendeou et al., 2009; Knoepke et al., 2013; Ransby & Swanson, 2003). These language comprehension skills comprise several component skills at the sentence and text level. In the following section, we will discuss studies that examined the potential impact of some of these component skills on reading comprehension problems. The studies included children with adequate word recognition but impaired comprehension skills or they controlled for word recognition skills statistically to investigate the unique contribution of sentence- and text-level skills to individual differences in reading comprehension.
Individual Differences in Syntactic and Semantic Integration Processes

To comprehend a sentence, simply decoding the words of the sentence and retrieving their meanings is not sufficient. The reader must integrate the individual word meanings into a coherent mental representation of the sentence according to its specific syntactic and semantic structure (e.g., Müller & Richter, 2014; Richter & Christmann, 2009). For example, the sentence *Katie sues Robert* contains exactly the same words as the sentence, *Robert sues Katie*. Based on the word meanings alone, a reader cannot determine the prosecutor and the respondent in the sentence. However, the syntactic structure of transitive English main clauses (subject-verb-object) reveals that in the first sentence *Katie* is the prosecutor and in the second sentence she is the respondent. In addition to the syntactic structure, a reader can also use the semantic context of a sentence to resolve, for example, syntactic or semantic ambiguities. In the sentence, *the bug has been killed/removed*, the interpretation of *bug* as either an insect or a technical error depends entirely on the semantic context of the sentence (insect: *killed*; technical error: *removed*).

Ample evidence exists showing a relationship between individual differences in syntactic and semantic integration processes and reading comprehension. For example, poor syntactic awareness, i.e. a reader’s “ability to reflect upon and to manipulate aspects of the internal grammatical structure of sentences” (Tunmer et al., 1987, p. 25) and deficient processes of semantic integration can result in reading difficulties in some poor readers. Byrne (1981) found a positive relationship between syntactic awareness and reading comprehension in poor vs. good beginning readers. In an act-out-task, the children were presented with spoken sentences, which were the same length but differed in grammatical structure complexity. In addition, children worked on a picture-choice task with spoken sentences varying in plausibility containing center-embedded relative clauses. Pictures matching plausible sentences were easy to find with the aid of the semantic context of the sentence, but pictures matching less plausible sentences required the aid of syntactic knowledge for their correct identification. The poor readers’ performance on the syntactically more complex sentences in the act-out task and on the less plausible sentences in the picture-choice-task was inferior to the good readers’ performance on these sentences. In contrast, the between-group performances were comparable for the less complex and plausible sentences. Similarly, Tunmer et al. (1987) found that older poor readers were less able to correct spoken sentences containing morphological or word-order violations or to supply a missing word in an auditory presented sentence compared to younger skilled readers of the same reading level. Poor readers also seem to have difficulties restructuring the words of a scrambled sentence.
Sources of Poor Reading Comprehension

back into their correct order (Nation & Snowling, 2000) and to perform poorly on Bishop’s (1983) test for the reception of grammar (TROG; Stothard & Hulme, 1992). In a longitudinal study with French children from Kindergarten to Grade 2, Casalis and Louis-Alexandre (2000) found that morpho-syntactic skills in Kindergarten, such as the ability to inflect nouns for gender or verbs for tense form, are predictive of sentence comprehension at the end of Grade 2. Plaza and Cohen (2003) demonstrated that syntactic awareness operationalized by a grammatical judgment and correction task was predictive of reading and spelling skills in French primary school children at the end of Grade 1. Moreover, syntactic awareness accounted for unique variance in reading and spelling even when phonological awareness, naming speed, and auditory memory were statically controlled. These studies suggest that individual differences in syntactic awareness and syntactic integration skills explain unique variance in reading comprehension and that deficient syntactic skills might cause reading difficulties.

In a reading time study with adult readers, Graesser et al. (1980) found that the syntactic complexity and the semantic complexity of sentences (independent from each other) had a greater retarding impact on slow readers compared to fast readers. Considering that the slower readers are likely to have lower reading skills, this finding suggests that poor readers need to invest a greater amount of cognitive resources to comprehend syntactically and semantically complex sentences. Investigating semantic integration skills, Hannon and Daneman (2004) found that less skilled readers tend to invest less cognitive effort in the establishment of coherence relations within a sentence in favor of establishing more global coherence relations. They presented poor and skilled readers with short texts containing a semantic anomalous term in the final sentence of the text, such as Amanda was bouncing all over because of too many tranquilizers/ sedatives/ tranquilizing sedatives/ tranquilizing stimulants (2004, p. 197). Poor readers were less likely to detect anomalies than skilled readers and they were less likely in particular to detect anomalies in internally incoherent noun phrases (e.g., tranquilizing stimulants) compared to internally coherent noun phrases (e.g., tranquilizing sedatives), indicating a rather shallow semantic processing of the meaning of noun phrases and sentences in poor readers.

Semantic information, in particular the semantic context of a sentence, can also be beneficial for poor readers with deficits in word-level processes, because the context helps these readers to recognize the words and infer their meaning. This explanation is the basic assumption of the interactive-compensatory model proposed by Stanovich (1980). The model is based on evidence from a number of inventive experiments that compared the word-
recognition performance of good vs. poor readers under different contextual manipulations (see also West & Stanovich, 1978). These experiments consistently revealed that the performance of the poor readers depended more heavily on the presence of a facilitating sentence context, whereas the good readers relied on their superior word-recognition skills rather than the sentence context. In a similar vein, Gernsbacher and Faust (1991, Exp. 4) demonstrated that poor readers extensively use a restricting semantic context when it facilitates word recognition (for similar results for dyslexic readers, see Nation and Snowling 1998, Exp. 2). Van der Schoot et al. (2009) found in an eye-tracking (Exp. 1) and in a self-paced reading study (Exp. 2) that poor 10- to 12-year-old Dutch readers used prior contextual information as effectively as skilled readers to resolve lexical ambiguities. However, in contrast to skilled readers, poor readers were less likely to correct an initial incorrect interpretation of an ambiguous word, indicating less efficient comprehension monitoring in poor readers.

Importantly, Gernsbacher and Faust (1991, Exp. 1) showed that poor readers have difficulties to suppress context-inappropriate meanings. The task was to judge the semantic relatedness of a sentence and a word that was presented after the final word of the sentence (e.g., He had a lot of patients). Poor readers showed a substantial and long-lasting interference effect in rejecting a probe word (calm) when it did not fit the sentence but was semantically related to a homophone of the final word (patience). In contrast, good readers exhibited this interference effect only when the probe word was presented immediately after the sentence. These results suggest an effective and rapid suppression of inappropriate word meanings by good but not poor readers.

In sum, the findings of the reported studies suggest that efficient syntactic and semantic integration processes are an important prerequisite for good text comprehension. If these processes are ineffective or deficient, the overall reading ability may be adversely affected.

**Individual Differences in Inference Making and Comprehension Monitoring**

Text comprehension goes beyond the sentence level by requiring the integration of information provided by several sentences into a coherent mental representation. According to Johnson-Laird (1981) and Van Dijk and Kintsch (1983), this mental representation consists of two qualitatively distinct levels. Readers need to construct a coherent representation of the semantic structure of the text (propositional text base), and they need to integrate text information and prior knowledge to build a mental model (Johnson-Laird, 1981) or situation
model (Van Dijk & Kintsch, 1983) of the circumstances described in a text. Thus, constructing a situation model (mental model) is essential for comprehending the text, and it requires several closely related cognitive activities, such as linking the contents of adjacent and distant sentences (Singer et al., 1992), using prior knowledge for drawing inferences (Graesser et al., 1994), predicting upcoming text (Van Berkum et al., 2005), monitoring the plausibility of the text content (Isberner & Richter, 2013), and monitoring the comprehension process (Nation, 2005). The key question is whether individual differences in these processes explain unique variance in overall reading comprehension in addition to readers’ word recognition skills. In a longitudinal study, Oakhill et al. (2003) and Cain et al. (2004) focused on the unique contribution of inference skills and individual differences in comprehension monitoring to reading comprehension. Inference skills can be defined as the ability to derive information from the text context and from world knowledge to enrich the mental representation of the text. Comprehension monitoring skills can be defined as the metacognitive ability to monitor the comprehension process and to detect comprehension problems as well as inconsistencies with the text or with prior knowledge (Baker, 1989). Oakhill et al. (2003) and Cain et al. (2004) presented children with several tasks that assessed inference-making skills, comprehension monitoring skills, verbal ability, working memory skills, and overall text comprehension. The ability to draw inferences and to monitor their comprehension process explained unique variance in reading comprehension even when verbal ability and single word recognition abilities were statistically controlled. These relationships were found in beginning readers aged 7 to 8 years (Oakhill et al., 2003) and also in older readers until the age of 11 (Cain et al., 2004). Although a substantial amount of variance in reading comprehension was explained by working memory capacity, this general cognitive ability failed to fully explain the effects of inference making and comprehension monitoring on reading comprehension. Instead, both higher-order cognitive component skills of text comprehension accounted for a unique portion of variance in children’s reading comprehension. In accordance with these findings, Van der Schoot et al. (2009) demonstrated that poor readers were less able to monitor their comprehension process than skilled readers. In contrast to good comprehenders, poor readers’ reading times on disambiguating information that followed a lexically ambiguous word were the same as when the information preceded the word. Moreover, they made more errors responding to comprehension questions when a lexically ambiguous word with a biased (not intended) meaning preceded the disambiguating region. The authors concluded that the poor readers are less likely to detect an interpretation error (as indicated by the lack of reading time increase on the disambiguating
Sources of Poor Reading Comprehension

The impact of inference skills on text comprehension has received ample attention in research. Bridging inferences that connect two pieces of information in a text, such as anaphoric (e.g., Garnham & Oakhill, 1985) and causal inferences (e.g., Singer et al., 1992), are especially important for constructing a coherent situation model. Cain and Oakhill (1999) and Cain et al. (2001) focused on individual differences in such text-connecting inferences and elaborative or gap-filling inferences, which refer to processes of “incorporating information outside of the text, i.e. general knowledge, with information in the text to fill in missing details” (Cain et al., 2001, p. 490). Seven- to 8-year-old children read short text passages and answered questions requiring the identification of literal assertions in the text, making text-connecting and gap-filling inferences. Cain and Oakhill (1999) found that poor readers drew fewer inferences of both types than good readers, whereas both groups performed equally well on literal assertions. To rule out the possibility that the poor readers’ inferior performance on the inference questions was due to a lack of necessary background knowledge, Cain et al. (2001) replicated the findings holding background knowledge constant. In this study, they provided children with background knowledge about a fictional planet named Gan to ensure that all children had the same background. As in the Cain and Oakhill study, the poor readers had significantly more difficulties drawing text-connecting and gap-filling inferences than the good readers. Moreover, poor readers’ performance on the inference questions could not be attributed to a lack of background knowledge.

These findings consistently suggest that word-level and text-level skills independently contribute to text comprehension variance. Oakhill et al. (2003) emphasized that determining the exact causes of reading difficulties and considering this individual pattern of deficits when planning remediation and intervention programs for poor readers is essential. Ideally, educators should take care to tailor such programs as accurately as possible to the needs and deficits of the individual reader. To accomplish this, the gross screening instruments that are typically used for diagnosing reading difficulties need to be augmented with more discriminative psychological tests that assess component skills of reading comprehension. One promising way to assess these skills is to measure the efficiency of the specific component processes of reading comprehension by using reaction-time measures in combination with well-defined reading tasks and test items that are constructed according to (psycho-)linguistic criteria (for an example, see the German-speaking test battery ProDi-L, Richter et al., in press).
Conclusion

This chapter discussed several problems concerning the common definition of dyslexia and its diagnostic value in identifying poor readers and their individual needs for training and intervention. In particular, we emphasized that the diagnosis of dyslexia bears no information about the cause of the individual reading deficit or the kind and extent of intervention that is required. Furthermore, we argued that the distinction between dyslexic and general backward readers based on the wide-spread discrepancy model of dyslexia is not empirically useful. One argument against the discrepancy model is that poor readers classified as dyslexic according to the discrepancy model perform the same on reading-related tasks as poor readers with a more general cognitive deficit. Consequently, both groups receive the same reading intervention. In that respect, a cognitive perspective on reading difficulties that examines component processes of reading at the word-, the sentence-, and the text level is far more promising. Even readers in the same age group differ greatly in the extent that they accurately and efficiently master these cognitive processes at all three levels. We argued that individual differences in word-, sentence- and text-level processes contribute uniquely to individual differences in reading comprehension. Against this background, we conclude that the potential causes for reading difficulties are multifaceted and heterogeneous. An important practical implication of this conclusion is that the success of intervention and remediation programs depends heavily on the identification of the specific type of cognitive deficit that causes reading difficulties in the poor reader.

Acknowledgments. The work on this chapter was supported by the German Federal Ministry of Research and Education (BMBF, grants 01GJ0986 and 01GJ1402B).
Chapter VII

General Discussion
General Discussion

The aim of this dissertation was to advance a more comprehensive understanding of the cognitive component processes involved in reading comprehension, their interrelatedness, and their development in beginning German readers. As reviewed and discussed in Chapter I, reading research has identified cognitive component processes at the word, sentence, and text level (e.g., Lenhard & Artelt, 2009; Müller & Richter, 2014; Perfetti, 1999, 2001; Perfetti et al., 2005; Richter & Christmann, 2009) that constitute reading comprehension. These processes include phonological recoding, orthographical decoding, access to word meanings, syntactic and semantic integration, establishing mental relations of local and global coherence, drawing knowledge-based inferences, and comprehension monitoring. I argued that successful reading comprehension depends on the successful accomplishment of these component processes and that comprehension can be impeded when one or more of these processes are inefficient or deficient. This dissertation contributes to the existing literature and empirical research on reading comprehension by exploring a subset of open research questions on the cognitive component processes of reading comprehension, their interrelatedness, and their development in beginning readers. In the following sections, the findings of the four empirical studies reported in this dissertation are summarized and discussed against the background of extant empirical studies and established models of reading comprehension and reading development.

Summary and Discussion of Study 1

The first empirical study presented in Chapter II examined the extent that phonological recoding skills (i.e., the ability to recognize written word forms via grapheme-to-phoneme translation) and orthographical decoding skills (i.e., the ability to recognize written word forms directly) uniquely contribute to reading comprehension at the sentence and text level and how these two skills develop in children learning to read in German. A cross-sectional study with German 2nd to 4th graders revealed that individual differences in both skills of visual word recognition explain unique and substantial amounts of variance in sentence and text comprehension in German primary school children across all three grade levels. This finding is consistent with the assumption of two routes of visual word recognition as proposed by Coltheart et al.’s (2001; Coltheart, 2005) DRC model. Orthographical
decoding explained more variance in reading comprehension (52%) than phonological recoding (26-36%) at all three grade levels, suggesting that by the end of Grade 2, German primary school children’s sight vocabulary is already well developed and that they recognize the majority of written word forms in age-appropriate texts directly via orthographical decoding. These results are in line with previous studies showing that visual word recognition skills are acquired early in German readers. Because of the high orthographic consistency in German, grapheme-to-phoneme conversion rules are easily acquired. Thus, primary school children’s phonological recoding skills are usually highly developed by the end of Grade 1 (e.g., Aro & Wimmer, 2003; Wimmer & Goswami, 1994). The ability to recode new and unfamiliar word forms early during reading acquisition allows children to build a sight vocabulary early on. Consequently, German beginning readers are able to recognize many written word forms in age-appropriate texts directly via orthographical decoding by the end of Grade 2.

One remarkable finding of Study 1 is that phonological recoding skills explain a substantial amount of variance in reading comprehension not only in less experienced readers in Grade 2 but also in more advanced readers in Grades 3 and 4. Moreover, no changes were observed in the relative contributions of phonological recoding and orthographical decoding to reading comprehension across grade levels. This finding conflicts with Frith’s (1985, 1986) three-stage developmental model, which proposes that readers shift completely from the alphabetic stage (phonological recoding) to the orthographic stage (orthographical decoding) only after the alphabetic stage has been mastered. However, our data indicated no shift to an orthographic stage in more advanced readers. Instead, our findings suggest that advanced German readers use both routes of visual word recognition to the same extent as less experienced readers. Our data could arguably reflect a state of intermediate overlap of both stages. Possibly, data from older children would have revealed a complete shift to the orthographic stage. Yet, this explanation seems rather unlikely, given that we found not even a tendency toward such a shift across the grade levels.

In contrast to the three-stage developmental model, the DRC model provides an explanation for the finding that phonological recoding skills contribute to reading comprehension not only in beginning but also in advanced readers. Although advanced readers recognize most written word forms via orthographical decoding, they still recognize word forms via phonological recoding when the word forms are unfamiliar or infrequent. Thus, the assumption is plausible that even advanced readers make use of both routes of visual word recognition during reading. However, the DRC model also implies that
developing readers recognize written word forms increasingly via orthographical decoding and decreasingly via phonological recoding to the extent that reading experience increases and sight vocabulary grows. Accordingly, a slight change should have occurred in the relative contributions of phonological recoding and orthographical decoding to reading comprehension across grade levels. Instead, their relative contributions did not change from Grade 2 to 4, indicating that the readers in our study made use of phonological recoding irrespective of their reading experience or their familiarity with particular words. This interpretation is consistent with Frost’s (1998) strong phonological model and Plaut et al.’s (1996) connectionist triangle model, which assume that phonological information is always and automatically activated along with orthographical and semantic information during visual word recognition.

The stable relationship between phonological recoding skills and reading comprehension across grade levels seems plausible, because German is characterized by highly consistent grapheme-to-phoneme mappings that make phonological recoding a highly reliable tool to recognize written word forms (e.g., Landerl et al., 1997; Wimmer & Goswami, 1994). Given the greater consistency of German compared to English orthography, the finding is not surprising that German readers rely more strongly and more permanently on phonological recoding skills to recognize written word forms than predicted by the DCR model or the three-stage developmental model, which were predominantly designed to explain processes of reading and reading acquisition in English.

Furthermore, given that phonological recoding skills are easily acquired in German, the finding that individual differences in phonological recoding skills explained comparatively little variance in reading comprehension (26-36%) as opposed to individual differences in orthographical decoding skills (52%) is not surprising. “In languages with a more transparent orthography, efficient grapheme-to-phoneme recoding strategies are acquired more easily […], and early differences in recoding skills should diminish because even impaired readers will successfully acquire such skills early on” (Pfost, 2015, p. 132; see also Landerl & Wimmer, 2008; Landerl et al., 1997; Seymor et al., 2003; Wimmer & Goswami, 1994). Consequently, only little variance in phonological recoding skills can account for variance in reading comprehension. This explanation is also consistent with Gough and Tunmer’s (1986; Hoover & Gough, 1990) simple view of reading, which implies that the power of decoding skills to predict reading comprehension decreases with increasing decoding skills (leaving the unexplained variance in reading comprehension to be accounted for by individual differences in listening comprehension). However, when testing this
assumption, we found that the relationship between phonological recoding skills and sentence and text comprehension was most strongly pronounced for children with highly developed phonological recoding skills and least strongly pronounced for children with poorly developed phonological recoding skills. This finding was somewhat unexpected and in contrast to the predictions derived from the simple view of reading. On closer examination, the weakly pronounced relationship between phonological recoding skills and reading comprehension in poor recoders appeared to be attributed to a few children with exceptionally poorly developed phonological recoding skills. These children were assumed to have exhibited major difficulties with grapheme-to-phoneme conversion and probably resorted to logographic or context information to support their poor phonological recoding skills (e.g., Frith, 1986; Nation & Snowling 1998; Stanovich, 1980; West & Stanovich, 1978; a more detailed discussion of compensatory strategies to support poor cognitive component skills of reading comprehension is provided later in this chapter). Consequently, phonological recoding skills should be weakly related to reading comprehension in these readers. Exclusive of these exceptionally poor recoders, the results might have been more consistent with the simple view of reading, that is, a decreasing relationship between phonological recoding skills and reading comprehension with increasing recoding skills.

Finally, the finding that a substantial amount of variance in reading comprehension remained unexplained by individual differences in phonological recoding and orthographical decoding skills is noteworthy. This is consistent with the assumption that reading comprehension comprises several cognitive component skills not only at the word but also at the sentence and text level (Lenhard & Artelt, 2009; Müller & Richter, 2014; Perfetti, 1999, 2001; Perfetti et al., 2005; Richter & Christmann, 2009). In addition to visually recognizing written word forms, readers need to access the word’s entries in the mental lexicon to retrieve their meanings. At the sentence level, word forms and meanings must be integrated syntactically and semantically to establish a coherent mental representation of the sentence. At the text level, readers must establish mental relations of local and global coherence between adjacent and distant propositions, they must integrate text information with prior knowledge, draw inferences, and monitor their comprehension process. The unexplained variance in sentence and text comprehension, which could not be accounted for by individual differences in phonological recoding and orthographical decoding skills, is most likely attributed to individual differences in these cognitive component skills of reading comprehension. This assumption is further elaborated in the **Summary and Discussion of Study 2**.
In sum, the findings of the first empirical study suggest that both phonological recoding and orthographical decoding make significant and unique contributions to reading comprehension in German 2nd to 4th graders and that their relative weight does not change across grade levels. At all grade levels, orthographical decoding skills explained substantially more variance in reading comprehension than phonological recoding skills suggesting that by the end of Grade 2 German readers have started to build a sight vocabulary and are able to recognize the majority of written words in age-appropriate texts directly via orthographical decoding. Nevertheless, our findings indicate that German readers continuously use phonological recoding to optimize visual word recognition.

Although the findings of our cross-sectional study do not allow for any conclusions about a causal link between readers’ abilities to recognize written word forms and individual differences in reading comprehension, the empirical literature discussed in Chapter I and in Chapter VI strongly suggests that the ability to recognize written words is a necessary prerequisite for successful reading comprehension and that deficient word recognition is a likely source for reading comprehension difficulties (e.g., Castles & Coltheart, 1993; Manis et al., 1996; Rack et al., 1992; Shankweiler et al., 1996; Shankweiler et al., 1999; Wimmer, 1996). Against this background, our findings have important practical implications for reading instruction, the diagnosis of individual deficits in poor readers, and reading intervention. Given that phonological recoding and orthographical decoding skills contribute significantly and uniquely to reading comprehension across all grade levels, practical exercises fostering both word recognition skills can be expected to enhance reading comprehension not only in beginning but also in more advanced readers and thus should be an integral part of reading instruction throughout primary school years. Moreover, the finding that the relative contributions of phonological recoding and orthographical decoding skills in reading comprehension did not change across grade levels suggests that both visual word recognition skills might be predictive of reading comprehension beyond Grade 4. Thus, practicing fundamental skills of visual word recognition could conceivably have a beneficial effect on reading comprehension beyond elementary school years.

The findings in Study 1 also bear important implications for the diagnosis of reading difficulties and the construction of reading intervention programs. Assuming that the relationship between both visual word recognition skills and reading comprehension is a causal one and that successful visual word recognition is at the core of reading comprehension (e.g., Perfetti, 1985; Gough & Tunmer, 1986), the findings suggest that deficits in one or both word recognition skills can cause severe reading comprehension difficulties. Consequently,
when children exhibit poor reading comprehension, their phonological recoding and orthographical decoding skills should be assessed selectively to be able to identify potential deficits in one or both word recognition skills and to be able to determine the severity of the deficits. On this basis, the development of a target-oriented reading intervention that is specially geared to improve deficits in phonological recoding, orthographical decoding, or both skills is possible, and its implementation will likely enhance reading comprehension.

Summary and Discussion of Study 2

The purpose of the second empirical study presented in Chapter III was to test one of the most prominent and influential theories in reading research, Gough and Tunmer’s (1986; Hoover & Gough, 1990) simple view of reading, by using a stringent methodology that overcomes several shortcomings and limitations of extant studies testing the simple view. In a cross-sectional study with German 3rd and 4th graders, the central assumptions of the simple view of reading were examined—the assumption that individual differences in decoding (D) and listening comprehension (C) account for all of the variance in reading comprehension (R) and the assumption that the multiplicative term D x C explains more variance in reading comprehension than the additive term D + C. Although the simple view of reading has been tested several times in the past (e.g., Hoover & Gough, 1990; Joshi & Aaron, 2000; Kendeou et al., 2009), our study stands out because of the optimized and methodologically stringent measures of its components. First, we accounted for the assumption that written word forms can be recognized not only via phonological recoding but also via orthographical decoding as soon as developing readers begin to store familiar word forms in sight vocabulary. The results of Study 1 indicated that both phonological recoding and orthographical decoding skills contribute to reading comprehension in Grades 2 to 4. The results of Study 1 also suggest that by the end of Grade 2 German primary school children’s sight vocabulary is already well developed and that they recognize the majority of written word forms in age-appropriate texts directly via orthographical decoding. Accordingly, we operationalized decoding (D) through a lexical decision task that assesses orthographical decoding skills. Although regular words in the lexical decision task can be recognized (and non-words can be rejected) correctly via both phonological recoding and orthographical decoding processes, irregular word forms can be recognized and pseudo-homophones can be rejected correctly only via orthographical decoding processes (see the description of the ProDi-L lexical decision task in Chapter I). The second aspect of the study that stands out is the use of parallel tasks and materials to assess reading and listening comprehension, which differed only in the modality of
presentation. This method was based on the formula \( R = D \times C \), which denotes that processes of reading and listening comprehension are basically the same after written word forms have been decoded. The advantage of the method is that it rules out the possibility of a task confound. Finally, not only was the reliability of the critical cognitive processes assessed but also the degree of \( D, C, \) and \( R \) automatization by recording response accuracies and response latencies for all three measures. A second aim of the study was to test the generalizability of the simple view of reading to German readers in Grades 3 and 4.

The results of the second empirical study provided only weak support in favor of the simple view of reading, which questions the central assumptions of the simple view and its generalizability to German developing readers. Although individual differences in decoding and listening comprehension explained a significant and substantial amount of variance in reading comprehension (28-38%), they were far from explaining all of the variance in reading comprehension as predicted by the simple view of reading. More than 60% of the variance in reading comprehension remained unexplained by the components of the model. This result differs with the findings from previous studies showing that more than two thirds of the variance in reading comprehension can be accounted for by individual differences in decoding and listening comprehension skills (e.g., Braze et al., 2007; Hoover & Gough, 1990; Johnston & Kirby, 2006).

Study 2 also revealed no evidence that the product term \( D \times C \) provides an increment to the explained variance in reading comprehension in addition to the additive term \( D + C \). This finding is inconsistent with the simple view of reading but consistent with findings reported by Chen and Vellutino (1997), Georgiou et al. (2009), and Savage and Wolforth (2007) who found that a multiplicative and an additive model predicted reading comprehension equally well. The only finding of Study 2 that could be interpreted in favor of a multiplicative model was the significant three-way interaction of decoding, listening comprehension, and grade level, which explained an additional 2.7% of the variance in reading comprehension when integrated test scores were used as a dependent variable. Upon closer examination, the relationship between listening and reading comprehension was found to increase with increasing decoding skills at Grade 4. This interaction is consistent with the simple view of reading, which implies that individual differences in reading comprehension must be attributed to individual differences in listening comprehension when decoding skills are highly developed. However, when we estimated the two-way-interaction of decoding and listening comprehension for both grade levels separately, the interaction term \( D \times C \) did not reach significance, neither at Grade 3 nor at Grade 4. At Grade 3, the relationship between
listening and reading comprehension even tended to decrease with increasing decoding skills, a pattern opposed to the multiplicative model of the simple view of reading.

Overall, our findings conflict with the simple view of reading in its most simplistic form and with several empirical studies that provide evidence in favor of a strong or extended version of the simple view (e.g., Braze et al., 2007; Hoover & Gough, 1990; Johnston & Kirby, 2006; Kendeou et al., 2009). The most likely explanation for the obtained differences between extant studies and the current results is the use of optimized, methodologically stringent measures of $D$, $C$, and $R$. First, unlike several other studies (e.g., Braze et al., 2007; Hoover & Gough; Joshi & Aaron, 2006), we used a lexical decision task (assessing orthographical decoding skills) instead of a non-word reading task (assessing phonological recoding skills) to measure decoding skills. Second, we used strictly parallel tasks and materials to assess reading and listening comprehension skills in contrast with previous researchers who had used different tasks (e.g., Joshi & Aaron, 2000; Kendeou et al., 2009). Finally, we assessed not only the reliability but also the degree of automatization of $D$, $C$, and $R$ by recording response accuracy and response latencies for all three measures. Considering these major methodological differences in testing the simple view, the discrepancy in results is not surprising. However, a stringent methodology does not directly explain why individual differences in decoding and listening comprehension explained only 28-38% of the variance in reading comprehension in our study. Given that the lexical decision task provides a highly appropriate and optimized measure of decoding skills in German 3rd and 4th graders and given the high similarities between tasks and materials used to assess reading and listening comprehension skills, the results arguably should have revealed an even stronger relationship between $D \times C$ and $R$ as compared to extant studies. Instead, more than 60% of the variance in reading comprehension remained unexplained.

A possible explanation might be that the readers in our study were able to compensate for poor decoding or poor comprehension skills in the reading comprehension task. Spoken sentence pairs were presented only once in the listening comprehension task, but readers had ample time to process the written sentence pairs in the reading comprehension task (see the description of the ProDi-L semantic verification task with sentence pairs in Chapter I). They could have slowed or paused reading, looked back, or reread the sentence pairs to compensate for poor decoding or comprehension skills (see the compensatory-encoding model, Walczyk, 1995; Walczyk, Marsiglia, John, & Bryan, 2004; see also the discussion in Kirby & Savage, 2008). Consequently, readers with inefficient decoding or comprehension skills might have performed better than expected on the reading comprehension task, which could have
General Discussion

*diminished the relationship of decoding and comprehension skills with reading comprehension (Walczyk et al., 2004). This discussion is elaborated in Study 3 discussion below.*

Another possible explanation for the considerable amount of unexplained variance is we tested the simple view of reading in a German sample with German materials. The majority of extant studies (e.g., Braze et al., 2007; Georgiou et al., 2009; Hoover & Gough, 1990; Johnston & Kirby, 2006; Joshi & Aaron, 2000; Kendeou et al., 2009; Tunmer & Chapman, 2012; Vellutino et al., 2007) tested the simple view of reading with English samples and materials. Unlike English, German is characterized by highly consistent grapheme-to-phoneme conversion (e.g., Landerl et al., 1997; Wimmer & Goswami, 1994; see also Summary and Discussion of Study 1), which facilitates and accelerates the acquisition of visual word recognition skills in German compared to orthographically opaque languages such as English (Aro & Wimmer, 2003; Wimmer & Goswami, 1994). As a consequence, individual differences in visual word recognition diminish early during reading acquisition in German (e.g., Pfost, 2015) and thus cannot account for much variance in reading comprehension. In this case, the simple view of reading predicts that individual differences in listening comprehension should account for most of the obtained variance in reading comprehension. Consistent with this assumption, Florit and Cain (2011) and Tobia and Bonafacci (2015) found that listening comprehension skills were a stronger predictor of reading comprehension than decoding skills in orthographically transparent languages. However, listening comprehension did not account for the remaining variance in reading comprehension in our study as predicted by the simple view of reading. Instead, more than 60% of the variance in reading comprehension remained unexplained by both decoding and listening comprehension skills.

In sum, our findings suggest that the simple view of reading in its most simplistic and strongest version could not withstand a methodologically stringent test, which could have strong theoretical implications. The substantial amount of unexplained variance in reading comprehension indicates that the model might need to be modified or elaborated with additional components. Joshi and Aaron (2000), for example, suggested to extend the simple view of reading formula \( R = D \times C \) by a processing speed component \( S \) to account for automatic word recognition processes in more advanced readers (see also Johnston & Kirby, 2006). However, including an additional speed component into the model seems pointless when an appropriate operationalization of \( D \) is used. The lexical decision task in our study primarily assessed the ability to recognize written word forms via more automatized
orthographical decoding processes. Moreover, integrated test scores were calculated from response accuracy and response latency measures of $D$, $C$, and $R$ (mean response accuracy divided by mean log-transformed response latency). These integrated test scores provide an optimal measure of both the reliability and the degree of automatization of cognitive processes. In this case, a further speed component would seem to be redundant.

Braze et al. (2007) proposed to include vocabulary size as another component into the model. In a study with young adults, they found that $D$ and $C$ accounted for most of the variance in $R$ (76%), as predicted by the simple view of reading, but vocabulary size explained another 6% beyond the contributions of $D$ and $C$. However, Tunmer and Chapman (2012) demonstrated in an exploratory factor analysis that vocabulary knowledge “is a component of oral language comprehension (i.e., $C$)” (p. 463). A possible explanation for Braze et al.’s finding might be that “the measure used to assess linguistic comprehension in the SVR model […] did not adequately assess the vocabulary component of $C$” (p. 460). Moreover, Tunmer and Chapman argued that the effect of vocabulary size would have vanished if Braze et al. had used context-free word recognition instead of non-word reading as measure of $D$. They found vocabulary size to be strongly correlated with word recognition but not with non-word reading skills. Given that the measure of $D$ in our study included context-free word recognition, an additional vocabulary component would have unlikely accounted for the unexplained variance in $R$.

A more promising extension of the simple view of reading would be a component accounting for compensatory strategies. As mentioned earlier, compensatory strategies are more easily applied during reading than listening comprehension. Readers can slow or pause reading, look back, or reread the text to compensate for inefficient decoding or comprehension skills (Walczyk, 1995; Walczyk et al., 2004). When reading longer written texts, readers can also use “headings and typographical cues” (Kirby & Savage, 2008, p. 79) to support reading comprehension. Under these conditions, even readers with poor cognitive component skills can perform well on text comprehension tasks. As a result, the relationship between the component skills of reading and reading comprehension would diminish. Thus, an additional component that accounted for readers’ use of compensatory strategies might have explained further variance in reading comprehension in Study 2. A conceivable alternative would be to design reading comprehension tasks in a way that prevents the use of compensatory strategies, for example, performing reading tasks under time-pressure (Walczyk, 1995; Walczyk et al., 2004). If readers are limited in their use of compensatory strategies during reading comprehension, $D$ and $C$ might account for more than 28-38% of the
variance in reading comprehension. Further research is warranted to show whether the simple view of reading can withstand a methodologically stringent test with a time-restricted measure of $R$.

Our finding that the multiplicative term $D \times C$ explained no additional variance in $R$ in addition to the additive term $D + C$ supports an additive as opposed to a multiplicative combination of $D$ and $C$. However, an additive model is meaningful only for children with at least rudimentary decoding and listening comprehension skills, which most of the children had in Study 2. For children without or with exceptionally poorly-developed decoding or listening comprehension skills, an additive model $R = D + C$ seems counterintuitive, because it implies that a moderate level of reading comprehension might not necessarily require both decoding and listening comprehension skills (Kirby & Savage, 2008). Reading comprehension would seem unimaginable if either $D$ or $C$ were not at all evident. This reasoning also holds for a model including a combination of the multiplicative and the additive term $R = D + C + D \times C$ as proposed by Chen and Vellutino (1997). The relationship between decoding and listening comprehension skills appears to be more complex than claimed by the strong version of the simple view of reading. Further research is needed to investigate the extent that both skills are related to each other and to reading comprehension and whether different models should be specified for readers with exceptionally poorly developed versus moderately to highly developed decoding and listening comprehension skills.

In sum, the findings from Study 2 suggest that the simple view of reading in its most simplistic and strongest version could not withstand a methodologically stringent test and that the model is not readily applicable to primary school children at Grades 3 and 4 who learn to read in German. Though we cannot entirely rule out the possibility that the findings would have been more compatible with the simple view if a time-restricted reading comprehension measure was used, it seems reasonable to assume that the model $R = D \times C$ is underspecified and should be enriched by further components. Nevertheless, our findings bear some practical implications with respect to reading development and intervention. The result that a substantial amount of variance in reading comprehension could be accounted for by individual differences in decoding and listening comprehension skills suggest that poor readers might benefit not only from interventions aiming to improve visual word recognition skills but also from trainings focusing on the elaboration of listening comprehension skills. Moreover, fostering both skills early during reading acquisition might help to pave the way for more successful reading development.
Summary and Discussion of Study 3

The third empirical study presented in Chapter IV investigated how German developing readers and adults process local coherence relations. More specifically, the aims of the study were to 1) explore how positive-causal and negative-causal coherence relations are processed in German children and adults, 2) to directly compare the processing of both coherence relations in spoken and written language processing, and 3) to investigate how the processing of positive-causal and negative-causal coherence relations develops throughout primary school years. Cross-sectional data from German primary school children and adults revealed that the processing of negative-causal coherence relations was cognitively more demanding than the processing of positive-causal coherence relations. Overall, children and adults provided less accurate responses when judging the coherence of spoken and written negative-causal sentence pairs compared to positive-causal sentence pairs. This finding supports the cumulative cognitive complexity approach (Evers-Vermeul & Sanders, 2009; Spooren & Sanders, 2008), which states that negative-causal coherence relations are cognitively more complex than positive-causal coherence relations, because they add a negation or contrast to the causal link between two propositions. Consequently, the cumulative cognitive complexity approach predicts that the establishment of negative-causal coherence relations requires more cognitive effort during text comprehension than the establishment of positive-causal coherence relations. Moreover, the results from Study 3 are consistent with the findings from extant studies that can be interpreted in favor of the cumulative cognitive complexity approach. Köhne and Demberg (2013, Exp. 1) and Millis and Just (1994, Exp. 4), for example, found that readers provided less accurate responses to comprehension questions after reading negative-causal compared to positive-causal sentence pairs, and Goldman and Murray (1992) demonstrated that readers had more difficulties appropriately filling in negative connectives into blank slots between two statements than filling in positive-additive or positive-causal connectives.

Furthermore, the accuracy data from Study 3 indicate that developmental trends in the processing of negative-causal coherence relations substantially lagged behind developmental trends in the processing of positive-causal coherence relations. Although there was an overall increase in response accuracy with increasing grade level for negative-causal and positive-causal sentence pairs, children at all grade levels responded with strikingly low accuracy to spoken and written negative-causal sentence pairs, whereas they performed well on positive-causal sentence pairs as early as Grade 1. Fourth graders even approached maximum scores for spoken positive-causal sentence pairs, whereas they provided only 63-67% correct
responses for negative-causal sentence pairs. Again, these findings are consistent with the cumulative cognitive complexity approach, which states that cognitively more complex negative-causal coherence relations are acquired later than cognitively less complex positive-causal coherence relations (Evers-Vermeul & Sanders, 2009; Spooren & Sanders, 2008). Although our results are not conclusive with respect to the acquisition order of positive-causal and negative-causal coherence relations, they suggest that the ability to establish both types of coherence relations still develops throughout primary school years (as shown by Cain & Nash, 2011; Cain et al., 2005). Moreover, the ability to establish negative-causal coherence relations seems to be fairly poorly developed at Grades 1 and 2 and its development seems to be still ongoing at the end of Grade 4. In contrast, the ability to establish positive-causal coherence relations is already highly developed at school entry and further improves until the end of Grade 4.

This developmental pattern is consistent with the results from the first study of Dragon et al. (2015, Study 1) in which German 2nd and 3rd graders exhibited severe difficulties judging the sensibility of negative-causal sentence pairs, whereas they had no difficulties judging positive-causal (and temporal) sentence pairs. However, results of their second study (Dragon et al., 2015, Study 2) suggest that children might have ignored both negative-causal and positive-causal connectives. Instead, they seemed to judge the sensibility of the presented sentence pairs primarily on the basis of the semantic compatibility of the sentence contents. Positive-causal connectives link two statements that are compatible with respect to world knowledge. Thus, this strategy is successful for positive-causal sentence pairs resulting in correct responses. Negative-causal connectives however link two incompatible statements, and when the connective is ignored, it results in systematically incorrect responses. The finding that first graders responded with high accuracy to sentence pairs containing positive-causal connectives but provided systematically incorrect responses (i.e., responses with below-chance level accuracy) for sentence pairs containing negative-causal connectives resembles the findings by Dragon et al. and suggests that at least the younger children in our study might have used a semantic world knowledge strategy. However, one finding contradicts this explanation. First graders in our study were able to correctly reject incoherent filler sentence pairs that included compatible statements linked by an inappropriate positive-causal connective instead of an appropriate positive-additive connective (e.g., *Der Vater schimpft viel. Darum hat der Vater einen Bart/*The father grumbles a lot. Therefore, the father has a beard). If they had simply ignored the connectives and instead had used a semantic strategy, they would have erroneously accepted all of these sentence pairs as
coherent, because the statements in these sentence pairs are not incompatible without the connective. However, the children performed fairly well on rejecting the incoherent filler sentence pairs that contained a positive-causal connective, whereas they performed poorly on filler sentence pairs that contained a negative-causal connective. Based on these findings, it seems reasonable to assume that the first graders in Study 3 did not use a semantic strategy as proposed by Dragon et al. Instead, the obtained differences in response accuracy between positive-causal and negative-causal sentence pairs seem to be due to differences in the processing difficulty of positive-causal and negative-causal connectives and coherence relations.

Finally, patterns of response accuracy were highly comparable in spoken and written language processing. This finding is consistent with predictions derived from the simple view of reading, which implies that higher-order processes of spoken and written language comprehension (such as establishing mental relations of local coherence) are basically the same after written word forms have been decoded. The finding that the processing of positive-causal and negative-causal coherence relations was highly comparable in both modalities is perfectly in line with this assumption. In contrast, the finding seems to be inconsistent with Perfetti’s (1985) verbal efficiency hypothesis. The verbal efficiency hypothesis claims that little cognitive resources remain for higher-level comprehension processes (such as establishing coherence relations) in beginning readers, because they still need sufficient cognitive resources to recognize written word forms. Consequently, beginning readers should have exhibited more difficulties with the processing of coherence relations (and of cognitively complex negative-causal coherence relations in particular) in written compared to spoken language comprehension. Instead, we obtained highly comparable patterns of response accuracy in spoken and written language processing. However, I would like to emphasize that a direct comparison of spoken and written language processing was possible only for children in Grades 3 and 4, because the visual version of the semantic verification task with sentence pairs was administered to 3rd and 4th graders only. Given the high orthographic consistency in German, a fairly high level of visual word recognition skill acquisition seems possible by then (e.g., Aro & Wimmer, 2003; Wimmer & Goswami, 1994; see also Summary and Discussion of Study 1). Consequently, these 3rd and 4th graders might have had sufficient cognitive resources available for the processing of even more complex negative-causal coherence relations. Thus, early differences in spoken and written language processing as predicted by the verbal efficiency hypothesis might have diminished by the end of Grade 3. Another explanation might be that the visual semantic verification task with sentence pairs
allowed readers to slow down or pause reading, look back, or reread the sentence pairs if necessary to support poor visual word recognition skills (Walczyk, 1995; Walczyk et al., 2004). These compensatory strategies might have diminished differences in the processing of coherence relations between spoken and written language processing. Accordingly, the finding that processing of coherence relations was highly comparable in both modalities does not necessarily question the predictions derived from Perfetti’s verbal efficiency hypothesis but should be interpreted with the limitations of the visual verification task in mind.

In contrast to the response accuracy data, response latencies cannot be readily interpreted in favor of the cumulative cognitive complexity approach. Adults provided slower responses to negative-causal compared to positive-causal sentence pairs only when sentence pairs were presented visually but not when they were presented auditorily. For children, however, longer response latencies to negative-causal compared to positive-causal sentence pairs were obtained only when sentence pairs were presented auditorily and only for incoherent sentence pairs. The expected difference for coherent spoken sentence pairs was found only at higher grade levels. Moreover, response latency patterns differed remarkably for spoken and written language processing and were inconsistent with predictions derived from the simple view of reading and from the verbal efficiency hypothesis. The difficulty in interpreting these results in comparison to a clearer interpretation of the response accuracy data in support of the cumulative cognitive complexity approach presents a quandary. A likely explanation is that processing latencies are more sensitive to factors, cognitive processes, or processing strategies that have an impact on processing fluency but not necessarily on the outcome (as indicated by response accuracy) of coherence establishing processes in the verification task. Differences in word frequency, for example, might have affected processing latencies unsystematically. Although both tasks included only words that were likely to be known by German children of the relevant age groups, we did not control for the individual degree of familiarity with specific words included in our stimulus material. Moreover, elaborative cognitive processes such as drawing inferences, the activation of prior knowledge, or processing strategies such as rereading single words, sentences, or whole sentence pairs (when presented visually) might have altered response latencies. As a result, possible differences could have been concealed in the processing of positive-causal and negative-causal coherence relations.

Response latencies are also more susceptible to distractions. The study was conducted in classrooms with several students at the same time. Some children were prone to distraction by their classmates, sometimes glancing outside the window or asking questions during the
presentation of a critical item. Although we excluded unusually short or long response latencies from the analyses, the exclusion of all response latencies that came about irregularly by our outlier detection procedure is not absolutely certain.

A more promising way to detect processing time difference between positive-causal and negative-causal coherence relations and connectives would be to use a more stringent method that assesses online reading latencies by means of self-paced reading or eye-tracking instead of recording response latencies after processing the whole sentence pair. In a self-paced reading experiment with adult readers, Knoepke et al. (2015) found no overall reading time differences between positive-causal and negative-causal sentence pairs. However, they obtained longer reading times when inconsistent information (e.g., easel) was detected in negative-causal sentence pairs (e.g., Gregor wants to become a musician like his father. Nevertheless, his parents buy him an easel for Christmas) as compared to consistent information (e.g., easel) in positive-causal sentence pairs (e.g., Gregor wants to become a painter like his father. Therefore, his parents buy him an easel for Christmas). Moreover, they found longer reading times at the sentence-final region (e.g., for Christmas) in negative-causal as compared to the positive-causal sentence pairs indicating increased difficulties when inconsistent information was integrated with the previous text to establish a coherent mental model. These findings suggest that online measures of reading times might reveal subtle processing differences between negative-causal and positive-causal coherence relations that might have been concealed in our study.

The findings by Knoepke et al. (2015) might also explain the core processes of the cognitive complexity of negative-causal coherence relations, a question that was neglected in the current study. Longer reading times on inconsistent information in negative-causal sentence pairs suggest that inconsistencies are validated against pertinent knowledge and recognized immediately during online-text comprehension (Isberner & Richter, 2013). Given that the validation outcome is negative, cognitive processing effort increases. Moreover, longer reading times at the end of the critical sentence pair indicate increased difficulties in integrating inconsistent information into a coherent mental model of the text. Knoepke et al. concluded that the validation and integration of information inconsistent with world knowledge is at the core of the cognitive complexity of negative-causal compared to positive-causal coherence relations. This two-step model of comprehending negative-causal coherence relations might also explain the surprising finding in Study 3 that longer response latencies are associated with less accurate responses for coherent negative-causal sentence pairs in children (both in spoken and written language processing) and for coherent and incoherent
negative-causal sentence pairs in adults (written language processing). Long response latencies conceivably indicate the detection of inconsistencies in negative-causal sentence pairs through immediate validation. However, the attempt to repair this inconsistency and to integrate the inconsistent information into the existing mental model of the text appears to fail fairly often resulting in lower response accuracy.

In sum, the response accuracy data in our study indicate that the processing of negative-causal coherence relations and connectives requires more cognitive effort than the processing of positive-causal coherence relations and connectives in spoken and written language processing for both children and adults. Moreover, we found developmental trends in the processing of negative-causal coherence relations to lag behind developmental trends in the processing of positive-causal coherence relations. These findings support the cumulative cognitive complexity approach. The ability to establish coherence relations (i.e., to mentally link the propositions of a text and to integrate them into a coherent mental model of the text) is a necessary requirement for successful reading comprehension (e.g., Cain et al., 2005, Study 2; Geva & Ryan, 1985; Hannon & Danemann, 2004; Long & Chong, 2001; see Chapter 1, Establishing Coherence for a detailed discussion). Difficulties establishing coherence relations can result in a lack of understanding spoken and written tasks or instructions in school and in miscommunication between students and teachers, which in turn affects student outcomes such as bad grades (see discussion in Dragon et al., 2015). Although our results are limited to the processing of positive-causal and negative-causal coherence relations, the findings that the comprehension of both coherence relations still develops throughout primary school years and that even 3rd and 4th graders exhibit major difficulties comprehending negative-causal coherence relations suggest that it might be beneficial to integrate practical exercises or strategy trainings concerning the establishment of coherence relations (and of cognitively complex negative-causal coherence relations in particular) into primary school curricula as early as possible. Moreover, learning materials and textbooks should be constructed with age-restricted abilities to establish specific coherence relations in mind to permit full comprehension of text contents.

Summary and Discussion of Study 4

The purpose of the fourth empirical study presented in Chapter V was to demonstrate the construct validity of process-oriented test instruments that selectively assess individual differences in cognitive component skills of reading comprehension. To this end, we used the ProDi-L grammaticality judgment task, which assesses individual differences in skills of
syntactic integration in German 3rd and 4th graders, as an example. The test is characterized by a well-defined task that includes items with carefully varied item features (syntactic complexity, grammaticality, and type of grammatical violation), which according to psycholinguistic research facilitate or impede processes of syntactic integration, resulting in items of varying difficulty. Moreover, the task includes no reading-unrelated cognitive skills such as picture recognition, language production, or comparing and memorizing response alternatives in multiple-choice questions. Finally, the task is designed to measure response accuracy as an indicator of syntactic integration processes’ reliability and response latency as an indicator of automatization degree. In a cross-sectional study with German 3rd and 4th graders, we used explanatory item response models to test the construct validity of the ProDi-L grammaticality judgment task. If the task assesses skills of syntactic integration, then experimentally varied item features, which are known to facilitate or impede syntactic integration, should predict empirically observed item difficulties (Wilson & De Boeck, 2004; see also Neeb et al., 2015, who used the same procedure to demonstrate the construct validity of a phonological comparison task). Analyses of the corresponding ProDi-H auditory grammaticality judgment task were also conducted (for Grades 1 to 4) to be able to directly compare syntactic integration skills in listening and reading comprehension.

Overall, the findings in Study 4 provide empirical evidence in favor of the construct validity of the ProDi-L and ProDi-H grammaticality judgment tasks. The test items of both tests detected individual differences in syntactic integration across the whole range of person abilities and differentiated person abilities best when syntactic integration skills were moderately or highly developed. As expected, empirically observed item difficulties varied as a function of experimentally varied item features. Response latencies increased with increasing syntactic complexity, whereas response accuracy decreased with increasing syntactic complexity. The only exception was that the latter negative relational pattern was not found in the response accuracy data of the auditory ProDi-H grammaticality judgment task. One explanation could be that even the most complex sentences in our study were not complex enough to affect response accuracy in spoken language processing, which requires less cognitive effort than written language processing. Apparently, sufficient cognitive resources are available in spoken language processing to deal with syntactically complex sentences resulting in overall high response accuracy for grammatical sentences (more than 90% correct responses). However, increased syntactic complexity was reflected in longer response latencies in the auditory grammaticality judgment task. This finding is consistent
with previous studies showing increased syntactic complexity to be associated with longer processing times (Ferreira et al., 1996; Graesser et al., 1980).

Children in the current study responded more slowly and provided less accurate responses to ungrammatical compared to grammatical test sentences. This finding is in line with extant studies demonstrating that syntactically inappropriate words in ungrammatical sentences were processed more slowly than words in grammatical sentences (Flores d’Arcais, 1982, Exp. 4; see also Baum, 1989, 1991) and that participants judged the grammaticality of ungrammatical sentences less accurately compared to grammatical sentences (Friederici, et al., 2002). However, children unexpectedly responded to sentences containing word order violations just as fast as to grammatical sentences in the auditory grammaticality judgment task. One explanation might be that word order information is such a highly reliable cue in young German children’s spoken language processing (Dittmar et al., 2008; Schipke et al., 2012) that they recognize word order violations easily and with high accuracy (>85% correct responses) in spoken sentences. A reasonable assumption is that word order violations occur more often in spoken than in written language comprehension. Consequently, the processing of spoken word order violations might have been comparatively easy, thus eliciting short response latencies.

Children provided the slowest and least accurate responses to sentences containing case-marking violations, whereas they responded fastest and with highest accuracy to sentences containing tense-form violations with sentences containing word-order violations in between. These findings are consistent with studies by Dittmar et al. (2008) and Schipke et al. (2012) who found that German primary school children are less sensitive to case-marking information compared to word-order information during sentence processing. Consequently, they exhibit more difficulties detecting violations of case marking compared to violations of word order. Moreover, as expected, children provided the fastest and most accurate responses to violations of tense form. This type of grammatical violation was presumably easiest to detect, because it contained two violations—a word-order violation and a morphological violation consisting of an incorrectly conjugated verb form in the sentence-final position. The only unexpected finding was that children needed as much time to respond to sentences containing tense-form violations than to sentences containing case-marking violations in the visual version of the task. Given that violations of verb tense form included a sentence-initial auxiliary and a mismatching sentence-final finite verb form, a reasonable assumption is that children regressed during reading from the finite verb form to the mismatching auxiliary to
recheck their syntactic analysis. This regression is possible only in the visual grammaticality judgment task, and it might have increased response latencies for this type of violation.

Finally, response accuracy increased and response latency decreased with increasing grade level, indicating that both tests were able to detect developmental changes in syntactic integration skills across primary school years. All results were highly comparable for the ProDi-L and the ProDi-H grammaticality judgment tasks. As discussed earlier (see Summary and Discussion of Study 3), this finding is consistent with the prediction derived from the simple view of reading that higher-order processes of spoken and written language comprehension (such as syntactic integration and establishing local coherence relations) are basically the same.

In sum, our findings provide evidence that demonstrates the construct validity of the ProDi-L and the ProDi-H grammaticality judgment tasks. Both tasks are adequate tools for assessing individual differences in skills of syntactic integration in spoken (Grades 1 to 4) and written language processing (Grades 3 to 4) in German primary school children and for detecting children with exceptionally poorly-developed syntactic integration skills. Given that syntactic integration is one of the cognitive component skills of reading comprehension (see Chapter I, Skills of Syntactic integration), test instruments assessing individual differences in skills of syntactic integration, such as the grammaticality judgment tasks of ProDi-L and ProDi-H, are of high practical relevance for the diagnosis of individual reading deficits. Several studies (e.g., Byrne, 1981; Casalis & Louis-Alexandre, 2000; Nation & Snowling, 2000; Plaza & Cohen, 2003; Richter et al., 2012; Stothard & Hulme, 1992; Tunmer et al., 1987) have demonstrated that individual differences in skills of syntactic integration are positively related to individual differences in reading comprehension, suggesting that poorly developed or deficient syntactic integration skills can cause poor reading comprehension (for a critical discussion of the relationship between reading comprehension and syntactic integration skills, see Oakhill & Cain, 1997). In an earlier study, Richter et al. demonstrated that both response accuracy and response latency measures in the ProDi-L grammaticality judgment task were predictive of 3rd and 4th grader’s reading comprehension. Response accuracy as an indicator of syntactic integration processes’ reliability was positively related to reading comprehension, whereas response latency as an indicator of automaticity degree was negatively related to reading comprehension. The findings by Richter et al. and the findings from Study 4 demonstrate that both response accuracy and response latency measures of the ProDi-L and the ProDi-H grammaticality judgment tasks are indicative of the effectiveness of syntactic integration processes in primary school children, which in turn, might affect overall
General Discussion

reading comprehension (see Chapter VI). If poorly developed or deficient syntactic integration skills contribute to poor reading comprehension, the grammaticality judgment tasks of ProDi-L and ProDi-H should be able to detect this deficit. Thus, both tasks provide a basis for the construction of adaptive and target-oriented reading interventions that are specially geared to foster syntactic integration skills, resulting in improved reading comprehension. A comprehensive diagnosis of reading difficulties, however, requires more than a single test that assesses individual differences in syntactic integration skills (see Summary and Discussion of Chapter VI for a more detailed discussion).

Summary and Discussion of Chapter VI

The aim of Chapter VI was to argue for the usefulness of characterizing poor readers by their deficits in the cognitive component skills of reading comprehension instead of classifying them as either dyslexics or as “garden-variety poor readers” (Stanovich 1988, p. 590). Characterizing poor readers as dyslexics based on an arbitrary aptitude-achievement discrepancy criterion (i.e., below age-average reading comprehension in the absence of any other cognitive deficit and adverse environmental factors, American Psychiatric Association 2005; World Health Organization 2010) creates at least two major problems. The first argument against this wide-spread discrepancy model is that poor readers classified as dyslexics differ very little from garden variety poor readers with more general cognitive deficits with respect to their need for reading intervention. Second, and most importantly, classifying poor readers as dyslexics lacks diagnostic value with respect to the underlying individual reading deficit and the required reading intervention. The literature review in Chapter VI strongly suggests that poor reading comprehension cannot be attributed to a single underlying deficit as assumed, for example, by the phonological-core variable-difference model (Stanovich, 1988). Instead, sources of reading difficulties seem to be multifaceted and heterogeneous. To the same extent that each cognitive component skill of reading comprehension at the word level (i.e., phonological recoding, orthographical decoding, and access to word meanings), at the sentence level (i.e., syntactic and semantic integration), and at the text level (i.e., establishing coherence, drawing inferences, and comprehension monitoring) contributes to reading comprehension, reading comprehension is impeded when one or more of these skills are deficient (e.g., Cain & Oakhill, 2004; Vellutino et al., 2004). From a cognitive point of view, characterizing poor readers according to their deficits in these cognitive components skills is more promising than categorizing them based on whether or not they have dyslexia. Consequently, a meaningful diagnosis of reading
difficulties requires a comprehensive test battery such as ProDi-L that assesses individual differences in the cognitive component skills of reading comprehension separately to detect individual patterns of reading deficits. Such a comprehensive process-oriented diagnosis of reading difficulties raises the possibility of developing adaptive and target-oriented reading interventions for improving specific deficient cognitive component skills and in turn overall reading comprehension.

Limitations of the Reported Studies and Directions for Future Research

The four empirical studies reported in this dissertation revealed new findings on the contributions of visual word recognition skills and listening comprehension skills to reading comprehension in German developing readers (Studies 1 and 2), on the processing of positive-causal and negative-causal coherence relations in developing German readers and adults (Study 3), and on the construct validity of the ProDi-L (and ProDi-H) grammaticality judgment task that assesses individual differences in syntactic integration skills in German primary school children in a process-oriented fashion (Study 4). However, these findings need to be interpreted with an understanding of the limitations in each study.

First, all analyses reported in each study were based on cross-sectional data of German 1st to 4th graders. Having used a cross-sectional design, the possibility that some of the grade level differences obtained in Studies 1 to 3 were due to uncontrolled group differences rather than actual developmental differences between grade levels cannot be ruled out. However, given that all analyses were based on comparatively large student samples (including students from several classes and schools in Frankfurt am Main, Cologne, and Kassel) and that the majority of our findings were consistent with predictions derived from the theories and approaches under investigation, the obtained developmental changes are highly unlikely the mere result of uncontrolled group differences. Nevertheless, replicating the results of Studies 1 to 3 with longitudinal data would add additional weight to our conclusions drawn from the developmental changes in cognitive component skills of reading comprehension throughout primary school years. In a current comprehensive longitudinal study on the development of cognitive component skills of reading and listening comprehension in primary school children, all subtests of the ProDi-L and ProDi-H test batteries were administered to two cohorts of German primary school children over a period of almost five years from school entry (beginning of Grade 1) until the end of Grade 4 (funded by the Federal Ministry of Education and Research, Bundesministerium für Bildung und Forschung, BMBF, grants 01 GJ 0985, 01 GJ 0986, 01 GJ 1206A, 01 GJ 1206B). The recently collected data are ideally
suited to investigating the research questions from Studies 1 to 3 on the basis of a longitudinal
design and to corroborate the reported findings in future analyses.

Second, the four empirical studies focused on the development of cognitive component skills in 1st to 4th graders and not in higher grade levels. Study 3 revealed that children’s comprehension of negative-causal coherence relations is still developing by the end of Grade 4, suggesting that some of the cognitive component skills of reading comprehension might not be fully developed by the end of primary school. Thus, future research on reading development should include secondary school students to assess developmental changes in cognitive component skills of reading comprehension beyond primary school age.

Third, individual differences in text comprehension skills and individual differences in the processing of positive-causal and negative-causal coherence relations were assessed via the ProDi-L and the ProDi-H semantic verification tasks with sentence pairs in Studies 2 and 3. A possible objection concerns the external validity of these tasks. Short texts consisting of only two sentences each might not be naturalistic and long enough to sufficiently assess complex text comprehension processes (such as establishing global coherence relations, making use of text-genre knowledge, or making use of reading strategies). However, comprehending sentence pairs requires not only the comprehension of both propositions separately but also more complex processes such as the establishment of local coherence relations, the activation and integration of world knowledge, and monitoring the comprehension process. Thus, I assume that the ProDi-L and the ProDi-H semantic verification tasks with sentence pairs are sufficient in assessing individual differences in text comprehension skills and in the processing of positive-causal and negative-causal coherence relations. Nevertheless, further research is warranted to investigate whether the obtained findings from Studies 2 and 3 can be replicated with longer and more naturalistic texts.

Finally, the generalizability of the reported findings to other languages and scripts was not addressed in the four studies. Most theories and models of the cognitive component skills of reading comprehension, reading acquisition, and reading development have been established on the basis of European languages and orthographies. As discussed in Chapter I, dual route models of visual word recognition, for example, appear to be highly suitable for alphabetic scripts, whereas they are probably not suitable to explain reading phenomena and reading development in non-alphabetic scripts such as Chinese and Japanese (e.g., Frost, 2012). Finding substantial differences between alphabetic and non-alphabetic scripts in visual word recognition processes would not be surprising, because this is where script differences matter most. However, higher-order cognitive processes of reading comprehension at the
sentence and text level and their development arguably would not differ substantially between the two types of language scripts. According to the simple view of reading, higher-order cognitive processes of listening and reading comprehension are basically the same after written word forms have been decoded (Gough & Tunmer, 1986; Hoover & Gough, 1990). Assuming that basal cognitive processes of listening comprehension are the same for all languages, sentence and text level processes of reading comprehension should not differ considerably between languages with different scripts. Nevertheless, future research needs to investigate whether cognitive component skills of reading comprehension, their interrelatedness, and development are comparable for European and non-European languages and scripts.

Conclusion

The aim of this dissertation was to advance a more comprehensive understanding of the cognitive component skills of reading comprehension in developing readers. The findings of Study 1 add to extant theoretical literature and empirical research on the relationship between visual word recognition skills and reading comprehension by showing that both phonological recoding and orthographical decoding skills make significant and unique contributions to reading comprehension in German 2nd to 4th graders. Most remarkably, phonological recoding skills explained a substantial amount of variance in reading comprehension even in more advanced readers in Grades 3 and 4. Theoretical implications concerning well-established models of visual word recognition and reading development such as Coltheart et al.’s (2001) DRC model and Frith’s (1985, 1986) three-stage developmental model have been discussed. Study 2 amplifies empirical research on one of the most prominent models of reading comprehension—the simple view of reading (Gough & Tunmer, 1986; Hoover & Gough, 1990). The results showed that the simple view in its most simplistic form could not withstand a methodologically stringent test that used optimized measures of the model components. Moreover, the simple view was not readily applicable to German 3rd and 4th graders. The findings suggest that the model is probably underspecified and might be enriched by additional components. Study 3 adds a valuable increment to research on children’s and adults’ processing of local coherence relations in written and spoken language comprehension. Supporting the cumulative cognitive complexity approach (Evers-Vermeul & Sanders, 2009; Spooren & Sanders, 2008), the analyses showed that comprehending negative-causal coherence relations requires more cognitive effort than comprehending positive-causal
coherence relations in German primary school children and adults. Moreover, comprehension
of both coherence relations seems to develop throughout primary school years with
developmental trends in the processing of negative-causal coherence relations lagging behind
developmental trends in the processing of positive-causal coherence relations.

In addition to the theoretical implications of the reported findings, Studies 1 to 3 bear
several practical implications with respect to reading instruction, the diagnosis of individual
reading deficits, and reading intervention. Practical exercises or systematic strategy trainings
that selectively target phonological recoding skills, orthographical decoding skills, listening
comprehension skills, and skills of establishing local coherence relations can be expected to
enhance reading comprehension not only in beginning but also in more advanced readers and
should thus be an integral part of reading instruction throughout primary school years.
Moreover, the results of Studies 1 and 2 suggest a relationship between poorly developed
visual word recognition and listening comprehension skills and reading difficulties. The
literature review in Chapter VI provides ample evidence that deficits in cognitive component
skills at the word, sentence, and text level can impede reading comprehension to the same
extent that these component skills contribute to reading comprehension (e.g., Cain & Oakhill,
2004). We emphasized the necessity to characterize poor readers according to their individual
deficits in these cognitive components skills to allow for the construction of adaptive and
target-oriented reading interventions that augment specific deficient cognitive component
skills, which in turn should improve overall reading comprehension. Such a meaningful
diagnosis of reading difficulties requires a comprehensive test battery, such as ProDi-L, that
assesses individual differences in cognitive component skills of reading comprehension
selectively in a process-oriented manner. Study 4 provided empirical evidence in favor of the
construct validity of one of the ProDi-L subtests (and the corresponding ProDi-H subtest)—
the grammaticality judgment task that assesses skills of syntactic integration. Explanatory
item response models demonstrated that the ProDi-L (and ProDi-H) grammaticality judgment
task is an adequate tool to assess individual differences in skills of syntactic integration in
German primary school children and to detect children with exceptionally poorly developed
syntactic integration skills. In this vein, the subtest provides an ideal basis for the construction
of reading intervention and remediation programs that specifically augment syntactic
integration skills to improve overall reading comprehension.

Although the results from the four empirical studies reported in this dissertation
provided answers to several open research questions concerning the cognitive component
skills of reading comprehension and their development throughout primary school years,
future research is still needed to corroborate and extend the findings to achieve a full and comprehensive understanding of reading comprehension, reading acquisition, and reading development. Such a comprehensive understanding is indispensable for an optimal composition of reading lessons, for the construction of learning materials, reading instructions, and textbooks. Most importantly, a comprehensive understanding of the cognitive component skills involved in reading comprehension is the basis for a meaningful diagnosis of individual patterns of reading deficits and for the construction of adaptive and target-oriented reading intervention and remediation programs to improve reading comprehension in poor readers. The findings from the four empirical studies in this dissertation constitute a step on the way to such a comprehensive understanding of the cognitive component skills of reading comprehension in developing readers.
References
References


References


Bäuerlein, K., Lenhard, W., & Schneider, W. (2012a). *LESEN 6-7 - Lesetestbatterie für die Klassenstufen 6-7* [LESEN 6-7 – Reading-test battery for Grades 6-7]. Göttingen: Hogrefe.


References


References


References


References


• References •


Universität Kassel, Fachbereich Humanwissenschaften
Erklärung zur kumulativen Dissertationen im Promotionsfach Psychologie

Erklärung über den Eigenanteil an den veröffentlichten oder zur Veröffentlichung vorgesehenen wissenschaftlichen Schriften innerhalb meiner Dissertationsschrift, Ergänzung zu § 5a Abs. 4 Satz 1 der Allgemeinen Bestimmungen für Promotionen an der Universität Kassel vom 13. Juni 2011:

I. Allgemeine Angaben:

Name: Knoepke, Julia
Institut: Institut für Psychologie, Fachbereich 01 Humanwissenschaften, Universität Kassel
Thema: Cognitive Component Skills of Reading Comprehension in Developing Readers

II. Nummerierte Aufstellung der eingereichten Schriften:


III. Darlegung des eigenen Anteils an diesen Schriften:

Zu Nr. 1, 2, 3 und 4: Die empirischen Untersuchungen basieren auf Daten, die im Rahmen des durch das Bundesministerium für Bildung und Forschung geförderten Verbundprojekts „Prozessbezogene Diagnostik des Lese- und Hörverstehens im Grundschulalter“ (Förderkennzeichen: 01 GJ 0985, 01 GJ 0986, 01 GJ 1206A, 01 GJ 1206B) unter der Leitung von Prof. Dr. Tobias Richter und Prof. Dr. Johannes Naumann erhoben wurden. Die hier verwendeten Lese- und Hörverstehens tests für
Grundschulkinder wurden durch Prof. Dr. Tobias Richter, Prof. Dr. Johannes Naumann, Dr. Maj-Britt Isberner und Yvonne Neeb konzeptualisiert und programmiert. Die Datenerhebung wurde durch Dr. Maj-Britt Isberner und Yvonne Neeb (Nr. 1) und durch Dr. Maj-Britt Isberner, Yvonne Neeb und Julia Knoepke (Nr. 2, 3 und 4) zu ungefähr gleichen Teilen durchgeführt (Eigenanteil: ca. 33%). Die Entwicklung der Konzeption der Manuskripte, die Literaturrecherche, die Auswertung der Daten, die Diskussion der Ergebnisse, die Beweisführung sowie das Niederschreiben der Manuskripte wurde überwiegend durch die Erstautorin Julia Knoepke (Eigenanteil: ca. 80%) und Prof. Dr. Tobias Richter (ca. 15%) durchgeführt. Die Koautor(inn)en Prof. Dr. Johannes Naumann, Dr. Maj-Britt Isberner, Yvonne Neeb (Nr. 1, 2, 3 und 4) und Prof. Dr. Sabine Weinert (Nr. 4) haben die Manuskripte vor der Einreichung bei der jeweiligen Fachzeitschrift überprüft und kommentiert (ca. 5%).

Zu Nr. 5: Die Entwicklung der Konzeption des Handbuchkapitels, die Literaturrecherche, die Ausarbeitung der Argumentation sowie das Niederschreiben des Handbuchkapitels wurde überwiegend durch die Erstautorin Julia Knoepke, M.A. (Eigenanteil: ca. 80%) und Prof. Dr. Tobias Richter (ca. 20%) durchgeführt.

IV. Anschriften der jeweiligen Mitautor(inn)en:

Prof. Dr. Tobias Richter: tobias.richter@uni-kassel.de
Prof. Dr. Johannes Naumann: j.naumann@em.uni-frankfurt.de
Dr. Maj-Britt Isberner: maj-britt.isberner@uni-kassel.de
Yvonne Neeb: neeb@dipf.de
Prof. Sr. Sabine Weinert: sabine.weinert@uni-bamberg.de

Datum, Unterschrift der Antragstellerin

........................................... ...........................................
Ort, Datum Unterschrift

Ich bestätige die von Frau Julia Knoepke unter Pkt. III abgegebene Erklärung:

Prof. Dr. Tobias Richter: ........................................... ...........................................
Ort, Datum Unterschrift

Prof. Dr. Johannes Naumann: ...........................................
Ort, Datum Unterschrift

Dr. Maj-Britt Isberner: ...........................................
Ort, Datum Unterschrift
<table>
<thead>
<tr>
<th>Name</th>
<th>Ort, Datum</th>
<th>Unterschrift</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yvonne Neeb:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prof. Dr. Sabine Weinert:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

........................................          ........................................
Ort, Datum                                                                                            Unterschrift
Acknowledgements

First of all, I would like to thank my advisor, Professor Dr. Tobias Richter. I feel very lucky that he offered me a position in his research group and took on the task of making me a ‘real psychologist’. In the past four years, he has been an infinite source of knowledge, ideas, and inspiration to me. He often pushed me to my limits while at the same time encouraging me to grow as a researcher. I am thankful for his relentless support and encouragement, for his critical questions, his constructive feedback, and for countless stimulating and inspiring discussions. I also thank Prof. Dr. Cordula Artelt and Prof. Dr. Mirjam-Ebersbach for their time and interest in this dissertation.

I would also like to offer my sincere thanks to my other co-authors—Dr. Maj-Britt Isberner, Prof. Dr. Johannes Naumann, Yvonne Neeb, and Prof. Dr. Sabine Weinert—for the close and productive collaboration, for fruitful discussions, and for their valuable feedback on the articles which are part of this dissertation.

Moreover, I am very grateful to the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) who supported the research presented in this dissertation, and to the coordinators of the Research Initiative Forschungsinitiative Sprachdiagnostik und Sprachförderung for several stimulating workshops and meetings. Many thanks also go to all the student research assistants for programming assistance and for collecting data, to the schools, teachers, and, of course, to all the children who were willing to participate in the reported studies.

I owe special thanks to my dear colleagues from the Department of Psychology at the University of Kassel—in particular to Dr. Maj-Britt Isberner and Dr. Isabel Lindner, who always supported me with their scientific expertise, with warm and encouraging words when I needed them most, and with “princess evenings” full of happiness and Berlin, Berlin.

Finally, I would like to thank my mother, Jana Knoepke, my father, Torsten Knoepke, my brother, and my grandparents, for their unconditional love and relentless support. Special thanks goes to all of my friends, in particular to Lisa Maringer and Judith Elias for their faith in me, to Victoria Bartlitz for stimulating (and quite entertaining) scientific discussions, and to Dr. Dominik Weber, who supported me when I moved to Kassel and who endured some of the sorrowful moments I experienced during my time as a PhD student. Last but not least, I want to thank my partner, Dr. Simon Schindler, who was there for me during the final phase of my dissertation and supported me with love, meditation, and—most importantly—calm.