

Lexical quality and reading comprehension in primary school children

accepted for publication in the journal *Scientific Studies of Reading*

Tobias Richter and Maj-Britt Isberner

Johannes Naumann and Yvonne Kutzner

University of Kassel

German Institute for International Educational
Research

Author note

The research reported in this article was supported by a grant from the German Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF, grant 01GJ0985). We would like to thank David Nitz for programming assistance and Lilly Andrich, Claudia Beck, Tatjana DeToia, Wanja von der Felsen, Christian Kirsch, Katrin Klein, Dominik Mathieu, Sarah Rom, Miriam Schwarzenthal, Vanessa Wergin, Barbara Werner, and Monika Wieland for their help in preparing stimulus materials and collecting data. Researchers who would like to inspect the items of the ProDi-L assessments used in this study are invited to send an e-mail to the first author.

Corresponding author:

Tobias Richter

University of Kassel, Department of Psychology

Holländische Str. 36-38

34127 Kassel, Germany

E-mail: tobias.richter@uni-kassel.de

Telephone: +49-561 804 3577

Fax: +49-561 804 3586

Abstract

In a cross-sectional study, we examined the relationship between the quality of lexical representations and text comprehension skill in German primary school children (grades 1-4). We measured the efficiency and accuracy of orthographical, phonological, and meaning representations by means of computerized tests. Text comprehension skill was assessed with a standardized reading test with questions requiring recognition of text information and inferencing. Both the accuracy of and the efficiency of access to the three types of lexical representations contributed to explaining interindividual variation in text comprehension skill. Results from a path-analytic model suggest a specific causal order of the three components of lexical quality with the quality of meaning representations partly mediating the effects of form representations.

Keywords: efficiency of lexical access, lexical quality, orthographical representations, meaning representations, phonological representations, reading comprehension

Lexical quality and reading comprehension in primary school children

According to the lexical quality hypothesis, reading comprehension skill strongly depends on the quality of lexical representations of words (Perfetti, 2007; Perfetti & Hart, 2001). In essence, the theory posits that for reading comprehension to function smoothly and successfully, the mental representations of words must be based on accurate lexical representations which can be retrieved rapidly, that is, without much cognitive effort. The three lexical representations, consisting of orthographical, phonological, and meaning components, are assumed to be of high quality when all three components are fully specified and tightly bound together so that the retrieval of one type of information (e.g., a word's spelling) also leads to the activation of the other types of information associated with the same word (e.g., its correct pronunciation and meaning). This latter assumption has attracted considerable research and has received support in studies with adult readers (e.g., experiments investigating form-meaning confusions, Perfetti, 2007). Evidence has shown, however, the components are typically not (yet) closely associated with one another in developing readers, forming loosely related dimensions of lexical skills rather than tightly bound constituents of the representations of words (Perfetti & Hart, 2002). This lack of association raises the important question of how these lexical skills act in concert to achieve good reading comprehension in developing readers. In the present study, we attempted to answer this question by investigating the relative contribution of the three main components of lexical quality (i.e., phonological, orthographical, and meaning representations) and the extent that they account for grade-level differences in the reading comprehension skill of primary school children. We also aimed to establish a path-analytic model of the component links to reading comprehension skill. These two aspects of the study and their theoretical rationale are subsequently discussed.

Lexical Quality and Grade-level Differences in Reading Comprehension Skill

The lexical quality hypothesis posits that high-quality lexical representations of words are a necessary precondition of skilled reading comprehension (Perfetti, 2007; Perfetti & Hart, 2001). Lexical quality is based on the availability of *accurate* lexical representations, which can be accessed *efficiently* during comprehension (verbal efficiency, Perfetti, 1985). Readers with poor lexical representations risk retrieving imprecise or incomplete lexical information during comprehension, resulting in the need to allocate more working memory capacity to word-level processes that is no longer available for higher-level comprehension processes such as knowledge-based inferences (Perfetti, 1985). Both problems can affect reading comprehension on the text level.

The link between lexical quality and reading comprehension skill is particularly important in developing readers. Reading instruction in primary school places a strong emphasis on word-level reading skills (e.g., phonics instruction, vocabulary instruction, and fluency training, National Reading Panel, 2000) and not so much on higher-level reading skills such as reading strategies. This emphasis is consistent with the hypothesis that most (if not all) of the differences in reading comprehension skill between grades 1 to 4 can be accounted for by individual differences in lexical quality. A number of studies have already shown that word-level verbal abilities and vocabulary knowledge in young readers are closely related to reading comprehension skill on the text level (e.g., Juel, Griffith, & Gough, 1986). The simple view of reading posits that reading comprehension skills are a product of word-level skills and listening comprehension, proposing not only that the knowledge and skills involved in visual word recognition are important for reading comprehension but also that they are the only skills which are specific to reading (Gough & Tunmer, 1986; Kendeou, Savage, & van den Broek, 2009). The present study builds on this research but takes a slightly different approach by replacing the gross

measure of decoding, which is typically used in studies on the simple view of reading, with highly selective measures that refer to the quality of phonological, orthographical, and meaning representations. By using separate measures of the quality of these representations, their joint contribution to grade-level differences in text comprehension skill can be estimated. Moreover, in contrast to previous studies which usually focus either on accuracy-based or time-based measures, we used computerized tasks to assess both the accuracy of and the efficiency of access to lexical knowledge.

Components of Lexical Quality and Reading Comprehension Skill

The manner in which the quality of phonological, orthographical, and meaning representations act in concert to contribute to the reading comprehension skill of primary school children is still not fully understood. Perfetti and Hart (2001) conceive of these representations as tightly connected constituents of word identity. Whenever one or several of the constituents are not well specified or the constituents are only loosely connected, lexical quality is low, resulting in problems for comprehension and other tasks that require word recognition such as spelling or reading aloud. Although the idea that word identity is based on lexical constituents is plausible, individual differences in lexical quality might also be viewed from a slightly different perspective. According to this perspective, the qualities of phonological, orthographical, and meaning representations are linguistic abilities, which form related but partly independent dimensions. Such a componential perspective might be particularly appropriate for beginning readers whose lexical representations are not (yet) tightly connected. For example, Perfetti and Hart (2002) found in a factor analysis of various lexical quality tasks that for less skilled (but not for skilled) adult readers orthographic knowledge, phonological knowledge, and meaning formed separate dimensions. Thus, it is feasible, for example, that primary school children possess accurate and rapidly retrievable phonological representations but poor meaning representations

for many words (cp. Nation & Snowling, 1998). Likewise, other patterns of strengths and weaknesses in various components of lexical quality can exist.

Based on a componential perspective on lexical quality, the quality of one type of lexical knowledge might be more directly relevant for reading comprehension than others. Cognitive models of lexical processing assume that phonological, orthographical, and meaning representations all play a role in visual word recognition (Coltheart, Rastle, Perry & Langdon, 2011) even though their relative importance differs depending on properties of the word (e.g., frequency and regularity) and the task (e.g., lexical decision or naming; Balota, Cortese, Sergent-Marshall, Spieler & Yap, 2004). Moreover, all three types of knowledge are relevant for reading comprehension, implying that individual differences in the quality of phonological, orthographical, and meaning representations should be correlated with reading comprehension skill. However, reading comprehension is fundamentally a process of extracting meaning from written text that relies heavily on the quality of meaning representations. Constraint satisfaction accounts imply that syntactic parsing should benefit from high-quality meaning representations as well (MacDonald & Seidenberg, 2006). In contrast, individual differences in the quality of phonological and orthographical representations exert their effects on reading comprehension in large parts because high-quality phonological and orthographical representations are instrumental for efficient access to word meanings (van Orden & Goldinger, 1994).

Against this background, assuming a path-analytic model that incorporates a direct effect of the quality of meaning representations on reading comprehension skill seems reasonable. However, effects of the quality of phonological and orthographical representations on reading comprehension skill should also be in part mediated by the quality of meaning representations. Nonetheless, high-quality orthographical and phonological representations benefit non-semantic aspects of text comprehension such as syntactic parsing as well (e.g., by releasing working

memory capacity). Hence, direct effects of the quality of these representations on reading comprehension are likely to emerge in addition to the hypothesized indirect effects. Finally, the accuracy of lexical representations is only moderately correlated with the speed with which these representations can be accessed, but both aspects of lexical quality affect reading comprehension. Accordingly, we assumed two separate mediational paths for the effects of accuracy and efficiency of lexical representations on reading comprehension skill. A newly developed set of standardized tests for the assessment of word-level reading skills in primary school children (in German), which combines accuracy and reaction time measures, was used to assess both the accuracy and speed of access to phonological, orthographic, and meaning representations within a common framework (Richter, Isberner, Naumann, & Kutzner, 2012; Richter, Naumann, Isberner, & Kutzner, in press).

Hypotheses

One purpose of this study was to examine the extent to which grade-level differences in reading comprehension skill of primary school children can be explained by individual differences in the quality of phonological, orthographical, and meaning representations. The lexical quality hypothesis implies that the development of proficiency in reading in the first school years corresponds to the development of lexical quality. Accordingly, we expected most (if not all) of the differences in reading comprehension skill that exist between grades 1 to 4 to disappear when differences in lexical quality are taken into account. Another major aim was to examine the joint contribution of the three components of lexical quality to reading comprehension skill. We estimated a path-analytic model that allowed testing two complex hypotheses. First, we assumed that because of their direct relevance for reading comprehension, the quality of meaning representations would partly mediate the effects of phonological and orthographical knowledge on reading comprehension skill. Second, we hypothesized that the

accuracy of lexical representations and the speed of access to these representations both contribute to reading comprehension skill via two largely separate mediational pathways.

Method

Participants

Participants were 247 primary school students from 10 schools in Cologne and Frankfurt am Main (Germany) between the age of six and 10 years ($M = 8.27$, $SD = 1.19$ years). The sample comprised 116 girls and 111 boys (gender information was missing for 20 participants) of which 25 (10.1% of the total sample) were in first grade, 83 (33.6%) in second grade, 57 (23.1%) in third grade, and 82 (33.2%) in fourth grade. According to their parents (or teachers when first language information from the parents could not be obtained – 17.4% of all participants), 189 participants (77.5%) had learned German as their first language (first-language information was missing for 3 participants). Among the students with a native language other than German, the largest groups were those who had learned Turkish (14.7%) and Albanian (14.7%) as their first language, followed by English (8.8%) and Italian (8.8%). The remaining students came from a large variety of linguistic backgrounds, none of which was shared by more than two participants. None of the participating students received special education services. Socio-economic status was measured according to the highest level of professional qualification attained between a participant's father and mother. The parents of 45 participants (18.2% of all parents who provided information on their professional qualification) had completed lower-level vocational training, 30 parents (12.2%) had received a degree from a technical college or vocational school, 91 parents (36.8%) had received a university degree (Bachelor or Master), 19 parents (7.7%) had received some other kind of professional training, and seven participants (2.8 %) had no formal professional qualifications. The distribution of qualification levels of the parents in our sample roughly corresponds to the distribution in the population of Germans aged between 31 and 40

(according to the German Microcensus 2010, Statistisches Bundesamt, 2012) with a slight over-representation of higher levels of professional achievement. The study was conducted at the end of the school year (May/June 2010).

Assessment of Lexical Quality

The quality of phonological, orthographical, and meaning representations were assessed with three newly developed computerized tasks (the word-level tasks of the German computerized reading skills test *ProDi-L*, Richter et al., 2012). The psychometric properties (reliability, factorial, and construct validity) of all tasks were tested extensively and successfully (Richter et al., 2012). Each of the tasks essentially relied on one of the three components of lexical quality which were the focus of this study. For all three tasks, accuracy as well as reaction time for yes/no-responses (provided with two response keys) were recorded to capture the accuracy of the respective type of knowledge and the speed of access to that knowledge. The tasks were embedded in a cover story of an extraterrestrial named Reli who wants to learn the earthlings' language. In the beginning, Reli introduced himself in an animated video clip and explained the general logic of the tasks:

Hello, my name is Reli. I am from the planet Lingua. I have come to your planet to learn more about your language. I have already learned a lot but I still feel uncertain about many things. Can you help me? On my planet, we do everything with computers. Therefore, you will see various words on the computer screen. Sometimes you will also hear them. I would like to learn more from you about these words.

Reli continued to explain the use of the response keys and then walked students through a general practice phase. Before each task, Reli explained the task to the students via an animated video clip. These instructions could be watched repeatedly by the students in case they did not

understand the task in the first trial. In each task, Reli asked the students to correct the errors that he made while trying to learn the earthlings' language. He also provided feedback for responses to two practice items on each task, which were presented before the actual test items.

Phonological knowledge. The quality of phonological representations was measured with a phonological comparison task based on 64 pairs of pseudowords. The first pseudoword in each pair was presented auditorily and the second one visually. Participants indicated whether the written pseudoword matched the spoken pseudoword. Reli explained the task to the students as follows:

You will soon hear a spoken word. This word does not really exist in your language; it is from my extraterrestrial language. After that, you will see a written word on the computer screen. Is the word that you have read the same as the word that you have heard? If it is the same word it is correct. Then press the key with the "J" and the green dot on it. If the words are not the same it is wrong. Then press the key with the "F" and the red dot on it.

The pseudowords were constructed by permuting a set of 80 syllables with a simple consonant-vowel structure (e.g., *gi*, *matozi*, *banufego*; for a similar structure, see Frith, Wimmer, & Landerl, 1996; Wimmer, 1996). Item difficulty was varied by the number of syllables (1-4) in each pseudoword. In 32 of the items, the phonological structure of the spoken and written word matched; in the other half, a mismatch appeared in one or two phonemes.

Orthographical knowledge. The quality of orthographical representations was assessed with a lexical decision task comprising 94 items (47 words and 47 pseudowords). Reli explained the task to the students as follows:

I have fooled around and made up a number of new words. Some of the words are very similar to actual words. Now I do not know any longer which words really exist and which ones I have made up! Can you help me? You will soon see

written words. If you see a word that really exists press the key with the “J” and the green dot on it. If the word does not exist press the key with the “F” and the red dot on it.

In the word stimuli, item difficulty was varied by varying the frequency of the word stimuli and the number of orthographical neighbors. In the pseudoword stimuli, item difficulty was varied by varying the similarity of the pseudowords to actual German words. Pseudowords similar to actual words were constructed by changing the first character of an existing word, such as the pseudoword *Bame* for the word *Name*. Pseudowords dissimilar to actual words were constructed by combining the syllables of two existing words with irregular spellings, such as the pseudoword *Chilance* which was constructed by combining the first syllable of the word *Chili* and the second and third syllables of the word *Balance*. All of the pseudowords were phonologically legal and were matched in length to the word stimuli.

Meaning. The quality of meaning representations was measured with a categorization task comprised of 32 items that included a spoken categorical word (e.g., *animal*) and a written word (e.g., *dog*) presented after a short delay of 200 ms. The participants’ task was to decide whether the written word fell into the category designated by the spoken word. This task essentially requires an access to word meanings (e.g., Thompson-Schill & Gabrieli, 1999). Reli explained the task to the students as follows:

You will soon hear a word such as *fruit*. Several kinds of fruits belong to this word. Apples, pears, and so on, belong to fruits. Subsequently, you will see another word, for example the word *banana*. You are supposed to decide whether the second word belongs to the first one or not. If the word you have read belongs to the word you have heard it is correct. In this case, press the key with the "J" and

the green dot on it. If the two words do not belong together it is false. In this case, press the key with the "F" and the red dot on it.

In half of the items, test word and category name matched (e.g., *Tiere-Hund* [*animal-dog*]), and in the other half they did not match (e.g., *Musikinstrumente-Drucker* [*musical instruments-printer*]). Matching and non-matching words were paralleled by length and frequency. All category names corresponded to commonly known everyday categories (on the basic or superordinate level according to Rosch, 1975), but the degree of familiarity of the categories for primary school children was varied systematically to generate items with different degrees of difficulty (children's familiarity with the words was estimated through three independent ratings from school teachers; $ICC(3,3)_{\text{absolute agreement}} = .84$; cf. McGraw & Wong, 1996). Within the 16 non-matching items, the degree of semantic association between the category name and the test word was varied as well, with a high degree of association in half of the items and a low degree of association in the other half of the items. The degree of semantic association was validated by computing the cosine values of a Latent Semantic Analysis (LSA) for the English translation of the word pairs (<http://lsa.colorado.edu/>; semantic space *General Reading up to 1st Year College*). The eight pairs of category names and test words with a high degree of association ($M = 0.27$, $SD = 0.10$) differed strongly from those with a low degree of association ($M = 0.03$, $SD = 0.03$), $t(15) = 6.83$, $p < .001$, $d = 3.91$.

Test scores. Mean response latencies were computed for each task on the basis of logarithmically transformed response latencies. When a participant responded unusually fast or slow (two standard deviations below or above the item mean response time across all items of this test), response latencies were replaced with the mean response time for this participant. Accuracy scores were computed according to the proportions of correct responses, which were arcsine-transformed to linearize their relationships with the other variables in the study (Cohen, Cohen, West, & Aiken,

2003). Reliability estimates (Cronbach's α) for the accuracy and latency of each task in the present sample are provided in Table 1 (main diagonal). In a study based on a larger sample of primary school children ($N = 536$), Richter et al. (2012) used confirmatory factor analysis to estimate and test two separate measurement models for the response latencies and the accuracy data. The measurement models included the three intercorrelated latent variables Phonological Knowledge, Orthographical Knowledge, and Meaning. Each latent variable was measured with four item parcels that included test items of the corresponding task. The models incorporated a strict homogeneity assumption by restricting the factor loadings of all four item parcels on their latent variable to be equal. Both the measurement model for the accuracy data and the one for the response latencies fitted the data very well (accuracy: $\chi^2(60) = 77.17, p = .07, RMSEA = 0.03, CFI = 1.00$; response latency: $\chi^2(60) = 113.90, p < .001, RMSEA = 0.04, CFI = 1.00$; for a rationale for using these indices of model fit, see the section *Results*). To obtain a model-based reliability estimate, we computed the Average Variance Extracted (Fornell & Larcker, 1981), which is the ratio of the indicator variance explained by the latent variable (the sum of the squared factor loadings) to the total variance of the indicator variables (the sum of the squared factor loadings plus the sum of the error variances), for each latent variable. For the accuracy data, the Average Variance Extracted estimates were .55 (meaning), .72 (orthographical knowledge), and .70 (phonological knowledge). For the response latencies, the Average Variance Extracted estimates were .92 (meaning), .94 (phonological knowledge), and .97 (orthographical knowledge). According to Bagozzi and Yi (1988), estimates of the Average Variance Extracted greater than .50 are usually considered adequate.

Assessment of Reading Comprehension Skill

Reading comprehension skill on the text level was measured with the sub test *Text Comprehension* of ELFE 1-6 (computerized version, Lenhard & Schneider, 2006). ELFE 1-6 is a standardized test widely used in Germany for assessing reading comprehension skill. The subtest *Text*

Comprehension includes 20 short texts with 4 multiple choice items each. The items require identifying specific information in texts, establishing anaphoric references across sentences, and global inferences. Test scores were computed by counting the number of correct responses. In all analyses, the raw test scores (unadjusted for grade level) were used.

Procedure

The study took place in a classroom environment. Children of the same class were tested simultaneously. Each child was seated in front of a laptop and wore headphones. An experimenter explained to the children that they would be performing reading and listening tasks with the extraterrestrial Reli who had come to earth to learn their language. They were also told that participation was voluntary and that they could ask questions when they did not understand a task. Experimenters had been familiarized with the tests and were trained to instruct the children and answer questions in a standardized and motivating manner. Task-specific instructions were given by Reli via the headphones. The assessment of phonological knowledge, orthographical knowledge, and meaning which we report here was embedded in a battery of reading and (for some children) listening comprehension tests on the word- and sentence-level, the overall length and difficulty of which was adapted to the grade level. Each test began with two practice trials (one positive and one negative example) for which feedback was given. During the subsequent test phase, children did not receive feedback on their responses. The first two trials of the test phase served as icebreaker trials and were discarded from the analysis. All subsequent trials were presented in random order. Afterwards, children completed the computer version of the reading comprehension test ELFE 1-6. Finally, a subsample of the children were administered paper-pencil-tests (not included in the analyses reported here) to assess vocabulary and intelligence. Testing was conducted in two separate 45-minute school lessons, with an interval of several days between each lesson.

Results

All hypothesis tests were based on a type-I error probability of .05. Data on reading comprehension were missing from the data set (2.8% of all data points) for 55 participants who were too slow in working on the phonological, orthographical, and meaning tasks and subsequently ran out of time on the subsequent reading comprehension test (12 from grade 1, 19 from grade 2, 13 from grade 3, and 11 from grade 4). Given that reaction times were recorded in the computerized tasks which were administered first, it is safe to assume that missingness was conditional on the predictor variables in the model but not systematically dependent on the missing variable itself (after controlling for the predictor variables). This type of missing data mechanism is called *missing at random* (Rubin, 1976; see also Enders, 2010). In this situation, multiple imputation of missing data (Rubin, 1987) is a suitable option of dealing with missing data, also in sample sizes and rates of missing data such as those in the present study (Graham & Schafer, 1999). In line with recent recommendations (Enders, 2010; Graham, Olchowski, & Gilreath, 2007), we used a large number of 100 imputed data sets, which were generated as well as pooled in the analysis step with Mplus 6 (Muthén & Muthén, 1998-2010). For each model, the variables in the imputation and analysis models were the same. However, none of the models had more parameter restrictions than the imputation model except for the full regression model (Model 3) reported in the section “Grade level differences in reading comprehension skill”. In addition, we reran all analyses with listwise deletion of cases with missing values. The results of these analyses were very close to the results obtained with the imputed data set. Based on the similarity of this comparison, they are not reported here (a report of the results based on the data set with listwise deletion is available from the first author upon request).

Several indices were used to evaluate model fit for the analyses based on Structural Equation Models. Given that the χ^2 statistic strongly depends on sample size and model

complexity, the χ^2 statistic and the associated p value were used as descriptive measures rather than as a significance test. A χ^2/df ratio less than 2 is usually regarded to indicate a good model fit and a ratio less than 3 to indicate an acceptable model fit. A p value greater than .05 was considered as good and a p value between .01 and .05 as an acceptable model fit (Schermelehen-Engel, Moosbrugger, & Müller, 2003). The Root Mean Square Error of Approximation (RMSEA) is a test of close fit, which should be less than .05 in models with good fit and less than .10 in models with acceptable fit (Browne & Cudeck, 1993). Finally, the Comparative Fit Index (CFI) was considered, which is a goodness-of-fit index based on the comparison of the χ^2 value of the hypothesized model to that of a more restrictive baseline model. The CFI compares the hypothesized model to the independence model, which assumes that all error variances and intercorrelations of latent variables are equal to zero and all factor loadings are equal to one. The CFI is suitable for assessing model fit in relatively small samples and ranges from 0 to 1. Values greater than .97 are usually regarded to indicate good fit whereas values greater than .95 are regarded to indicate acceptable fit (Schermelehen-Engel et al, 2003).

Descriptive statistics of all variables, correlations, and partial correlations with grade level partialled out are provided in Table 1. After controlling for grade level, the intraclass correlation for comprehension (students nested within classes) was low ($ICC = .02$).

Covariance Structures in Children With and Without German as Their First Language

Reading development of children with a foreign language background may differ from those who learn to read in their first language. For this reason, we conducted a multisample analysis with Lisrel 8 (Jöreskog & Sörbom, 1996) to test whether the covariance structure of the study variables differed between participants with German as their first language ($n = 189$) and participants with another language than German as their first language ($n = 55$). A strict model

which assumed invariance across groups for all variances and covariances of the seven variables in the study showed an overall acceptable fit to the data, $\chi^2(28) = 54.56, p < .01$, RMSEA = 0.09, CFI = .96. An inspection of the modification indices revealed that the differences between the two groups were mainly due to the fact that the variances of three variables (text comprehension, accuracy of meaning representations, and speed of access to orthographic knowledge) were slightly higher in the sample of children who had learned another language than German as their first language. A more liberal model in which the variances of these three variables were allowed to vary across groups showed a good fit to the data, $\chi^2(25) = 35.60, p = .08$, RMSEA = 0.06, CFI = .98. In sum, the covariance structures in the two groups of participants were assumed to be equivalent and thus data from both groups were included in the following analyses.

Correlations of Lexical Quality and Reading Comprehension Skill

As expected, all three components of lexical quality were correlated with text-level reading comprehension skill, with the strengths of the relationship ranging from medium to high (Table 1). Notably, not only the accuracy rates but also the latencies of responses to the lexical quality tasks were related to reading comprehension skill. The correlations slightly dropped but remained significant when grade level was partialled out. The strongest decrease in strength of relationship was observed in the negative correlation of the response latency in the phonological comparison task with reading comprehension skill, which dropped from -.65 to -.20 when grade level was partialled out.

Grade-level Differences in Reading Comprehension Skill

To test the hypothesis that lexical quality can account for grade-level differences in reading comprehension skill, we used a series of three nested regression models with grade level included in Step 1, the quality of phonological and orthographical knowledge (accuracy and

efficiency of access) included in Step 2, and the quality of meaning representations (accuracy and efficiency of access) included in Step 3 (Table 3). These three nested regression models provide hints at possible mediation relationships between the predictor variables (cf. the stepwise procedure of mediation analysis, Baron & Kenny, 1986). Step 1 provides an estimate of the proportion of variance in reading comprehension skill which may be explained by class level. Step 2 provides a test of whether the quality of form representations (phonological and orthographical knowledge) alone can account for the differences between grade levels. According to the stepwise procedure of mediation analysis proposed by Baron and Kenny (1986), the idea that the quality of phonological and orthographical knowledge mediates class-level differences would be supported if class level was no longer significant after including these variables in the model. Likewise, Step 3 provides a first test of whether and to what extent effects of the quality of phonological and orthographic representations on reading comprehension skill are mediated by the quality of meaning representations.

In Step 1, grade level accounted for 20% of the variance in reading comprehension skill. Reading comprehension skill monotonically increased with participants' grade level (see Table 2 for descriptive statistics of all variables by grade level). However, both the accuracy and speed of responses to the lexical quality tasks also increased with grade level (Figure 1). When the indicators of the quality of phonological and orthographical knowledge were included as predictors in Step 2, grade level no longer contributed to the explanation of variance in reading comprehension skill. Instead, the accuracy of phonological and orthographical knowledge exerted significant effects on reading comprehension skill. The proportion of explained variance more than doubled to 53%. In Step 3, the accuracy and the efficiency of access to meaning representations had additional significant effects on reading comprehension skill but the effects of the accuracy of phonological and orthographical knowledge remained strong and significant.

The proportion of explained variance increased to 57%. This pattern of results suggests that the quality of meaning representations does not fully mediate the effects of these variables. Still, it may be a partial mediator, a possibility which is explored in the path-analytic model estimated in the next step of analysis. In sum, grade level differences disappeared when indicators of lexical quality were included in the model. The three components of lexical quality accounted for a considerable proportion of variance in reading comprehension skill that by far exceeded the proportion accounted for by grade level.

Components of Lexical Quality and Reading Comprehension Skill: A Path-analytic Model

Finally, we used structural equation modeling to test the hypothesized path-analytic model that assumed the quality of meaning representations to mediate the effects of the quality of orthographical and phonological knowledge on reading comprehension skill. Two separate mediational pathways were assumed for the accuracy of lexical knowledge and the efficiency of access to this knowledge. The Maximum Likelihood procedure implemented in MPlus 6 (Muthén & Muthén, 1998-2010) was used for parameter estimation. Considering that the purpose of the model was to explain individual differences in reading comprehension skill that exist between students within each grade level, we used residuals with grade level partialled out as the basis for parameter estimation (Table 1, correlations above the main diagonal). The parameter estimates (standardized solution) are provided in Figure 2. In line with the expectations, the accuracy and speed of access of orthographical and phonological knowledge exerted substantial and significant indirect effects on reading comprehension skill through the quality (accuracy and efficiency of access) of meaning representations (Sobel tests for all four indirect effects: $Est./SE > 2.26$, $p < .01$, one-tailed). Moreover, the accuracy of orthographical knowledge and the accuracy of phonological knowledge had direct effects on reading comprehension skill. In addition to the hypothesized relationships, a path from the accuracy of orthographical representations to the

latency of access to meaning representations was included in the model (Figure 2). With this modification, the model had an excellent fit to the data, $\chi^2(6) = 10.10$, $p = .12$, RMSEA = 0.05, CFI = 0.99. To further corroborate the hypothesized mediation model, we tested two alternative models in which (a) the two indicators of the quality of orthographical knowledge and (b) the two indicators of the quality of phonological representations served as mediators. Neither of the two models fit the data well, even when all possible direct effects of the distal predictors to reading comprehension skill were permitted (quality of orthographical representations as mediator: $\chi^2(5) = 39.79$, $p < .001$, RMSEA = 0.17; quality of phonological representations as mediator: $\chi^2(5) = 21.18$, $p < .001$, RMSEA = 0.11).

Discussion

By demonstrating that grade level differences in reading comprehension skill could be fully accounted for by individual differences in lexical quality, our results underscore the importance of lexical representations for reading comprehension skills in developing readers. Overall, indicators of three different types of lexical knowledge – phonological, orthographical, and meaning representations – together explained nearly 60% of text-level reading comprehension skill. Remarkably, not only the accuracy of these representations was related to reading comprehension skill but also the speed with which they could be retrieved. This finding is consistent with the idea that efficient access to lexical representations releases working memory resources which can be used for resource-dependent higher-level comprehension processes (Perfetti, 1985).

Given the correlational and cross-sectional character of our data, any causal conclusions are tentative and need to be substantiated by further research which includes longitudinal data and training experiments. As a heuristic for this research, the path-analytic model tested here

suggests a specific causal order of the three components of lexical quality. The qualities of phonological and orthographical representations were linked to reading comprehension skill through the quality of meaning representations. High-quality representations of word meanings seem to be essential to the comprehension of written texts but also seem to depend on high-quality phonological and orthographical representations (whereas the reverse does not hold). One consequence of this asymmetric relationship of form and meaning representations is that reading comprehension difficulties can arise because of deficits in semantic representations even when the quality of phonological and orthographical representations is high (cf. Nation & Snowling, 1998, 1999). In addition, the path-analytic results suggest that the quality of semantic representations might serve as a mediator in reading development that explains how good phonological and orthographical skills in beginning readers eventually lead to good text comprehension skills (as suggested by the predictive power of rapid naming for reading development, Landerl & Wimmer, 2008). Similarly, the integration of phonological, orthographical, and meaning representations to closely connected constituents, which characterizes high-quality lexical representations of developed readers, might proceed in such a way that representations of word forms are established first, whereas their integration with meaning representations follows at a later point. To test these assumptions, longitudinal studies are needed that employ tasks that would tap into the relationships of various lexical constituents (including tasks that test for form-meaning confusions, Perfetti, 2007).

Interestingly, the accuracy of both orthographical representations and phonological representations exerted direct effects on reading comprehension skill over and above the indirect effect from the accuracy of meaning representations. The direct effect of orthographical representations was much stronger than that of phonological representations. This pattern of results is consistent with the idea that high-quality orthographical and phonological

representations are relevant for non-semantic aspects of sentence and text processing (e.g. syntactic parsing). In addition, although models of visual word recognition agree that both phonological and orthographical information can be used to recognize written words, orthographic representations of word forms clearly become more and more relevant as reading skill develops. In particular, these representations become the primary basis for recognizing frequent words (Coltheart et al., 2001). Moreover, the quality of orthographical representations reflects the amount of reading practice (Stanovich & West, 1989), which might contribute to its strong relationship with reading comprehension skill. Consonant with these and the present results, Hersch and Andrews (2012) found that spelling ability, which taps into the quality of orthographical representations, predicted sentence comprehension in skilled readers, in particular reduced reliance on context in word access.

Finally, it should be noted that the accuracy in the phonological knowledge task and in the meaning task was high overall and associated with a variance that was lower than the variance in the orthographical knowledge task. The lower variance could have been the result of the relatively low internal consistency of the phonological knowledge task (Table 1). Thus, we cannot rule out completely that the slightly weaker relationship of this component with reading comprehension skill might have resulted from the way this component was assessed. However, the correlations of the accuracy rates in both the phonological knowledge task and in the meaning tasks with comprehension were substantial and did not disappear when all three types of knowledge were included in the same model. This pattern of effects suggests that the results were not biased by ceiling effects in the phonological knowledge tasks and the meaning tasks.

Another core finding of this study was that the accuracy of lexical representations and the efficiency of access to these representations contributed to reading comprehension skill via two (largely) separate paths. Both types of variables were only moderately correlated with each other,

suggesting that accuracy and speed of access should be regarded as two different and partly independent facets of lexical quality. A practical implication of this finding for educational settings is that a detailed assessment of reading skills on the word level should go beyond screening tests which often focus on the accuracy of lexical representations or mix accuracy and speed of access when responses are to be provided within a certain time frame (e.g., the word-level subtests of the Woodcock Reading Mastery Test, Woodcock, McGrew, & Mather, 2001, vocabulary tests, word attack measures; or CBM measures of reading, Deno, 1985). Similarly, researchers interested in the role of lexical skills in reading should take care to use measurements that clearly distinguish the accuracy of the underlying knowledge and the speed with which these representations can be retrieved. Both the accuracy and the speed of access to lexical representations seem to contribute to reading comprehension skill, but the effects for accuracy are much larger compared to speed. This pattern of findings suggests that accuracy is the more relevant aspect of lexical representations, underscoring the primacy of code quality stressed by the lexical quality hypothesis (Perfetti & Hart, 2001) which marks an advancement of the earlier verbal efficiency theory. Reading comprehension skill does not benefit much from lexical representations which are low in code quality, regardless of how fast they can be retrieved from long-term memory.

Aside from the cross-sectional design of this study, other aspects should be kept in mind as potential limitations to the generalizability of results. One potential limitation is that the study was not based on a random sample of primary school children (although the socioeconomic background of the children in the sample appears to have matched the general population quite well) and that the subsamples drawn in grades 1 and 3 are relatively small. Thus, a replication with a larger sample with evenly distributed age groups would be desirable to strengthen the generalizability of the results. A second limitation is that specific tasks were used to assess each

of the quality of phonological, orthographic, and meaning representations. Consequently, the extent to which the current results are task dependent remains unclear. A third limitation is that the indicators of lexical quality used in this study are likely to be correlated with component skills of comprehension on the sentence and text level. Previous research with primary school children has shown that metacomprehension and inference skills explain a considerable amount of variance in reading comprehension even when word-level skills had been controlled for (e.g., Cain, Oakhill, & Bryant, 2004; see also Oakhill & Cain, 2012). Likewise, as posited by the Simple View of Reading (Gough & Tunmer, 1986), oral language comprehension skills beyond the word level predict reading comprehension over and above decoding skills (Kendeou, Savage, & van den Broek, 2009; Kendeou, van den Broek, White, & Lynch, 2009). The present study did not consider any component skills of comprehension beyond the word level as predictors of reading comprehension skill. For this reason, a considerable amount of the large proportion of variance explained by indicators of lexical quality in the present study might have been shared variance with higher-order language comprehension skills. In other words, the unique contribution of lexical quality might be smaller when a broader array of linguistic skills is taken into account.

We are currently undertaking a longitudinal study with two cohorts that will be followed from the beginning of primary school until the end of fourth grade. In this longitudinal study, the lexical quality assessments used in the present study will be applied along with other tests that assess various reading comprehension skills ranging from the word to the sentence level. With this design, we will be able to test the causal model suggested by the present results with a more stringent analyses and determine the unique contribution of aspects of lexical quality to reading comprehension. The clear and consistent results of the current study make us optimistic about this endeavor.

References

- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Sciences, 16*, 74-94.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283-316.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.
- Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*, 230-258.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*, 31-42.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Coltheart, M., Rastle, K., Perry, C., & Langdon, R. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204-256.
- Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.
- Enders, C. G. (2010). *Applied missing data analysis*. New York: Guilford.
- Fornell, C., & Larcker, D.F. (1982). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*, 39-50.
- Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German-

- and English-speaking children. *Scientific Studies of Reading*, 2, 31-54.
- Gibbs, P., & van Orden, G. C. (1998). Pathway selection's utility for control of word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1162-1187.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6-10.
- Graham, J.W., Olchowski, A.E., & Gilreath, T.D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8, 206-213.
- Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 1–29). Thousand Oaks, CA: Sage.
- Hersch, J., & Andrews, S. (2012). Lexical quality and reading skill: Bottom-up and top-down contributions to sentence processing. *Scientific Studies of Reading*, 16, 240-262.
- Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*, 78, 243–255.
- Kendeou, P., van den Broek, P., White, M. J., Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, 101, 765-778.
- Kendeou, P., Savage, R., & van den Broek, P. (2009). Revisiting the Simple View of Reading. *British Journal of Educational Psychology*, 79, 353-370.
- Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology*, 100, 150-161.

- Lenhard, W., & Schneider, W. (2005). *ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässer* [ELFE 1-6: A reading comprehension test for grades one to six]. Göttingen, Germany: Hogrefe.
- MacDonald, M. C., & Seidenberg, M. S. (2006). Constraint satisfaction accounts of lexical and sentence comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.). London: Academic Press.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46.
- Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's Guide* (6th ed.) Los Angeles, CA: Muthén & Muthén
- Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language, 39*, 85-101.
- Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition, 70*, B1-B13.
- National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Rockville, MD: National Institute of Child Health and Human Development.
- Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16*, 91-121.
- Perfetti, C. (1985). *Reading ability*. New York: Oxford University Press.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357-383.

- Perfetti, C., & Hart, L. (2001). The lexical bases of comprehension skill. In D. Gorfien (Ed.), *On the consequences of meaning selection* (pp. 67-86). Washington, DC: American Psychological Association.
- Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam: John Benjamin.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56-115.
- Richter, T., Isberner, M.-B., Naumann, J., & Kutzner, Y. (2012). Prozessbezogene Diagnostik von Lesefähigkeiten bei Grundschulkindern [Process-based measurement of reading skills in primary school children]. *Zeitschrift für Pädagogische Psychologie*, *26*, 313-331.
- Richter, T., Naumann, J., Isberner, M.-B., & Kutzner, Y. (in press). *ProDi-L: Prozessbezogene Diagnostik von Lesefähigkeiten bei Grundschulkindern* [ProDi-L: Process-based assessment of reading skills in primary school children] [Computerized test]. Göttingen, Germany: Hogrefe.
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*, 532-547.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581-592.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*, 23-74.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*, 402-433.

- Statistisches Bundesamt (2012). *Bildungsstand: Bevölkerung 2010 nach Bildungsabschluss und Altersgruppen in Deutschland* [online document] [Educational background: Population in 2010 by educational achievement and age group in Germany]. Retrieved April 25, 2012, from <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Bildungsstand/Tabellen/BildungsabschlussAlterBB.html?nn=50760>
- Thompson-Schill, S. L., & Gabrieli, J. D. E. (1999). Priming of visual and functional knowledge on a semantic classification task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 41-53.
- Van Orden, G. C., & Goldinger, S. D. (1994). The interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1269–1291.
- Wimmer, H. (1996). The nonword reading deficit in developmental dyslexia: Evidence from children learning to read German. *Journal of Experimental Child Psychology*, *61*, 80-90.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.

Table 1

Descriptive Statistics, Zero-Order Correlations (Above Main Diagonal), Partial Correlations with Class Level Partialled Out (Below Main Diagonal) and Reliability Estimates (Cronbach's α , Main Diagonal) for All Variables

	Raw		Transformed		Correlations (Above Main Diagonal), Partial Correlations (Below Main Diagonal) and Cronbach's α (Main Diagonal)						
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7
1 Phonological Knowledge (ACC)	.83	0.13	1.17	0.17	(.59)	.43	.47	.02	-.23	-.22	.51
2 Orthographical Knowledge (ACC)	.67	0.21	1.01	0.27	.32	(.79)	.46	-.38	-.65	-.57	.72
3 Meaning (ACC)	.87	0.12	1.24	0.19	.40	.28	(.76)	.46	-.38	-.65	.57
4 Phonological Knowledge (RT)	1994	607	7.46	0.29	.12	-.27	-.08	(.99)	.46	-.38	-.65
5 Orthographical Knowledge (RT)	2096	1041	7.48	0.45	-.10	-.50	-.06	.52	(.97)	.46	-.38
6 Meaning (RT)	1941	982	7.38	0.40	-.07	-.35	-.08	.66	.57	(.97)	-.49
7 Text Comprehension	10.44	4.90	–	–	.44	.62	.42	-.20	-.37	-.33	(.92) ^a

Note. RT: Reaction Time in ms (raw and after logarithmic transformation of reaction times to individual items). ACC = Proportion of correct responses (raw and after arcsine transformation). Correlations and partial correlations were computed for logarithmically transformed reaction times and arcsine-transformed proportions of correct responses. $|r| > .12: p < .05$; $|r| > .16: p < .01$; $|r| > .21: p < .001$ (two-tailed).

^aReported by Lenhard & Schneider (2005) for an independent sample of 100 primary school children (grades 2 to 4)

Table 2

Descriptive Statistics by Grade Level for All Variables

	Raw data				Transformed data (arcsine/ log. transformed)			
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 1	Grade 2	Grade 3	Grade 4
	<i>M (SE_M)</i>	<i>M (SE_M)</i>	<i>M (SE_M)</i>	<i>M (SE_M)</i>				
1 Phonological Knowledge (ACC)	.77 (.03)	.80 (.01)	.84 (.02)	.88 (.01)	1.09 (0.04)	1.12 (0.02)	1.19 (0.02)	1.23 (0.01)
2 Orthographical Knowledge (ACC)	.42 (.03)	.54 (.02)	.73 (.02)	.82 (.02)	0.71 (0.04)	0.85 (0.02)	1.08 (0.03)	1.21 (0.02)
3 Meaning (ACC)	.77 (.03)	.83 (.01)	.92 (.01)	.91 (.01)	1.10 (0.04)	1.17 (0.02)	1.31 (0.02)	1.31 (0.02)
4 Phonological Knowledge (RT)	2443 (157)	2086 (67)	1968 (76)	1781 (52)	7.61 (0.08)	7.52 (0.03)	7.44 (0.04)	7.36 (0.03)
5 Orthographical Knowledge (RT)	3420 (296)	2591 (129)	2054 (126)	1583 (64)	7.83 (0.11)	7.68 (0.05)	7.44 (0.05)	7.21 (0.03)
6 Meaning (RT)	3285 (366)	2143 (83)	1818 (78)	1410 (52)	7.80 (0.12)	7.53 (0.03)	7.36 (0.04)	7.12 (0.03)
7 Text Comprehension	6.60 (0.67)	8.35 (0.40)	11.89 (0.59)	12.70 (0.57)	--	--	--	--

Note. RT: Reaction Time in ms (raw and after logarithmic transformation of reaction times to individual items). ACC = Proportion of correct responses (raw and after arcsine transformation).

Table 3

Parameter Estimates for Nested Regression Models with Reading Comprehension Skill as Outcome Variable and Class Level and Lexical Quality Variables as Predictors.

	Model 1			Model 2			Model 3		
	<i>Estimate</i> (SE)	<i>Est./SE</i>	ΔR^2	<i>Estimate</i> (SE)	<i>Est./SE</i>	ΔR^2	<i>Estimate</i> (SE)	<i>Est./SE</i>	ΔR^2
Intercept	6.20 (1.16)	5.35		4.97 (8.30)	0.60		5.56 (8.44)	0.66	
Grade 2 vs. 1 (dummy-coded: 1 vs. 0)	2.24 (1.29)	1.74*	.02	-0.28 (1.12)	-0.25	.00	-0.58 (1.12)	-0.52	.00
Grade 3 vs. 1 (dummy-coded: 1 vs. 0)	5.51 (1.33)	4.14***	.08	-0.14 (1.20)	-0.12	.00	-0.77 (1.21)	-0.64	.00
Grade 4 vs. 1 (dummy-coded: 1 vs. 0)	6.56 (1.27)	5.17***	.12	-0.57 (1.23)	-0.46	.00	-1.29 (1.26)	-1.02	.00
Phonological Knowledge (ACC)				6.61 (1.87)	3.50***	.03	4.83 (1.94)	2.48**	.03
Phonological Knowledge (RT)				-0.69 (1.24)	-0.55	.00	0.58 (1.30)	0.45	.00
Orthographical Knowledge (ACC)				16.11 (2.46)	6.54***	.10	13.68 (2.52)	5.44***	.08
Orthographical Knowledge (RT)				-2.13 (0.95)	-2.25*	.01	-2.00 (1.03)	-1.94*	.03
Meaning (ACC)							4.80 (1.60)	3.01**	.04
Meaning (RT)							-1.56 (0.87)	-1.79*	.02
Goodness of fit	$R^2 = .20$ (0.05)	4.06***		$R^2 = .53$ (0.05)	11.00***		$R^2 = .57$ (0.05)	12.02***	

Note. * $p < .05$, ** $p < .01$, *** $p < .001$ (one-tailed).

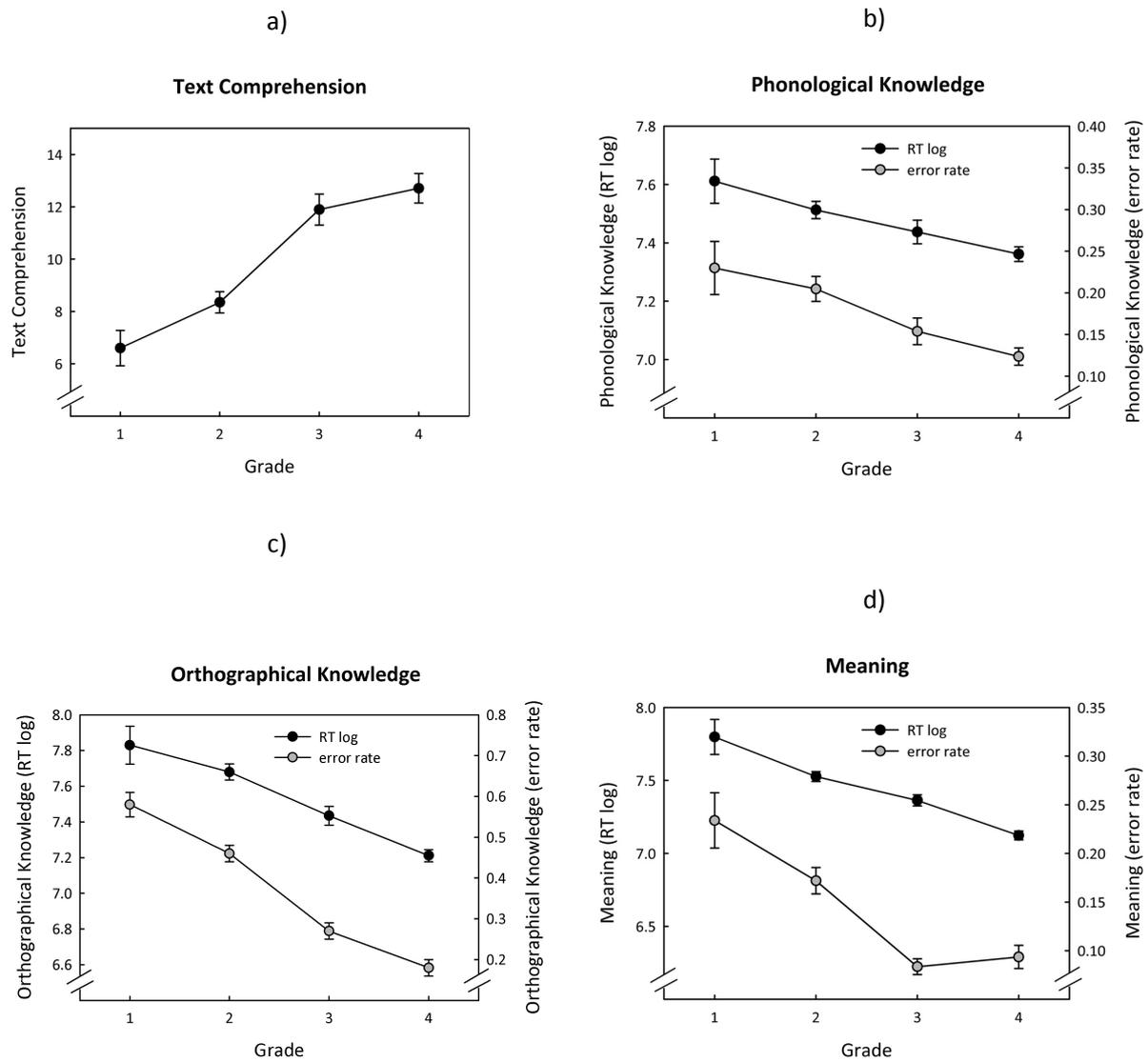


Figure 1. (a) Grade level differences in the reading skill on the text level (ELFE 1-6, Lenhard & Schneider, 2005), accuracies (error rates) and reaction times (logarithmically transformed) of (b) phonological knowledge, (c) orthographical knowledge, and (d) meaning representations.

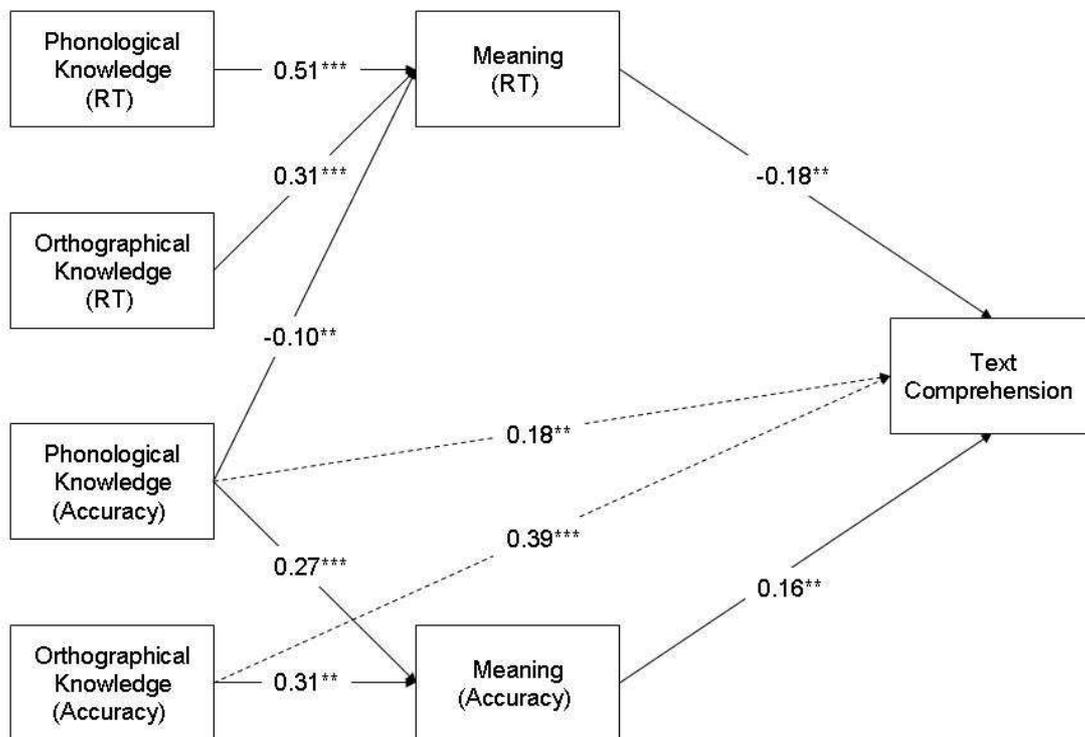


Figure 2. Parameter estimates (standardized solution) for the hypothesized mediation model.

Dashed lines represent direct effects. * $p < .05$, $p < .01$, $p < .001$ (two-tailed).