Effects of Oral Language and Decoding Skills on Reading Comprehension Performance across Multiple Assessments: A Longitudinal Study

Svetlana Kapalková[1], Kamila Polišenská[2], Andrej Mentel[3], and Lydia Vencelová[4]

[1] Department of Communication Disorders, Faculty of Education, Comenius University in Bratislava, Slovakia, https://orcid.org/0000-0002-0052-6884

[2] Division of Human Communication, Development and Hearing, The University of Manchester, UK

[3] Institute of Social Anthropology, Faculty of Social and Economic Sciences, Comenius University in Bratislava, Slovakia, https://orcid.org/0000-0002-8556-7625

[4] Banská Štiavnica, Slovakia

Correspondence should be addressed to Kamila Polišenská, Division of Human Communication, Development and Hearing, The University of Manchester, Oxford Road, Manchester M13 9PL, UK

kamila.polisenska@manchester.ac.uk https://orcid.org/0000-0001-7405-6689

Statements and Declarations: No potential competing interest was reported by the authors.

Acknowledgments: This work was supported by grants APVV-20-0126 and APVV-0410-11 from the Slovak Research and Development Agency (APVV).

Keywords: Reading comprehension, Simple view of reading, Oral comprehension, Decoding, Transparent orthography, Assessment

Abstract

Reading comprehension (RC) is a multi-faceted construct but is often assessed with a single instrument. Previous research has highlighted that commonly used RC tests are only mildly correlated and vary in the skills they assess, including the differential contribution of oral language and decoding to children's performance. Our study, framed within the Simple View of Reading model, examined the contribution of underlying component skills for multiple RC measures and evaluated whether the contribution of decoding and oral language skills changes according to the RC test used and developmental level. Two hundred Slovak-speaking children were assessed across two time points, using multiple RC tests and measures of decoding and oral language skills. The RC tests showed weak to moderate correlations, echoing findings from other languages. At the end of Year 1, the contribution of decoding and oral language to RC was similar in the Slovak transparent orthography. At the end of Year 2, the contribution of oral language had increased threefold, while decoding remained unchanged from Year 1. Crucially, there were also differences between the tests, with some more reliant on oral language. The results highlight the potential benefits of increasing understanding of the differential effects of the component skills in commonly used RC assessments as an aid to interpretation of children's scores. Such an approach could not only identify children with poor scores, but also pinpoint where weaknesses lie in the underlying components so intervention targets could be formulated accordingly.

Introduction

Research has demonstrated a relationship between reading comprehension (RC) and academic achievement (Cain & Oakhill, 2006). Cain (2009) discussed how the consequences of poor reading comprehension go beyond written narrative understanding and are revealed in the national assessments of English, Maths and Science, where poor comprehenders scored lower than good comprehenders. RC is a complex cognitive construct consisting of multiple component skills and accurately measuring RC remains difficult because of the involvement of multiple student-, task-, and situation-level factors (Clemens et al., 2021).

In order to increase understanding about the contribution of different components of RC, we have conceptualized our study within The Simple View of Reading (SVR) model (Gough & Tunmer, 1986; Hoover & Gough, 1990). According to the SVR, a child having difficulty in comprehending written text might also struggle with decoding (sounding out and/or recognizing written words) or with oral language comprehension which Hoover and Gough (1990) see as "the ability to take lexical information (i.e., semantic information at the word level) and derive sentence and discourse interpretations" (p. 131). Deficiency in either written decoding or oral comprehension (OC) can cause poor RC and some children have deficits in both areas. Another advantage of the SVR model is that it can be applied to different developmental stages. There is evidence that there are changes between decoding, OC and RC, with the role of oral language increasing as children's decoding skills improve with age and reading ability (e.g., Garcia & Cain, 2014).

Decoding and OC have been shown to be significant sources of variance in RC across different languages and types of orthographies (e.g., Ho et al., 2012; García & Cain, 2014; Memisevic et al., 2020; Nation, 2019). Florit and Cain (2011) analysed data from 33 studies, across nine different orthographies, and with children who had different amounts of schooling. Their analyses split the data into groups of children with 1-2 years and 3-5 years of reading instruction. Their results suggested that the role of the two components of the SVR changed according to the transparency of orthography and length of schooling. For English and its deep orthography, reading accuracy was more important than OC in the early stages (1-2 years of instruction), but listening comprehension was a better predictor compared to nonword reading for children with 3-5 years of instruction. In contrast, listening comprehension was a strong predictor with transparent orthographies, even in the early years of instruction. The study by Florit and Cain (2011) is an important contribution to the discussion of the effects of orthographic transparency on reading development framed within the SVR, but it is a cross-sectional study, so replicating its findings with longitudinal evidence based on cross-linguistic or individual-language studies could be beneficial. Our study, which evaluates reading skills in Slovak (a language with a transparent orthography), contributes by evaluating a range of RC measures with varying reliance on decoding and OC. It is also the first study to design and compare RC measures for Slovak-speaking children and assess the contribution of the components of SVR for those measures cross-sectionally and longitudinally.

Reading and reading comprehension in Slovak-speaking children

Children growing up in Slovakia do not generally start formal literacy instruction before they enter primary school (usually in the September following a child's 6th birthday). In the first few months of school, children will focus on learning the alphabet and basic phonics skills, and they can usually decode and read simple texts by the end of Year 1. By the end of Year 2, most children will be able to read fluently at the speed of approximately 70-80 words per minute (Caravolas, 2017).

There is little research into reading skills beyond the word-level in Slovak speaking children, as discussed by Caravolas (2017) in her chapter on *Learning to read in Czech and Slovak*. While some studies have been conducted in Czech which is typologically related and has a similar orthography, research in Slovak is more limited and focuses primarily on precursors to literacy and predictors of decoding. Indeed, Caravolas (2017, 2022) called for longitudinal studies including a variety of comprehension reading paradigms which would allow evaluation of models developed for English to establish if those models are also valid for learners of other orthographies, including Slovak. Slovak is a language with a transparent orthography, as demonstrated by Caravolas et al. (2012). Their study reported the estimated consistency of grapheme-phoneme mapping across four languages, but here we concentrate on English as a reference point and Slovak. Consistency values were derived from children's printed word corpora and ranged from 0 to 1. The value for English was .72, while it was .90 in Slovak, demonstrating higher consistency of grapheme-phoneme mappings in Slovak compared to English. Although our study focuses on RC, the guiding framework is the SVR model and decoding is a critical component within this model. Therefore, the next section reviews precursors of decoding skills in Slovak-speaking children, before turning to RC comprehension assessments in Slovak.

Zubáková and Mikulajová (2021) analysed precursors of early literacy in Slovak in a sample of 211 children. The children were 6 years of age on average at the point of initial testing (the year before entering primary school) and were then followed for 2.5 years. The study evaluated the early precursors of reading and writing, and children were assessed on nonverbal IQ, phonological awareness, letter knowledge, RAN, and working memory. The results showed that two months after school entry, phonological awareness and letter knowledge were the strongest predictors, explaining almost half of the variance on a one-minute word reading test. After 9 months of formal reading instruction (end of Year 1), RAN was identified as the strongest predictor, followed by phonological awareness, but with letter knowledge no longer a significant predictor. This pattern of results also held at the end of Year 2, which was the final testing point. In addition to concurrent predictors, the study also assessed RAN and phonological awareness skills as longitudinal predictors of reading skills. As assessed by a 1-minute word reading test, RAN was the strongest longitudinal predictor for reading skills in Slovak children, followed by phonological awareness which was a significant but less strong predictor. In the early stages of reading acquisition in Slovak, it appears that phonological awareness and letter knowledge are strong predictors. However, these predictors are soon replaced by RAN which becomes the strongest predictor, and this is also the case longitudinally.

There are two tests used in Slovak clinical practice to assess RC. Váryová (2012) combines assessment of decoding skills with comprehension questions but with the same reading stimulus. The test can be only administered if the whole of the text is read aloud. It consists of six multiple-choice questions, asking about information explicitly presented in the text and a child chooses from three possible answers. There is a possibility to follow up with three open questions. The test is administered individually and the child answers orally. The second test (Mikulajová et al., 2012) which is used to assess RC is a cloze reading test where children are required to select missing words in short passages.

Both tests are problematic for assessing reading comprehension. The tests assess decoding skills and fluency at the same time as comprehension. This is by design in the test created by Váryová (2012), but with the Mikulajová et al.'s (2012) test, the cloze reading test confounds poor decoding skills with weak comprehension (Caravolas et al., 2019). Váryová (2012) herself points out the 1972 text used in this test is archaic and includes concepts which are unfamiliar to children nowadays. It should also be noted that while the internal consistency and reliability of the decoding parts of the test are reported to be high, no validity or reliability is provided for the comprehension part, making it difficult to evaluate its psychometric properties and therefore its suitability.

The only experimental study beyond word-level reading skills in Slovak was a cross-linguistic study investigating the effects of orthographic consistency on the development of RC across three consistent orthographies (Czech, Slovak and Spanish) and inconsistent English orthography. Caravolas et al. (2019) compared precursors of decoding and language comprehension across four different languages: English (N=179, age range 53-66, mean=60.20, SD=3.67), Spanish (N=188, age range 61-73, mean=66.76, SD=3.65), Slovak (N=194, age range 62–81, mean 71.85, SD=3.78) and Czech (N=135, age range 64-80, mean=72.00, SD=3.73). In their study, variations in decoding skills in Year 1 (i.e., first year of reading instruction when children enter the formal school system at the age of 6 years) predicted RC levels in Year 2. For Slovak, pre-school language skills also predicted RC skills in Year 2, but this was not the case for English. In Slovak, RC was only assessed with a single multiple-choice reading measure that was based on the selection of one correct option out of three pictures, and this might have led to a stronger reliance on decoding, as the authors pointed out (p. 397).

Clearly there is limited knowledge about reading comprehension development in Slovak. The information that is available comes from international comparative tests (PISA, Avvisati et al., 2019), and Slovak children achieved low comprehension scores in the 2018 results: 458 points (bottom 8th place); average OECD = 487 points, range 471-523. To make informed recommendations on how to improve children's performance, greater insight into the RC of Slovak-speaking children would be valuable and more background information about the RC measures designed in Slovak and children's developmental trajectories on those measures would benefit educators and clinicians.

Differences in Reading Comprehension Measures

A growing number of studies have highlighted the complexity of RC and how assessment measures may fail to identify the same children as poor comprehenders when different measures are used (e.g., Clemens et al., 2021; Francis et al. 2006; Keenan et al., 2008; Keenan & Meenan, 2014). For example, a study by Rimrodt et al. (2005) compared the results of three RC tests, and only 9.4% of the sample were identified as having a RC deficit by all three tests. Keenan and Meenan (2014) compared four tests used to diagnose a reading deficit. The different formats of the tests were: multiples-choice questions, open-ended questions, selection from a set of pictures to match the text, and a cloze test. Within their sample of 995 English-speaking children, age range 8-18 with a median age of 11.17 years, only 20 children were consistently identified by all four tests, and overall, “the odds were less than half” (p.133) that a child would receive a reading deficit diagnosis if a different test had been used.

More recent studies have also found few correlations between RC tests and consistency of clinical labels (e.g., Cutting & Scarborough, 2006). Collins et al. (2018) carried out a meta-analysis on English-speaking studies, looking at sources of variance that contribute to RC test scores. Their results concluded that one source of variance is the use of different response formats. Collins and colleagues found that the achievement gap between poor and typically developing readers was larger for picture selection and open-ended questions than for retell tasks. Therefore, the choice of the reading measure may affect the size of the gap between poor and typical score, raising implications for individual students as well as services. Collins et al. (2018) called for further research into reliability and validity in the assessment of reading measures, and this should be extended to other languages and writing systems.

Colenbrander et al. (2017) compared two commonly used comprehension assessments: the Neale Analysis of Reading Ability (NARA, Neale, 1999) and the York Assessment of Reading for Comprehension (YARC, Snowling et al., 2012) with ninety-five English-speaking children aged 8 to 12. They found that NARA comprehension scores were more dependent on decoding skills than YARC scores. The NARA included more comprehension questions and passages ranged more widely in difficulty. Similar to other studies, Colenbrander et al. (2017) reported that 15-34% of children received different diagnoses across tests. The authors suggested that the differences between the tests may be even larger in younger children because of differences between the tests' reliance on decoding ability, leading to calls for further research with a younger age group. The sample in Colenbrander et al. had 4 to 7 years of schooling, providing an opportunity for reading and comprehension to undergo crucial development. Other factors might have accounted for the differences found. For example, NARA was always administered before YARC within one testing session, children of different ages answered different questions which may have targeted different aspects of comprehension/decoding, and there were different discontinuation procedures (NARA had a discontinuation rule, while YARC did not). Therefore, more research is needed with a narrower age group and measures that are similar within the sample studied. The focus of Colenbrander et al. (2017) study was on consistency of the diagnosis by different RC tests, but greater insight into the components of the tests would have been useful, for example, whether the test components reflect more decoding or OC skills, as was suggested by a study by Nation and Snowling (1997) who compared a different pair of tests.

Nation and Snowling's (1997) study compared two tests: the Neale Analysis of Reading Ability (NARA; Neale, 1989) and the Suffolk Reading Scale (Hagley, 1987) to understand the role of decoding and OC in those assessments. The study found that the RC tests assessed different aspects of the reading process and that performance in both tests was influenced by decoding skills, but listening comprehension only accounted for additional variance in the NARA. This early finding highlights how the role of OC interacts with type of the reading test.

Keenan et al. (2008) further investigated the comparability of popular RC measures used in research and clinical practice in the United States to examine if these tests measure the same construct. Their cross-sectional study also had a developmental aspect, with the aim of establishing the extent of developmental differences in test measures. The study included 510 children speaking English as their first language, aged 8 to 18 years (median = 10.5). The researchers compared test performance on different reading measures, ages, and reading ability levels, with the correlation between RC measures ranging from r = .31 to r = .71 and the majority between r = .41 and .51. This result was seen by the authors as "rather modest given that they are all purporting to be measures of the same construct."

(p.288). Those correlations were repeated with a larger sample (N = 995, age range 8-18 years, mean age 11.17) in a study by Keenan and Meenan (2014) and remained modest.

When comparing the role of decoding skills and listening comprehension across different ages and reading measures, Keenan et al. (2008) replicated a previously reported finding that decoding skills accounted for more variance in younger children compared to older children (e.g., Catts et al., 2005). The novel aspect of their finding was highlighting large discrepancies between the reading tests in relation to developmental differences, once again confirming inconsistencies between the reading tests but also that the lack of consistency might be exacerbated by age/developmental stage of the child. Keenan et al. (2008) evaluated developmental differences in a cross-sectional study with children across a wide age range (8-18 years). As the study was cross-sectional rather than longitudinal, changes in individual children and their responses to particular tests cannot be taken into account within a repeated-measures design. Keenan et al. suggested that the greater reliance on decoding in younger children is more pronounced in some tests. However, it is not clear if this finding is specific to the age group Keenan et al. studied (youngest children were 8 years, but majority older) and/or related to English's inconsistent orthography which is harder to master compared to more transparent orthographies (see Caravolas, 2022 for review).

In contrast to the studies presented so far, Zuilkowski et al. (2019) found high correlations among different RC measures in a large sample of 5,389 Kenyan children (age range 4–14 years; mean age 7.53 and SD = 1.32), using English and Kiswahili, the most common languages in urban Kenyan settings. Three measures were compared (RC 60-second passage, RC 180-second passage, sentence comprehension) within and across the two languages. They interpreted the high correlations between the measures as support for the claim that the tasks were likely measuring the same underlying skill, and argued that one measure was sufficient in this context. They reasoned additional measures cost valuable time and resources but were unlikely to provide additional information. In order to reconcile this divergent finding with the previous research, it should be noted that the measures compared within Zuilkowski et al.'s study were a shorter and a longer version of the same task, which showed very high correlations (r=.87 in English; r=.75 in Kiswahili), while the correlations of the 60-s and 180-s passage with the sentence comprehension reading measure correlated less, although were still high (r=.78 and r=.74 in English; r=.69 and r=.59 in Kiswahili). Also worth noting is the distribution of the scores and internal reliability of the tasks. It might be the case that the higher correlations were influenced by poor performance on the tasks, for example the mean percentage score for the Kiswahili RC 60-second passage was 30.90 (SD=29.49), and 14.98 (SD=17.74) for the RC 180-second passage, with Cronbach alpha reported for those tasks as .44 and .64 respectively. Therefore, the correlations in Zuilkowski et al. (2019) might not necessarily point to the fact that their measures assessed the construct of RC more consistently than other studies which reported larger discrepancies between RC measures, such as Keenan et al. (2008).

The Present Study

Our study compares RC measures developed for Slovak-speaking children by evaluating the processes the measures rely on (following the SVR model's two main components: decoding and OC). We set out to evaluate if the role of decoding skills and OC skills for RC performance varies with the reading measure used and the developmental stage. We compare comprehension measures at the early stages of reading instruction (end of school Year 1 and Year 2) to provide insight into the important window of early learning of reading. In contrast, many previous studies looked at middle or final years of primary school and/or across secondary school years (e.g., Calet et al., 2020; Clemens et al., 2021; Colenbrander et al., 2017). Because our sample included children followed longitudinally, we could determine not only whether there were differences between what the tests relied on to predict RC (decoding vs. OC) but also whether this differed with developmental level.

Research questions:

- What are the roles of *decoding* and *oral language comprehension* in different measures of RC (*Slony, Deixy, Plagat, Ostrov* and *Jurko*)?
- Do the roles of the *decoding* and *oral language comprehension* components change with age (Year 1 vs. Year 2)?

In keeping with the SVR and previous research, we expect the importance of role of decoding in RC to decline as decoding skill improves with age and experience (e.g., Fuchs et al., 2012; García & Cain, 2014). Based on findings of decoding skill accounting for more variance in RC of younger children (e.g., Catts et al., 2005; Keenan et al., 2008), we predict that the decoding component will be more prominent in explaining variance in Year 1 compared to Year 2 in our longitudinal study. Keenan et al. (2008) reported the age differences in a cross-sectional study of English-speaking children aged 8-18 years. However, it is not clear if a similar pattern would be found in earlier stages of reading development (6-8 years) or with the small difference in progression (between Year 1 and Year 2) in our study. The current study is conducted in Slovak, a transparent orthography where decoding is mastered relatively quickly (Caravolas et al., 2019), and it is possible that clear developmental differences might not be detected. Following Florit and Cain (2011) and their finding on transparent orthographies and the years of schooling, we expect to see a strong early role for OC.

Keenan et al. (2008) found differences were accounted for by decoding based on age but also on the comprehension measure used. By using a longitudinal design and a range of comprehension measures across two time points, we are able to evaluate the hypothesis that differences between comprehension tests might be even larger in younger children as the variance explained by decoding and OC changes as reading competence progresses. We predict that decoding and oral language skills will play a different role, depending on the RC measure used. For Slovak, Caravolas et al. (2019) reported that RC relies on decoding. However, their study only used one measure of a multiple-choice task which relies on the decoding component more. By including a variety of measures, we are the first study to be able to address the question of whether the reliance of decoding was due to the measure used. Based on the RC studies with English that employ multiple-choice questions or cloze questions, greater variability in RC is explained by decoding measures, while oral language skills such as listening comprehension play a stronger role with measures that use open-ended questions (Keenan et al., 2008). For our multiple-choice test *Slony*, we therefore predicted a stronger role for decoding, particularly in Year 1 when the children were at the earlier stages of literacy acquisition. On the other hand, we expected more variance to be explained by OC for our open-ended question tasks *Plagat* and *Jurko*. Our other RC tasks (*Ostrov*, *Deixy*) have answer formats that had not been studied with respect to the SVR components and the predictions were therefore less clear. We expected a more balanced contribution of both decoding and oral language, with the role of oral language contribution strengthening at Year 2.

In summary, our study builds on previous research and also extends it in important ways. It is the first study to longitudinally explore if the key components of the SVR that explain variance on RC measures vary with age and a measure used, with a focus on the early stages of literacy instruction. The investigation extensively compares reading measures in a language with a transparent orthography and it is the first study to consider education and clinical implications of RC measures developed for the Slovak language. Given poor comprehenders are often difficult to detect and the lack of suitable tools in Slovakia, it is important to evaluate tests that have clinical potential and could serve as assessments in the diagnostic process in Slovakia.

Method

Participants

The sample was comprised of 200 children (105 girls, 95 boys) who all spoke Slovak as their first language. Nine children had been exposed to a second language. The children were recruited from six state primary schools in Slovakia (four schools in Bratislava N = 152, two schools from rural areas in western Slovakia N = 48 children). Children were tested at the end of the first year of primary school (mean age = 88.7; SD = 4.39, range = 80 - 102 months), which is the first year of literacy instruction, and then retested approximately 12 months later at the end of Year 2 (mean age = 100.34; SD = 4.28, range = 92 - 113 months).

Materials

In the absence of standardized tests, many of the language tasks used were designed and/or adapted for the purpose of this study. Reliability for the measures (Guttman's λ_6 and Cronbach's alpha) consisting of multiple items is reported in Table 1 together with other descriptive statistics.

Nonword reading test. To assess decoding efficiency, a pseudoword reading test was adapted. The nonword reading test created for this study was built on similar principles as Caravolas (2018), but we also considered phonology in addition to the orthographic form, ensuring that items were legal combinations both orthographically and phonotactically. In addition, items in our test were ordered randomly rather than increasing in length. The test consisted of 69 words which varied in length (24 one-syllabic, 25 two-syllabic, 11 three-syllabic, 7 four-syllable, and 2 five-syllable) and syllabic structure (CV, VC, CVC, CVCC, CCVC, CCCVC structure, where C = consonant and V = vowel). Children had one minute to read as many nonwords as they could, and one point was awarded for each nonword read correctly.

Test for Reception of Grammar Version 2 (TROG-2; Bishop, 2003). Experimental version adapted for Slovak and previously used in Polišenská et al. (2018). The test consists of 20 blocks, with each block testing a specific grammatical structure. There are four items within each block and 80 items in total. The lexical items used in the test were early acquired items to avoid confounding lexical and grammatical comprehension. In this picture-matching task, a child is presented with four pictures and one sentence and asked to choose the picture that matches the meaning of the sentence. Responses were scored as correct if the child chose the matching picture. The total number of items answered correctly was used as the score.

Reading comprehension tests

We included five tasks designed to measure RC skills. Two of the tasks were repeated at the end of Year 1 and Year 2 (*Slony, Deixy*), one was administered only at the end of Year 1 (*Plagat*), and two were designed for more advanced readers at the end of Year 2 (*Ostrov*, *Jurko*). As we were limited by the length of the testing session, one task from Year 1 was dropped at Year 2. Therefore, at the earlier stages of reading instruction when children's capacity for testing was more limited (e.g., by attention span), we asked children to read three texts (*Slony, Deixy, Plagat*), and when children were older we asked them to read four texts (*Slony*, *Deixy*, *Ostrov* and *Jurko*). All the RC comprehension measures were designed by the authors of the current study and followed the principles of measures commonly used for assessing RC. Cain (2010) discusses different formats and pros and cons of the variety of the measures. Our aim was to include a range or different types of tasks (e.g., multiple-choice, open questions, true/false statements) and different types of texts in a variety of writing styles (e.g., fiction as well as science texts, narrative as well as expository prose) as is the case for example in the standardised Gates–MacGinitie Reading tests or comparative assessment frameworks such as *Progress in International Reading Literacy Study* (PIRLS; Mullis & Martin, 2019). For example, PIRLS requires children to work with different types of visual information such as graphs, photos, and maps. Likewise, our study aimed to reflect this aspect of literacy in our tasks, for example our tasks *Plagat* and *Ostrov* required children to work with visuospatial information. In summary, we designed tasks that varied the types of texts as well as formats of the answers.

Slony (translation: 'elephants'). This is an expository text in which children read a short passage about elephants' teeth. The passage consisted of 20 sentences, divided into three parts, and was 180 words in total (mean length of a sentence was 9.00 words, SD = 3.60). The first and second parts were each followed by three questions and the third part by two questions. Each question came with four possible answers and the child had to select a correct answer. There were 32 items in total, each scored as correct/incorrect.

Deixy (translation: 'deixis'). The design of this task was motivated theoretically as well as clinically. Yuill and Oakhill (1988) showed that less skilled monolingual English-speaking 8-year-old comprehenders were considerably poorer than skilled comprehenders when assessing anaphors in reading. Several practitioners also observed and reported that struggling with anaphors comprehension leads to reading difficulties in English (e.g., Letchford & Rasinski, 2021, Mesmer & Rose-McCully, 2018) and our own observations highlighted the difficulties in Slovak. Therefore, we designed a task which targeted deictic terms covering people, objects, and places. Children were given a short narrative text of sentence pairs which were 85 words in total (mean sentence length was 5.31 words, SD = 1.89), with 8 paragraphs in total. Nouns from the first sentence of the pair were replaced with pronouns in the second sentence and children were asked to identify the pronoun referents. Children provided the answers in writing and were awarded a point for each correctly matched deixis and person/object/place. The maximum score was 16.

Plagat (translation: 'poster'). Children were given a poster with information about a meeting (place, date, time, rules, and the program). The text had 15 sentences, 87 words in total (mean sentence length was 5.80 words, SD = 3.00). Children were asked to answer seven questions and provide a rationale for their answers in writing. The task had a time limit of 45 minutes. A point was available for each correct answer and another for a correct justification; the maximum score was 14 points.

Ostrov (translation: 'island'). This task combines working with text and visual-spatial skills like those needed for map reading. Children were given a map of a fictional island with information about its inhabitants and characteristics that was presented in a booklet accompanying the map. Some text was in the map (19 sentences, 125 words, mean sentence length = 5.95 words, SD = 3.43) and some in the booklet (43 sentences, 323 words, mean sentence length = 7.51 words, SD = 3.69). The task had four parts. Children provided answers in writing or by drawing on a scoring sheet. In part 1, they were asked to select items from a list to take with them to the fictional island. The answers followed from the text rather than experience or world knowledge. For example, the booklet stated that they should take an old, unfinished sandwich because that is what one of the inhabitants of the island likes to eat. The maximum score was 10 points. In part 2, children were asked to draw a portrait of the queen from the island and the maximum score was 24 points. Children had to follow the text rather than their experience to correctly follow the instructions (for example, the instructions stated that the queen had one eye, one tentacle, a headband with a name written backwards, etc.). In part 3, children were asked two open questions. In part 4, children were asked to decide if four different written sentences were correct or not. Although the content of the four parts was related, how children provided the answers (underlining, writing, drawing) differed across the parts and therefore answers were analysed separately rather than as a composite score.

Jurko (Jurko is a proper name in Slovak). Children were given a narrative text adapted from the story *Henry and the Red Stripes* by Eileen Christelow, translated into Slovak. This task was inspired by a listening comprehension task designed by Nicolopoulou and Ünlütabak (2017). The original task was administered orally to 4-7-year-old English-speaking children during a book-reading session, with the experimenter asking 11 questions, including questions about false belief. Our task, *Jurko*, was translated by the authors of the present study from English to Slovak as the story had not previously been published in Slovak. One of the co-authors of this article is also a published children's book author and all are familiar with children's texts, ensuring the text was age-appropriate. Further adaptations included the delivery mode and the number of questions asked: in *Jurko* children read the text themselves and were asked 13 questions, with some questions about the text (e.g., who the story was about), while other questions required children to adopt perspectives of the characters and consider what they might be thinking (e.g., why is the fox afraid to eat the rabbit? – the text implies he looks different and therefore the fox thinks he might be ill). The text was comprised of 54 sentences, 416 words in total (mean sentence length was 7.43 words, SD = 3.41). Each correct answer was awarded 1 point; the maximum score was 13.

Procedure

The measures were administered at the children's school. The RC measures were presented to a whole class by speech and language therapists or speech and language therapy students who received training and supervision. Children were asked to read the RC tasks silently. The nonword reading measure and TROG were administered individually. All testing sessions took place within one week at the end of the school Year 1 and 2. The group testing was always scheduled as the first activity of the school day, followed by an individual testing session, with each session lasting maximum of 45 minutes. The tests were administered in the following order in Year 1: *Slony* (session 1), *Deixy* (session 2), and *Plagat* (session 3). A similar procedure was repeated for Year 2: *Ostrov* (session 1), *Jurko* (session 2), *Deixy* (session 3), *Slony* (session 4). This was followed by a nonword reading measure and TROG, which were both administered individually after group assessment.

Statistical analysis

The test battery consisted of two types of data. The nonword reading task assessed the number of nonwords that were read correctly per minute. All other measures were scored as correct or incorrect. Because of the dichotomous scoring, reliability estimates – Guttman's λ_6 and Cronbach's alpha – were based on the tetrachoric correlation matrices (Zijlmans et al., 2018). The descriptive statistics and reliability estimates were calculated using the functions of the package psych (Revelle, 2020) working in the statistical environment R (R Core Team, 2021). Some data was missing for the nonword reading test (N = 2 in Year 1 and N = 16 in Year 2) due to technical problems with the recording device, so the number of observations varied across the time points for this task (see Table 1). In statistical analyses which involved the nonword reading test as a predictor (i.e., decoding predictor), missing data were excluded listwise.

Following the research design, there were two independent variables (the number of correctly read nonwords per minute as a measure of the decoding; a score in the TROG test as a measure of OC). The dependent variables consisted of a set of test scores from the RC measures. Two of the RC measures were administered at both time points (*Slony* and *Deixy*), whereas other RC tests (*Plagat, Ostrov, Jurko*) were appropriate for testing only at a single time point, therefore the progress between the years in RC cannot be assessed directly on all measures. Because the dependent variables were mutually correlated, a multivariate linear model with analytic *p*-values was used. The appropriate function mvlm() is available in the MVLM package (McArtor, 2017) working in the statistical environment R (R Core Team, 2021). The dependent variables are represented by matrix Y, whereas the predictors x_1 to x_n appear in the formula using the syntax usual for linear models:

```
model <- mvlm(Y ~ Decoding + Oral.comprehension, data = data)
```

In our analysis, matrix Year 1 has three columns (measures of RC at Year 1 – *Slony, Deixy, Plagat*) and matrix Year 2 has five columns (measures of RC at Year 2 – *Slony, Deixy, Ostrov1, Ostrov2, Jurko*). A direct comparison of the change in the impact of the predictors on RC rates can only be made for the *Slony* and *Deixy* tests because only those tests were administered to all children in both years of testing. To assess the impact of the predictors in a repeated measures design, a linear mixed effects model was used in which we defined the individual child identifier as a random effect. Models for both outcome variables were fitted using the lmer() function available in the lme4 package (Bates et al., 2015). The general formula for both outcome variables is the same:

```
model <- lmer(Y ~ (Decoding + Oral.comprehension)*Time + (1|ID),
data = data)
```

Variable Y stands for the score in the tests *Slony* and *Deixy* respectively. We further assessed the developmental aspect of our research question by considering the relative effect sizes of both predictors, i.e., decoding and OC, across the RC measures. As an effect size estimate, the MVLM package offers Pseudo-R^2 coefficients (omnibus and conditional).

Results

Descriptive statistics and test reliability estimates are shown in Table 1. Both reliability estimates (Guttman's λ_6 and standardized Cronbach α) appeared to be excellent (> .9) or very good (> .8). Two subtests showed reliability below .8 and were excluded from further analyses (*Ostrov3* and *Ostrov4*). To assess the patterns of association between the various measures, Pearson correlations were computed as shown in Table 2.

Table 1

Descriptive statistics and reliabilities for all measures

	N	Mean	SD	Median	Min	Max	Skewness	Kurtosis	SE	Std. Alpha	Guttman's Lambda $_6$
Decoding Y1	198	18.23	8.95	17	4	56	1.21	2.16	0.64		
Decoding Y2	184	34.32	12.28	34	8	68	0.07	-0.50	0.91		
Oral comprehension Y1	200	62.37	6.28	63	43	74	-0.71	0.04	0.44	0.94	0.96
Oral comprehension Y2	200	66.71	4.92	68	52	75	-0.65	0.08	0.35	0.95	0.96
Slony Y1	200	19.02	4.76	19	0	28	-0.83	1.14	0.34	0.83	0.88
Slony Y2	200	23.40	3.52	24	7	31	-0.97	2.15	0.25	0.83	0.84
Deixy Y1	200	4.47	4.77	3	0	16	1.14	0.16	0.34	0.97	0.97
Deixy Y2	200	8.78	6.21	9.5	0	16	-0.18	-1.56	0.44	0.99	0.99
Plagat Y1	200	3.81	2.83	4	0	11	0.49	-0.44	0.20	0.87	0.94
Ostrov1 Y2	200	8.01	2.64	9	0	10	-1.67	2.11	0.19	0.95	0.97
Ostrov2 Y2	200	15.79	7.84	19	0	24	-0.80	-0.71	0.55	0.98	1.00
Ostrov3 Y2	200	1.42	1.17	1	0	4	0.21	-1.11	0.08	0.71	0.71
Ostrov4 Y2	200	2.02	1.29	2	0	5	0.17	-0.58	0.09	0.59	0.65
Jurko Y2	200	10.97	2.33	12	0	13	-2.09	5.80	0.16	0.90	0.92

Table 2

Zero-order Pearson correlations between variables

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Decoding Y1	—										
2. Decoding Y2	.79**	—									
3. Slony Y1	.42**	.35**	—								
4. Slony Y2	.27**	.29**	.31**	—							
5. TROG Y1	.19**	.10	.28**	.50**	—						
6. TROG Y2	.19**	.11	.21**	.43**	.79**	—					
7. Deixy Y1	.21**	.17*	.24**	.31**	.23**	.27**	—				
8. Deixy Y2	.28**	.20**	.27**	.45**	.54**	.52**	.46**	—			
9. Plagat Y1	.25**	.36**	.26**	.27**	.34**	.35**	.36**	.36**	—		
10. Ostrov1 Y2	.11	.21**	.13	.28**	.31**	.30**	.11	.22**	.16*	—	
11. Ostrov2 Y2	.24**	.37**	.29**	.36**	.40**	.39**	.19**	.35**	.36**	.33**	—
12. Jurko Y2	.20**	.33**	.16*	.50**	.39**	.39**	.21**	.38**	.27**	.30**	.37**

*$p < .05$. **$p < .01$.

RQ1: What is the role of *decoding* and *oral language comprehension* for different measures of RC?

In order to estimate the extent decoding and OC predict RC, a multivariate multiple regression model was fitted. The predictors (the number of nonwords read per minute and the OC test score TROG) were associated with multiple measures of RC. As can be seen in Table 2, the measures of RC for each year of study were correlated. Therefore, it would not be appropriate to individually repeat the regression model with each outcome variable because of the inflation of Type I error probability. Table 3 presents the output of the multivariate multiple regression model at the end of Year 1.

Table 3

Multivariate Linear Model output at the end of Year 1

	Statistic	df	Pseudo-R^2	*p*
Omnibus Effect	16.849	2	.147	< .001
Intercept	4.716	1		.007
Decoding	17.534	1	.077	< .001
Oral comprehension	10.370	1	.045	< .001

Both predictors (decoding and OC) are statistically significant, with similar predictive value at the end of Year 1. The effect size of both predictors together is small, with the predictors explaining only about 15 % of the variance. The conditional effects of individual predictors (while fixing the other predictor) are also low: Pseudo-R^2= .045 for OC and Pseudo-R^2 = .077 for decoding. Similar interpretations can be inferred from the estimates of the regression coefficients for individual predictors and corresponding dependent variables, as shown in Table 4. The first row of Table 4 contains the overall variance explained by the model for each dependent variable. The remaining rows list the variance of each outcome as explained by the conditional effect of each predictor. Thus, these results represent the effect size of each predictor on each dependent variable. From the results reported in Table 3, it can be inferred that there was a higher effect of decoding on RC at the end of Year 1. However, taking into account the effects of both predictors on the three different RC measures presented in Table 4, it appears that this stronger role of decoding might be driven by the RC test *Slony*. Different results emerged at the end of Year 2, see Table 5 and Table 6.

Table 4

Variance explained by the model (R^2) for individual RC measures at the end of Year 1

	Slony	*Deixy*	*Plagat*
Omnibus Effect	.213	.083	.145
Decoding	.139	.028	.037
Oral comprehension	.038	.039	.082

Table 5

Multivariate Linear Model output at the end of Year 2

	Statistic	df	Pseudo-R^2	*p*
Omnibus Effect	31.300	2	.258	< .001
Intercept	22.230	1		< .001
Decoding	16.300	1	.067	< .001
Oral comprehension	38.220	1	.158	< .001

Table 6

Variance explained by the model (R^2) for individual RC measures at the end of Year 2

	Slony	*Deixy*	*Ostrov1*	*Ostrov2*	*Jurko*
Omnibus Effect	.245	.296	.114	.330	.251
Decoding	.058	.019	.031	.083	.082
Oral comprehension	.161	.257	.071	.213	.142

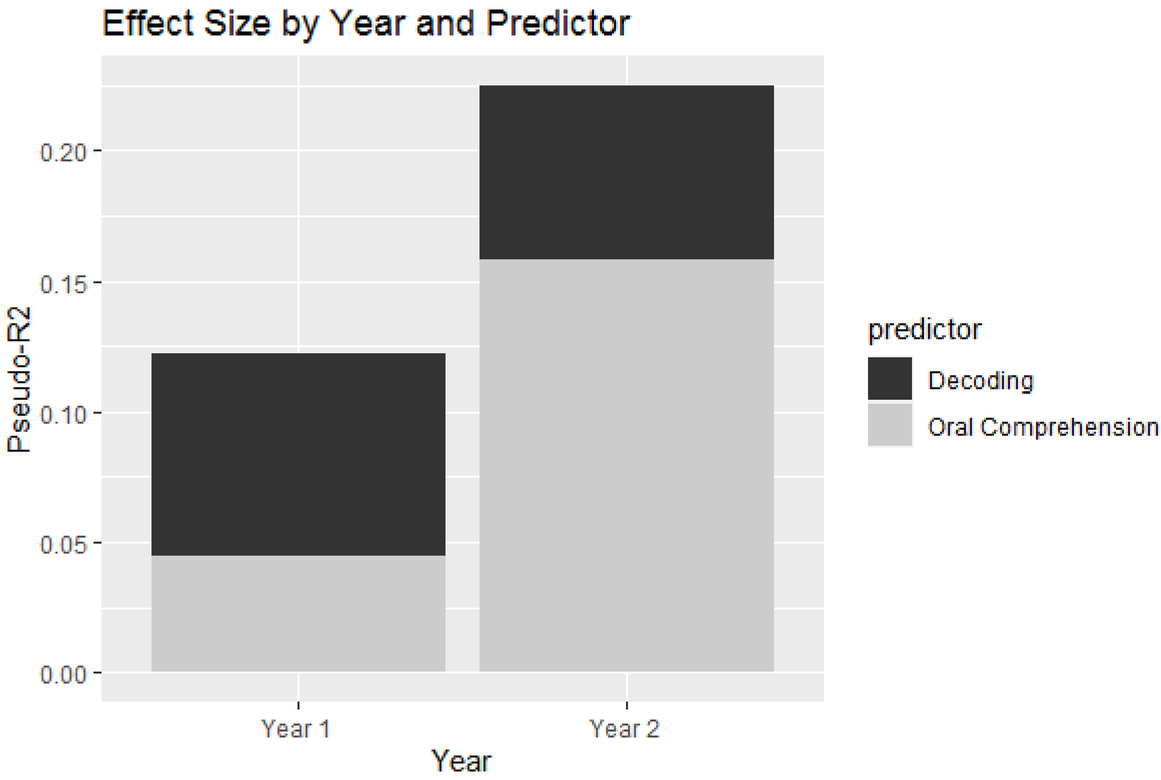
At the end of Year 2, the overall effect size was larger, with predictors explaining almost 26 % of the variance. Changes were observed regarding the conditional effects of individual predictors: The importance of OC increased to conditional Pseudo-R^2 = .158. On the other hand, the effect of the decoding ability remained similar to what was found at the end of Year 1, with conditional Pseudo-R^2 = .067. The estimates of the effect sizes for each predictor and each dependent variable offer a similar picture, see Table 6. The influence of OC at the end of Year 2 appears to be noticeably higher for each RC measure tested (*Slony*, *Deixy*, *Ostrov1*, *Ostrov2*, *Jurko*).

RQ2: Does the role of the *decoding* and *oral language comprehension* components change with age (Year 1 vs. Year 2)?

As stated above, the developmental change can be inferred from the estimates of conditional effect sizes for both predictors. It appears that the impact of OC increased at the end of Year 2. The conditional effect size for the predictor OC at Year 2 was higher compared to the corresponding effect size at Year 1 (Year 1 Pseudo-R^2 = .045, whereas Year 2 Pseudo-R^2 = .158). On the other hand, the effect of decoding was more constant across time: the conditional effect size for the nonword reading task at Year 1 was similar to the corresponding effect size at Year 2 (Year 1 Pseudo-R^2 = .077, whereas Year 2 Pseudo-R^2 = .067), as can be seen from Figure 1. The estimates of variance explained by the model for each dependent variable provide a more detailed view: the scores in tests *Slony* and *Deixy* show a lowering of the R^2 for the measure of decoding ability from Year 1 to Year 2 (from .139 to .058 for *Slony*; from .028 to .019 for *Deixy*). On the other hand, the R^2 for OC increases for both RC scores (from .038 to .161 for *Slony*; from .039 to .257 for *Deixy*). Other RC tests are not directly comparable because they were not repeated at both time points.

Figure 1.

The proportion of variance explained by predictors at the end of Year 1 and Year 2 for two types of RC measures administered at both time points (Slony and Deixy).



These interpretations based on the effect sizes are supported by the results of the linear mixed-effect models used for the longitudinal comparisons. The results are presented in Table 7 and Table 8, for the tests *Slony* and *Deixy* respectively. The linear mixed-effect model with the outcome variable *Slony* shows a significant effect of *Decoding* and significant *Decoding*Time* and *Oral comprehension*Time* interactions. To further aid the interpretation, the two significant interactions were plotted (see Figure 2a for the predictor Decoding and Figure 2b for the predictor Oral comprehension). As can be seen from the figures, Decoding explained more variance in the RC measure *Slony* at Year 1 compared to Year 2, while OC showed the opposite pattern, explaining less variance at Year 1 compared to Year 2.

Table 7

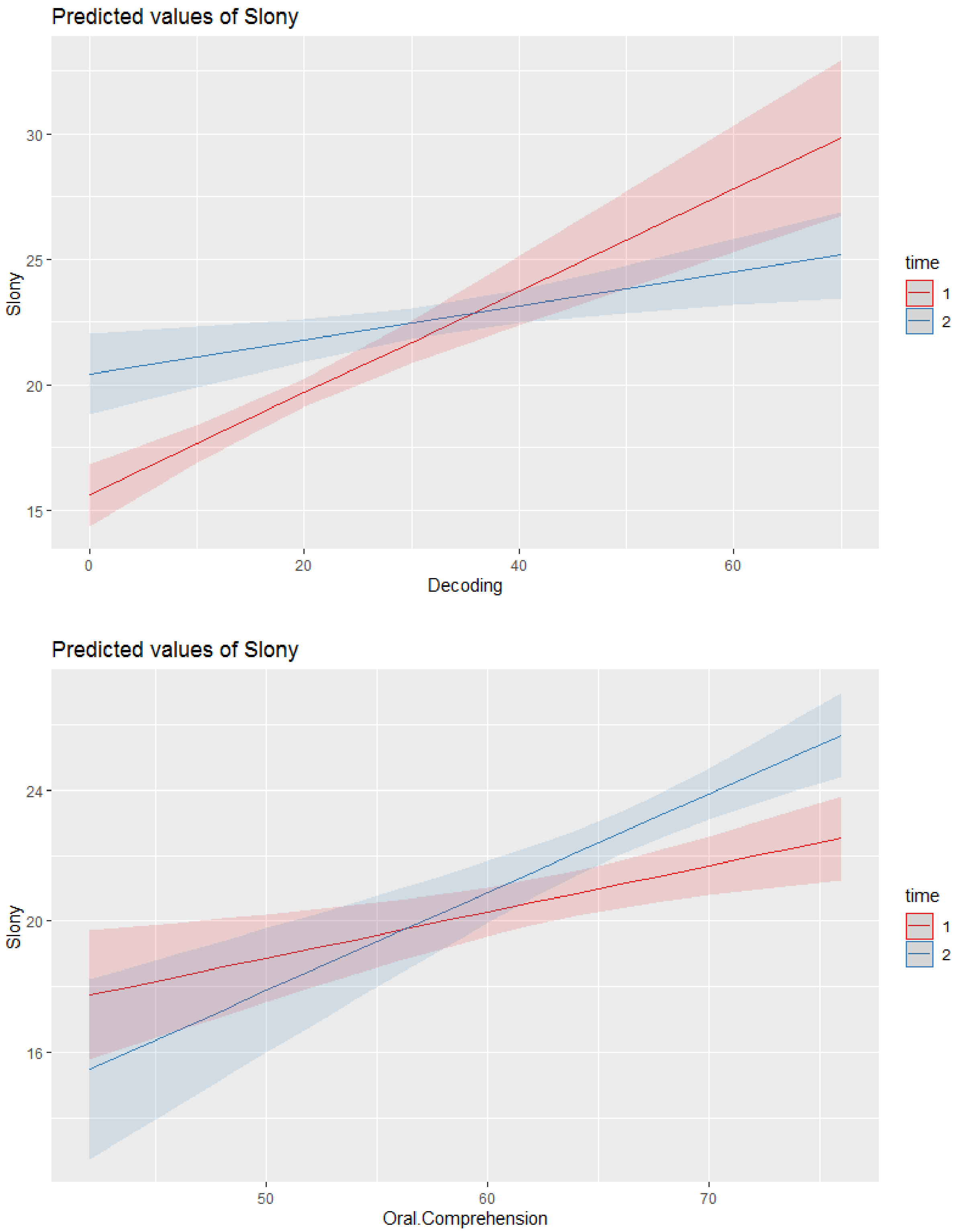
Output for the linear mixed-effect model for the outcome variable Slony

Effect	Estimate	95% CI		*p*
		LL	UL	
Fixed effects				
Intercept	11.97	-.19	24.12	.054
Decoding	.34	.22	.46	<.001
Oral comprehension	-.02	-.21	.18	.856
Time	-5.39	-13.94	3.16	.216
Decoding * Time	-.14	-.21	-.07	<.001
Oral comprehension * Time	.16	.03	.29	.019
Random Effects				
σ^2	11.93			
$\tau_{00\ ID}$	2.17			
ICC	.15			
Marginal R^2 / Conditional R^2	.379 / .475			

Note. Number of participants = 200, number of observations = 382. CI = confidence interval; *LL* = lower limit; *UL* = upper limit; ICC = Intraclass correlational coefficient

Figure 2a and 2b.

Interaction plots for two predictors and the score in the test Slony taken at two time points in a repeated-measures design.



The results for the RC measure *Deixy* are different to the results for the RC measure *Slony*. The linear mixed-effect model with the outcome variable *Deixy* showed a significant effect of *Oral comprehension*, *Time* and a significant *Oral comprehension*Time* interaction. To further aid the interpretation, the interactions were plotted (see Figure 3a for the predictor Decoding and Figure 3b for the predictor OC). Whereas the role of decoding did not change at Year 1 compared to Year 2, there was a significantly higher contribution of OC to the outcome variable at Year 2 compared to Year 1 (Table 8).

Table 8

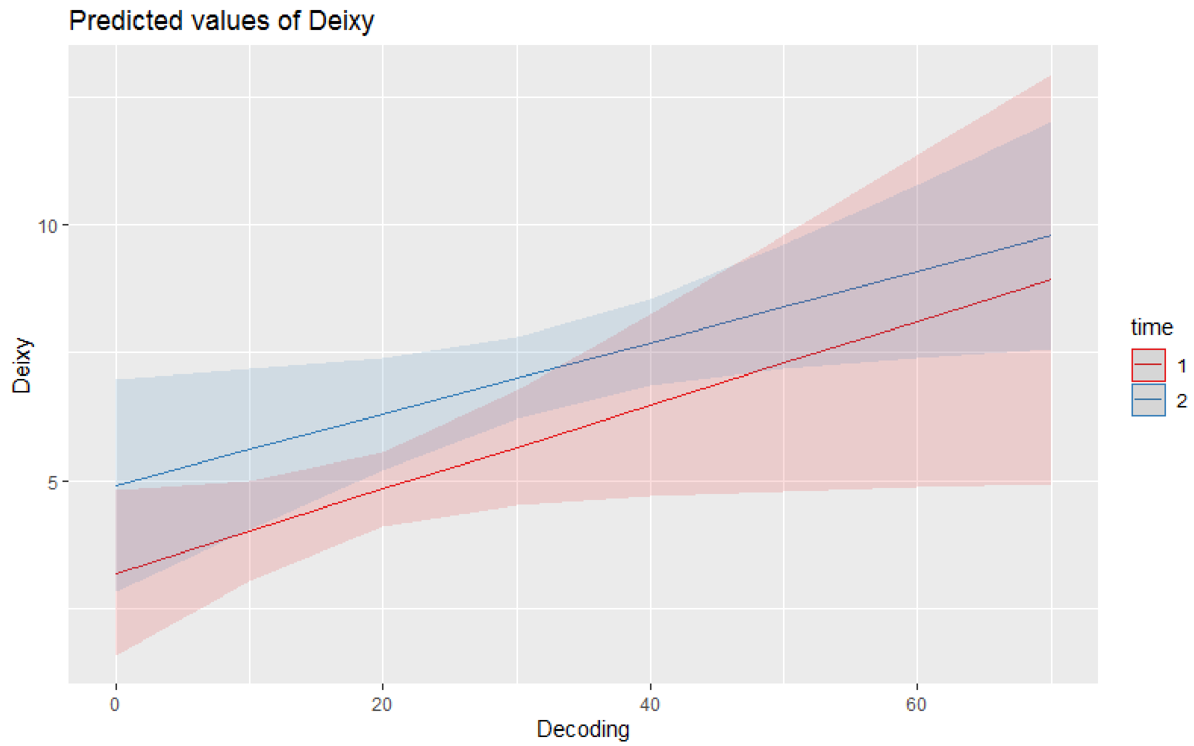
Output for the linear mixed-effect model for the outcome variable Deixy

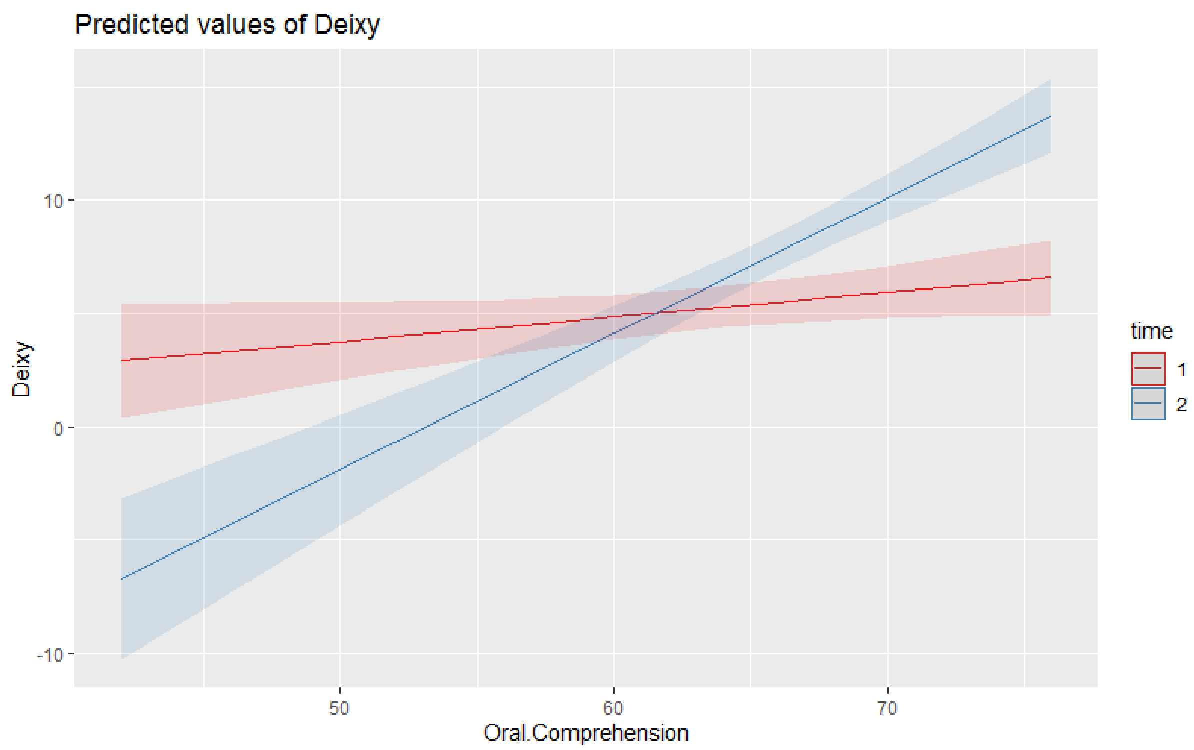
Effect	Estimates	95% CI		*p*
		LL	UL	
Fixed effects				
Intercept	26.18	11.67	40.69	<.001
Decoding	.09	-.06	.24	.217
Oral comprehension	-.38	-.62	-.15	.001
Time	-29.99	-40.18	-19.80	<.001
Decoding * Time	-.01	-.10	.07	.773
Oral comprehension * Time	.49	.34	.65	<.001
Random Effects				
σ^2	16.16			
$\tau_{00\ ID}$	8.39			
ICC	.34			
Marginal R^2 / Conditional R^2	.283 / .528			

Note. Number of participants = 200, number of observations = 382. CI = confidence interval; *LL* = lower limit; *UL* = upper limit; ICC = Intraclass correlational coefficient

Figure 3a and 3b.

Interaction plots for two predictors and the score in the test Deixy taken at two time points in a repeated-measures design.





Discussion

This study set out to evaluate a number of RC measures in Slovak and tease apart the contribution of the main components of RC skills as presented in the SVR - decoding and oral language comprehension (as measured by grammatical comprehension on a sentence level in this study). The study investigated if differences between RC tests depend on developmental level and/or type of a measure. This is the first longitudinal study which has evaluated differential effects of decoding and OC contribution in diverse RC measures.

Understanding which components are assessed by a specific RC measure and how different measures maybe comparable to each other has important implications for educational and clinical practice in Slovakia as well as future research and theoretical understanding of RC. The time required to administer tests is a consideration, and researchers and clinicians often wonder if one test instead of several would be sufficient while ensuring that the test is providing results which can be interpreted appropriately.

Our study replicated findings from studies with the English language that showed weak to moderate correlations between different reading measures. Looking at the relationships between our three reading measures (*Slony, Deixy, Plagat*) administered at Year 1, the correlations were also relatively low or moderate (between r = .24 and r = .36, see Table 2 for details). The strength of the relationship improved for some of the tests in Year 2 when five reading measures were administered (*Slony, Deixy, Plagat, Ostrov1, Ostrov2, Jurko*), with the correlations ranging from r = .22 to r. = 50. The figures from Year 2 come close to the correlations reported in Keenan et al. (2008) for English. Similar correlational coefficients were also reported by Calet et al. (2020) for Spanish on their comprehension measures (PROLEC-R and their other two tests r = . 30 –.33). Although the tests in our study were designed to measure the same construct (i.e., RC), it does not appear that one test can be favoured over the others, therefore based on the data available, it would not be advisable in the interest of time to only administer one test. Previous research has mainly highlighted the differences between RC measures in English speaking countries (US - Keenan et al., 2008, UK - Nation & Snowling, 1997, Australia - Colenbrander et al., 2017), but studies showing similar results from other languages have also started to appear (e.g., Spain - Calet et al., 2020, Kenya - Zuilkowksi et al., 2019, Italy - Florit et al., 2021). Our study contributed to this literature by presenting novel evidence from a group of Slavic languages which have been under-investigated in this respect.

Next, we turn to the question of unpicking which components of the SVR are being assessed by different RC tests and if this varies depending on the test used and the developmental level. Our predictors explained 15% of variance on RC scores at the end of Year 1 and 26% variance at the end of Year 2. While the role of decoding remained largely similar (Year 1 - variance explained 8%; Year 2 - variance explained 7%), the role of OC appeared stronger in Year 2 compared to Year 1 (16% compared to 5%). The findings from the cross-sectional analyses were mirrored by the results from the mixed-effect models, carried out with two measures (*Slony* and *Deixy*) that were administered to all participants at Year 1 and Year 2. Decoding had a significant effect on *Slony* and also interacted with time, with its role shown to be more prominent for Year 1 compared to Year 2. The role of oral comprehension also became more prominent with time, demonstrated by significant interactions for both *Slony* and *Deixy*.

This finding is in line with reports from languages with transparent orthographies, such as Italian, Portuguese or Finish. Tobia and Bonifacci (2015) who evaluated the components of the SVR in a large cross-sectional study showed that OC was a stronger predictor of RC for Italian-speaking children from the age of 6 years than both reading speed and accuracy. As Tobia and Bonifacci discussed, their findings were partially in contrast with previous findings which reported a primary role of decoding, as measured by reading speed rather than accuracy, in the first years of schooling for transparent orthographies (Florit & Cain, 2011). Similarly, Cadime et al. (2016) found that listening comprehension was a stronger predictor than word recognition for children both in grade 2 and grade 4 in European Portuguese which was described in their study as an intermediate-depth orthography. Torppa et al. (2016) reported a similar finding for Finish, showing that the effect for listening comprehension was stronger compared to the effect of reading fluency as measured by reading tasks in which a child was required to use “detailed fluent decoding” (p. 187). This finding adds to the cross-linguistic evidence and highlights the importance of diversity for understanding general processes children use during reading acquisition.

These studies have contributed to our understanding of changes with age and different orthographies. However, they were either cross-sectional or only included a single measure of RC and therefore could not address the question of change directly. Our study was longitudinal and included multiple measures of RC and therefore allowed the opportunity to examine if the role of the components changes with the developmental level. Overall, the role of decoding and OC was similar after 1 year of schooling; after two years of schooling the role of OC tripled, while decoding stayed largely the same. While there might be a limited benefit of decoding practice after certain levels are reached, there are benefits of oral language knowledge which continues to develop. Oral language will progress naturally as children even of reading age still need to acquire more complex syntactic structures, derivational morphology, enrich their vocabulary, improve pragmatic skills and inferencing. While exposure will ensure progress for most, it is important to provide extra support to those who might need it and target children who might become poor comprehenders. Taken together, our study adds to the growing body of evidence on transparent orthographies, showing that decoding already plays a weaker role by the second year of reading instruction compared to OC. While previous studies addressed this issue across different orthographies, they did not use different measures of RC which vary in dependence on OC and decoding. As shown for example by Keenan et al. (2008) with English, the nature of the RC measure can significantly affect such results.

Looking at the individual reading measures provides further insight into the role of the decoding and OC components of the SVR, and how their role depends on the measure used. At the end of Year 1, both *Deixy* and *Plagat* from the current study showed similar (4% vs. 3% for *Deixy*) or slightly stronger effect of OC vs. decoding (8% vs. 4% for *Plagat*). However, the measure *Slony* showed the opposite pattern, with decoding showing a stronger effect (14%) compared to OC (4%). This finding highlights the importance of the choice of reading measure. In line with our prediction, we found that the multiple-choice subtest *Slony* was more reliant on decoding, while *Plagat,* where children had to provide short answers, relied more on OC less than one year after formal reading instruction started.

The picture is different again after two years of reading instruction. Here, we found a stronger effect of OC for all five measures of RC *(Slony, Ostrov1, Ostrov2, Jurko, Deixy)*, with *Deixy* showing a particularly strong OC effect. A possible explanation for this strong link might be the choice of a proxy measure for OC - the TROG which is a measure of grammatical comprehension without reliance on contextual support. From all the reading measures used in the current study, the *Deixy* measure is probably the closest to TROG in its focus on morphosyntactic knowledge which would be necessary, but not sufficient, to successfully tackle the task. As seen from Table 6, the role of decoding remained largely the same for *Deixy* across Year 1 (3%) and Year 2 (2%), but the role of OC increased dramatically (4% and 26%). This suggests that it was not decoding skills that provided constraints on RC for *Deixy* at the end of Year 1; rather the success of this task appeared to be more influenced by oral language skills. In contrast, on the multiple-choice measure of RC *Slony*, the relations between OC and decoding changed over the course of development. In this case, it appears that decoding played a slightly stronger role than OC in Year 1, even in a transparent orthography like Slovak, but by the end of the second year of instruction when children are better at decoding, RC was more dependent on OC skills. This may be again due to relatively quick acquisition of reading in transparent orthographies (e.g., Torrpa et al., 2016).

In general, decoding seems to be the main determinant of RC in deep orthographies, particularly in early phases of reading development, but listening comprehension/oral language seems to be a stronger predictor of RC in more transparent orthographies across a wide range of reading development phases. The novel contribution of this study was showing that the previously described pattern additionally depends on the measure used. While this has been suggested in previous studies, it has not been evaluated in a longitudinal study with multiple reading measures in a transparent orthography. As predicted by the SVR, the role of RC depends on *decoding* x *oral language comprehension* and we have unpicked each of these contributions depending on the measure used and developmental level of the reader. Given decoding skills are acquired relatively quickly in Slovak and that they have a limited role once a certain required level is reached, our findings point to the importance of OC for RC skills. As oral language skills develop (see Marková & Mikulajová, 2011 for Slovak), this developmental progress in the oral language domain will inevitably have an impact of RC.

Limitations

We only used one measure as a proxy for OC and it focused on sentence-level comprehension. TROG is the only OC test available for Slovak-speaking children and validated on a larger sample with the age range needed for our study. It was also selected because the morphosyntactic comprehension skills that TROG evaluates are relevant to Slovak, a highly inflected language. In short, there were limited options in Slovakia, unlike English-speaking countries where numerous options are available. Other studies have used TROG to assess grammatical comprehension (e.g., Bishop & Adams, 1990; Caravolas et al., 2019; Language and Reading Research Consortium, & Chui, 2018). However, selecting any single measure as a proxy for the OC component in the Simple View of Reading may be limiting and affect our findings. Future studies could select other measures of OC, such as vocabulary, morphology, listening comprehension as was chosen e.g., by Clarke et al. (2010) or Cutting and Scarborough (2006), to see if similar findings can be obtained. More recently, some studies have opted for a combination of language measures (e.g., Lervåg et al., 2018) and this might also be a way towards capturing a more complex language comprehension profile. In addition, we have not included other independent measures previously shown to affect RC, such as verbal working memory (e.g., Lepola et al., 2012), and these could be incorporated into future studies.

Implications

Our findings have implications for educational and clinical settings. Firstly, our findings add to the body of research highlighting the importance of oral language support at primary school. While children are often well supported when it comes to speech difficulties in Slovakia or possibly other countries, difficulties with spoken oral language might be more hidden and receive less support and attention compared to disordered and or/delayed speech. The focus of reading instruction is often on decoding, and comprehension difficulties might remain unnoticed for some time. Fricke et al. (2013) and Keenan et al. (2014) stress that if poor comprehenders can be identified early, intervention is likely to be more effective. Even if teachers are aware, the link between OC and RC might not be explicitly included in their curriculum and therefore little focus and time becomes allocated to this area. There is a large body of research supporting the importance of oral language skills, such as narrative skills, inference making, grammar, and vocabulary for reading skills. This line of thinking is also supported by randomised controlled trials showing the positive effects of OC on RC (Clarke et al., 2010; Fricke et al., 2013).

Secondly, we evaluated a number of novel reading measures and asked if it was possible to use them interchangeably or focus on one rather than multiple measures in order to reduce the time needed for administration, scoring, and interpretation. Our findings are in line with reports for English and how the measures might be tapping different aspects of RC which also changed with a developmental level. Therefore, we do not recommend using a single measure, particularly if the measure is being used to arrive at a clinical diagnosis. For research, our findings highlighted that it is important to closely examine what measures (e.g., multiple-choice questions, open questions, similarity to the oral language measure) are used to evaluate RC. This will allow researchers to better understand and explain the differences and inconsistencies between studies. Given these findings, it would be useful to consider and label specific components which explain RC and not just operate with the umbrella term RC. For example, in our test *Deixy*, 26% of performance was explained by sentence-level OC skills in 7-8-year-old children. If a child had a low score on the subtest *Deixy*, the intervention could target sentence comprehension in oral language in the first place. Specifically, children could practice changing pronouns and full nouns/noun phrases which co-refer. This could be supported by using gestures and/or tokens which replace objects/nouns/pronouns or role playing, strategies known to be used for oral language therapy and support. In conclusion, if information about which components are predominantly assessed in the commonly used RC tests is available, this could make the interpretation of the results and the planning of support and intervention more targeted.

References

Avvisati, F., Echazarra, A., Givord, P., & Schwabe, M. (2019). *Progamme For International Student Assessment (PISA) Result from PISA 2018*. Retrieved June 11, 2022, from https://www.oecd.org/pisa/publications/PISA2018_CN_SVK.pdf

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bishop, D. V., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 31*(7), 1027–1050. https://doi.org/10.1111/j.1469-7610.1990.tb00844.x

Bishop, D. V. (2003). *Test for Reception of Grammar: Version 2: TROG-2.* Harcourt Assessment.

Cadime, I., Rodrigues, B., Santos, S., Viana, F. L., Chaves-Sousa, S., do Céu Cosme, M., & Ribeiro, I. (2017). The Role of word recognition, oral reading fluency and listening comprehension in the simple view of reading: a study in an intermediate depth orthography. *Reading and Writing, 30*(3), 591-611. https://doi.org/10.1007/s11145-016-9691-3

Cain, K. (2009). Children's reading comprehension difficulties: A consideration of the precursors and consequences. In C. Wood & V. Connelly (Eds.), *Contemporary perspectives on reading and spelling. new perspectives on learning and instruction* (pp. 59-75). Routledge. ISBN 9780415497169

Cain, K. (2010). *Reading Development and Difficulties*. BPS Blackwell Publishing Ltd.

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology, 76,* 683–696. https://doi.org/10.1348/000709905X67610

Calet, N., López-Reyes, R., & Jiménez-Fernández, G. (2020). Do reading comprehension assessment tests result in the same reading profile? A study of Spanish primary school children. *Journal of Research in Reading, 43*(1), 98-115. https://doi.org/10.1111/1467-9817.12292

Caravolas, M. (2017). Learning to Read Czech and Slovak. In L. Verhoeven & C. Perfetti (Eds.), *Learning to read across languages and writing systems* (pp. 347-370). Cambridge University Press. https://doi:10.1017/9781316155752.015

Caravolas, M. (2018). Growth of word and pseudoword reading efficiency in alphabetic orthographies: Impact of consistency. *Journal of Learning Disabilities*, *51*(5), 422-433. https://doi.org/10.1177%2F0022219417718197

Caravolas, M. (2022). Reading and Reading Disorders in Alphabetic Orthographies. In M. Snowling, C. Hulme, & K. Nation (Eds.), *The science of reading: A handbook* (pp. 327-353). John Wiley & Sons Ltd. https://doi.org/10.1002/9781119705116.ch15

Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and reading comprehension ability. *Scientific Studies of Reading, 23*(5), 386-402. https://doi.org/10.1080/10888438.2019.1580284

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior, S., Mikulajová, M., Seidlová-Málková, G., & Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science, 23*(6), 678–686. https://doi.org/10.1177/0956797611434536

Caravolas, M., Mikulajová, M., & Kucharská, A. (2019). Developmental Dyslexia in Czech-Slovakian. In L. Verhoeven, C. Perfetti, & K. Pugh (Eds.), *Developmental dyslexia across languages and writing systems* (pp. 96-117). Cambridge University Press.

Catts, H. W., Hogan, T. P., & Adolf, S. M. (2005). Developmental changes in reading and reading disabilities. In H.W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 38-51). Erlbaum.

Clarke, P. J., Snowling, M. J., Truelove, E. & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science, 21*, 1106–1116. https://doi.org/10.1177/0956797610375449.

Clemens, N. H., Hsiao, Y. Y., Lee, K., Martinez-Lincoln, A., Moore, C., Toste, J., & Simmons, L. (2021). The differential importance of component skills on reading comprehension test performance among struggling adolescent readers. *Journal of Learning Disabilities*, *54*(3), 155-169. https://doi.org/10.1177%2F0022219420932139

Colenbrander, D., Nickels, L., & Kohnen, S. (2017). Similar but different: differences in comprehension diagnosis on the Neale Analysis of Reading Ability and the York Assessment of Reading for Comprehension. *Journal of Research in Reading, 40*(4), 403-419. https://doi.org/10.1111/1467-9817.12075

Collins, A. A., Lindström, E. R., & Compton, D. L. (2018). Comparing students with and without reading difficulties on reading comprehension assessments: A meta-analysis. *Journal of Learning Disabilities, 51*(2), 108-123. https://doi.org/10.1177/0022219417704636

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*(3), 277-299. https://doi.org/10.1207/s1532799xssr1003_5

Elwer, S., Keenan, J.M., Olson, R.K., Byrne, B. & Samuelsson, S. (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology, 115(3)*, 497–516.https://doi.org/10.1016/j.jecp.2012.12.001

Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies?. *Educational Psychology Review, 23*(4), 553-576. https://doi.org/10.1007/s10648-011-9175-6

Florit, E., Roch, M., Dicataldo, R., & Levorato, M. C. (2020). The Simple View of Reading in Italian beginner readers: Converging evidence and open debates on the role of the main components. *Learning and Individual Differences, 93,* 1-10. https://doi:10.1177/0022219412442154

Francis, D.J., Snow, C.E., August, D., Carlson, C.D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading, 10*(3), 301–322. https://doi.org/10.1207/s1532 799xs sr1003_6

Fricke, S., Bowyer-Crane, C., Haley, A., Hulme, C., & Snowling, M. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry, 54*, 280–290. https://doi.org/10.1111/jcpp.12010

Fuchs, D., Compton, D., Fuchs, L., Bryant, J., Hamlett, C., & Lambert, W. (2012). First-grade cognitive abilities as long-term predictors of reading comprehension and disability status. *Journal of Learning Disabilities, 45*, 217–231. . https://doi:10.1177/0022219412442154

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research, 84*(1), 74–111. https://doi.org/10.3102/0034654313499616

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Hagley, F. (1987). *Suffolk Reading Scale*. NFER-Nelson.

Ho, C. S. H., Chow, B. W. Y., Wong, S. W. L., Waye, M. M., & Bishop, D. V. (2012). The genetic and environmental foundation of the simple view of reading in Chinese. *PloS ONE*, 7(10), e47872. https://doi.org/10.1371/journal.pone.0047872

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127–160. https://doi.org/10.1007/BF00401799

Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities, 47*(2), 125-135. https://doi.org/10.1177/0022219412439326

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and OC. *Scientific Studies of Reading, 12*(3), 281-300. https://doi.org/10.1080/10888430802132279

Language and Reading Research Consortium, & Chui, Y. D. (2018). The simple view of reading across development: Prediction of grade 3 reading comprehension from prekindergarten skills. *Remedial and Special Education, 39*(5), 289–303. https://doi.org/10.1177/0741932518762055

Lepola, J., Lynch, J., Laakkonen, E., Silvén, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4–6-year-old children. *Reading Research Quarterly, 47*(3), 259-282. https://doi.org/10.1002/rrq.020

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development, 89*(5), 1821-1838. https://doi.org/10.1111/cdev.12861

Letchford, L., & Rasinski, T. (2021). Moving beyond decoding: Teaching pronoun resolution to develop reading comprehension. *The Reading Teacher, 75*(2), 233– 240. https://doi.org/10.1002/trtr.2038

Marková, J., Mikulajová, M. (2011). Oral sentence comprehension in Slovak-speaking, typically developing children, in children with a Specific Language Impairment, and in children with dyslexia. *Journal of Speech and Language Pathology, 1*(1), 27-47.

McArtor. D. B. (2017). *MVLM: Multivariate Linear Model with Analytic p-Values. R package version 0.1.4*. Retrieved from https://CRAN.R-project.org/package=MVLM

Memisevic, H., Dedic, A., Biscevic, I., Hadzic, S., Pasalic, A., & Malec, D. (2020). Identifying predictors of reading speed and reading comprehension in Bosnian. *Applied Neuropsychology: Child*, 1-10. https://doi.org/10.1080/21622965.2020.1815023

Mesmer, H. A., & Rose-McCully, M. M. (2018). A closer look at close reading: Three under-the-radar skills needed to comprehend sentences. *The Reading Teacher*, *71*(4), 451-461. https://doi.org/10.1002/trtr.1639

Mikulajová, M., Váryová, B., Vencelová, L., Caravolas, M., Škrabáková, G. (2012). *Čítanie, písanie a dyslexia s testami a normami*. Slovenská asociácia logopédov.

Mullis, I.V.S., & Martin, M.O. (2019). *PIRLS 2021 assessment frameworks*. Boston College.

Nation, K. & Snowling, M.J. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67*, 359–370. https://doi.org/10.1111/j.2044-8279.1997.tb01250.x.

Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties, 24*(1), 47-73. https://doi.org/10.1080/19404158.2019.1609272

Neale, M. D. (1989). *The Neale Analysis of Reading Ability-Revised*. NFER.

Neale, M. D. (1999). *The Neale Analysis of Reading Ability. (Third edn)*. ACER.

Nicolopoulou, A. & Ünlütabak, B. (2017). Narrativity and mindreading revisited: Children's understanding of theory of mind in a storybook and in standard false belief tasks. In N. Ketrez, A. C. Küntay, Ş. Özçalışkan, & A. Özyürek (Eds.), *Social environment and cognition in language development: Studies in honor of Ayhan Aksu-Koç* (pp. 151-166). Trends in Language Acquisition Research Series (TiLAR): John Benjamins.

Polišenská, K., Kapalková, S., & Novotková, M. (2018). Receptive language skills in Slovak-speaking children with intellectual disability: understanding words, sentences, and stories. *Journal of Speech, Language, and Hearing Research, 61*(7), 1731-1742. https://doi.org/10.1044/2018_JSLHR-L-17-0029

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Revelle, W. (2020). *psych: Procedures for Personality and Psychological Research*. Northwestern University. R package version 2.0.12. Retrieved from https://CRAN.R-project.org/package=psych

Rimrodt, S., Lightman, A., Roberts, L., Denckla, M. B., & Cutting, L. E. (2005, February). *Are all tests of reading comprehension the same?*[Poster presentation]. The annual meeting of the International Neuropsychological Society, St. Louis, MO.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E. & Hulme, C. (2012). *York Assessment of Reading for Comprehension: Passage Reading. (Australian edn).* GL Assessment.

Tobia, V., & Bonifacci, P. (2015). The simple view of reading in a transparent orthography: The stronger role of OC. *Reading and Writing, 28*, 939–957 https://doi.org/10.1111/j.2044-8279.1997.tb01250.x.

Torppa, M., Georgiou, G. K., Lerkkanen, M. K., Niemi, P., Poikkeus, A. M., & Nurmi, J. E. (2016). Examining the simple view of reading in a transparent orthography: A longitudinal study from kindergarten to grade 3. *Merrill-Palmer Quarterly, 62*(2), 179-206. https://doi.org/10.13110/merrpalmquar1982.62.2.0179

Váryová, B. (2012). Skúšky čítania pre mladší školský vek [Tests of reading for the younger school years]. In M. Mikulajová, B. Váryová, & L. Vencelová, M. Caravolas, & G. Škrabáková (Eds.), *Čítanie, písanie a dyslexia (s testami a normami) [Reading, writing and dyslexia (with tests and norms]* (pp.126–159). Slovenská asociácia logopédov.

Yuill, N., & Oakhill, J. (1988). Understanding of anaphoric relations in skilled and less skilled comprehenders. *British Journal of Psychology, 79*(2), 173-186. https://doi.org/10.1111/j.2044-8295.1988.tb02282.x

Zijlmans. E. A.. van der Ark. L. A.. Tijmstra. J.. & Sijtsma. K. (2018). Methods for estimating item-score reliability. *Applied Psychological Measurement, 42*(7), 553-570. https://doi.org/10.1177/0146621618758290

Zubáková, M., & Mikulajová, M. (2021). Strong Precursors of early literacy in Slovak language. *Psychológia a patopsychológia dieťaťa [Psychology and Psychopathology of a Child]*, 55(1), 3–25. https://doi:10.4149/ppd_202101

Zuilkowski, S. S., Piper, B., Kwayumba, D., & Dubeck, M. (2019). Examining options for reading comprehension assessment in international contexts. *Journal of Research in Reading, 42*(3-4), 583-599. https://doi.org/10.1111/1467-9817.12285