Titre: Measures of comprehension for Czech first- to fourth-grade pupils

Auteur(s): Anna Kucharská & Klára Špačková Charles University, Prague, Czech republic

Publication: Perspectives actuelles sur l'apprentissage de la lecture et de l'écriture/ Contributions about learning to read and write - Actes du Symposium international sur la litéracie à l'école/International Symposium for Educational Literacy (SILE/ISEL) 2015

pages: 141 - 167

Directeurs: Marie-France Morin, Denis Alamargot et Carolina Gonçalves.

Éditeur: Les Éditions de l'université de Sherbrooke, 2016.

ISBN: 978-2-7622-0355-4

URI: http://hdl.handle.net/11143/10233

DOI: https://doi.org/10.17118/11143/10233

Measures of comprehension for Czech first- to fourth-grade pupils[1]

Anna Kucharská & Klára Špačková, Charles University, Prague, Czech republic

Abstract: The present findings are drawn from the project Reading comprehension – Typical Development and its Risks aiming to map the developmental dynamics of the reading comprehension skills of Czech children. In this paper, we focus on the issue of comprehension measures. Because of the lack of such measures in the Czech Republic, the research team designed three new tools: an oral reading comprehension (Rabbits), a silent reading comprehension (Going on a trip), and a listening comprehension test (Little Star). All of these tests have an identical structure, similar content features and assess the literal and inferential comprehension of a story.

First, we report on the reliability and validity of these tests using data from a study involving 467 first- to fourth-graders. Second, we compare the comprehension scores (global, implicit, and explicit) and investigate the differences between grades in the patterns of these comprehension skills. We discuss the possibility of the application of these measures as diagnostic tools in education and research.

Keywords: comprehension tests, reading comprehension, listening comprehension, silent reading, oral reading, literal and inferential comprehension

Résumé : Ces résultats de recherche sont tirés du projet Compréhension de lecture- Développement typique et ses risques, dont le but principal est d'établir la dynamique des habiletés en compréhension de lecture chez les enfants tchèques. Dans cette publication, nous mettons l'emphase sur les problèmes des mesures de compréhension. Dû au manque de telles mesures en République Tchèque, l'équipe de recherche a conçu trois nouveaux outils: un test de lecture à haute voix (Rabbits), un test de lecture silencieuse (Going on a trip), et un test d'écoute (Little Star). Tous ces tests ont une structure identique et des caractéristiques similaires et évaluent la compréhension littérale et déductive d'une histoire.

Premièrement, nous évaluons la fiabilité et la validité de ces tests en prenant les données d'une étude impliquant 467 élèves de la première à la quatrième année. Ensuite, nous comparons les résultats de compréhension (généraux, implicites et explicites) et étudions les différences entre les

---

classes suivant les modèles de ces habiletés de compréhension. Nous discutons de la possibilité d'appliquer ces mesures comme outils de diagnostic en éducation et recherche.

Mots-clés : test de compréhension, compréhension de lecture, compréhension à l'écoute, lecture silencieuse, lecture à haute voix, compréhension littérale et déductive.

INTRODUCTION

It is well accepted today that reading ability cannot be treated as a unitary construct (Scarborough, 2005). Not only for research purposes but also for educational reasons, we have to distinguish between word and text-level reading skills. Word-level reading skills develop as the child learns how to decode print and relate to the mechanism of reading. In alphabetic languages they involve knowledge about phoneme-grapheme correspondences but also orthographic knowledge and word-specific semantic reading skills (Plaut, 2005). Text-level reading skills go beyond reading single words and refer to text comprehension and writing. Though word-level reading skills are essential for understanding the meaning across sentences of the connected text, other skills, such as inference, comprehension monitoring, prior knowledge, and knowledge about text structure, are also required (e.g. Cain, Oakhill, & Bryant, 2004; van den Broek & Espin, 2012).

To estimate word-level and text-level reading skills, different indicators are commonly used. Though it may seem easy to define an adequate measure in the case of word-level reading skills (decoding and sight word recognition), extensive research based on theoretical models of word reading (e.g. Coltheart, 2005; Plaut, 2005) shows that different formats and tasks should be used in order to receive sufficient information about a subject's word recognition/decoding skills. The assessment process therefore often includes monitoring the reading of regular and irregular words and also pseudowords. Also while decoding measures based on simple model of reading assess the accuracy of real word and pseudowords reading (Hoover & Gough, 1990), numerous studies confirm that including the indicator of reading speed improves the prediction models of reading comprehension significantly (e.g. Cutting & Scarborough, 2006; Caravolas, Volín, & Hulme, 2005). Recently, the unique contribution of word reading in and out of the context has also been discussed (Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Ardoin et al., 2013).

When one is trying to grasp reading comprehension and describe its development, the situation becomes even more complicated. Understanding is not just something "that is" or "that is not". As Kamhi (2005) points out, it is more like a continuum from literal to creative and comparative levels of interpretation and this is what makes it very difficult to assess. When addressing reading comprehension van den Broek and Espin (2012) points out it is important to distinguish between the product (the mental representation of the text) and the processes that occur as the reader proceeds through the text. Although theoretical models differ in identification of the critical source of comprehension, there is a consensus in classification of these processes in two categories: lower and higher level processes. It is well documented that lower level skills such as word recognition and decoding represent the major determinant of reading ability in early elementary grades but in the later grades the correlation between word-level reading skills and reading comprehension declines (Curtis, 1980; Hoover & Gough, 1990) and higher level language skills such as inference making become more important (Oakhill & Cain, 2012)

To assess reading comprehension in English various tests are used by practitioners and researchers. Traditional tests assess the ability to answer question (often in multiple format) about text passages (e.g. NARA[2], GORT[3]) or complete missing words in text passages (e.g. WJPC[4]). While all measure reading comprehension, the variety of formats and administration procedures raise concerns about reading comprehension tests accuracy and interchangeable use. Moreover, in the light of current body of research, frequent criticisms occur about the unidimensionality of the current available assessments, their failure to implicate the multidimensional and multicomponenental construct of reading comprehension in assessments design and explain individual differences in processing system and component skills (e.g. Hannon & Daneman, 2001; Sweet, 2005; Sabatiny, O'Reilly, & Deane, 2013)

Just as the situation in other countries, Czech researchers and counsellors have paid more attention to word decoding than to the comprehension process. While this has changed elsewhere in the last two decades, not much has changed in research and counselling practice in our country. Up to now, there have only been a few studies mapping the development of reading comprehension, and not much is known about individual variability in comprehension skills. Reading comprehension is conceptualised in relation to the level of automaticity of decoding processes and intelligence. Matějček (1995) gives a minimum of 60-70 correctly read words per minute as a sufficient level of decoding skills regarding the mechanism of reading to be automatized enough to pay attention to the content of reading. It is believed that this is the level at which a child starts reading for amusement and is able to learn by reading. For normally developing readers we observe this level of decoding by the end of second grade and it has been common practice that when this reading speed is reached, special support interventions end. It would be misleading to claim that during the intervention, specialists did not care about the improvement of reading comprehension, but the main and most systematic training was concentrated on decoding related skills – letter recognition and sounding, syllable reading, etc. No systematic intervention programme for the improvement of comprehension has been evaluated or introduced to educational professionals.

Currently, there is only one research report where we can find valuable information about the reading comprehension of Czech children in the early grades of school (second-fifth). In their study Caravolas, Volín, & Hulme (2005) conducted a multiple-group path model of the predictors of reading comprehension in Czech and English. For both languages, the data revealed four unique predictors of reading comprehension: reading speed, phoneme awareness, vocabulary, and Raven's matrices. However, as important as the results are, the study has several limits.

First, when the reading comprehension score was being estimated, a timed cloze test modelled on the Gates Basic Reading test (Gates, 1958) was used. A cloze test consists of short passages that are

---

2. Neale Analysis of Reading Ability (Neale, 1997)

3. Gray Oral Reading Test (Wiederholt & Bryant, 2001)

4. Woodcock-Johnson Passage Comprehension subtest (Woodcock, McGrew, & Mather, 2001)

usually graded in terms of length, vocabulary, and general knowledge. In each passage, there are some missing words. While the test is filling in the missing words (often in a multiple-choice version) it is assumed that he/she has to understand the meaning of the text and activate several cognitive resources, such as lexical knowledge, word retrieval, and knowledge about syntax and text structure (Rathvon, 2004). Though we find this type of measure of reading comprehension useful because of its ease of administration and good reliability and validity, the comprehension that is assessed is restricted to reading sentences or short passages. It does not inform us about the child's ability to construct the global meaning of a text. In contrast, it is complex texts that children encounter and are required to comprehend most often in order to reach successful academic achievement. We just cannot assume that good performance on a cloze test also means good performance on a test in the format of reading a story or that the converse is true. Instead, we must carefully consider the selection of the measure according to the type of comprehension we aim to assess. A growing body of evidence warns against the interchangeable use of reading comprehension measures and contests their universal validity (Cutting & Scarborough, 2006; Cain & Oakhill, 2006; Keenan, Betjemann, & Olson, 2008).

Second, in the past decade, an increasing number of teachers and schools have been using a new instructional method for reading[5] (genetic, spelled reading). along with the traditional reading instruction method (analytic-synthetic, syllabic reading), It is necessary to reflect this situation and also to include children instructed by the genetic method of reading in the research sample and analyse the possible impact of this instructional approach on the development of literacy.

Third, even though the path model analysed the unique contribution of many skills and processes to reading comprehension, there are others that need to have their role investigated for possible intervention purposes. Besides morphological awareness, the issue of listening comprehension in particular deserves much more attention in the diagnostic and educational process in the Czech Republic.[6] Listening comprehension develops from early childhood and reflects the ability to understand language. Although it is not clear whether reading and listening comprehension represent the same or different processes (Samuels, 1987; Sticht, Beck, Hauke, Kleiman, & James, 1974), many research studies declare that there is a close association between listening comprehension and later achievement in reading (e.g. Hoover & Gough, 1990).

Reflecting these three concerns and on the basis of rich research evidence from abroad, we prepared a project called Reading Comprehension – typical development and its risks, that aimed not only to map the early development of reading comprehension but also to explore internal and external factors influencing reading comprehension in the Czech language and educational context. In this article, we focus on the investigation of new comprehension tasks designed to provide addi-

---

5. Both methods will be described later in the text.

6. So far no Czech data regarding the development of listening comprehension or its relation to reading comprehension is available. Neither is there any standardised way of measuring listening comprehension.

tional information (along with the cloze test) about early reading comprehension. The first part of this article overviews the reading comprehension construct and issues regarding its assessment. In the second part, we present descriptive data analysis and discuss further steps in the process of the development of a new set of formal comprehension measures.

THEORETICAL BACKGROUND

Construct of reading comprehension

Considering the product, reading comprehension is usually defined as the mental representation of a text after reading it; the more successful the comprehension, the more coherent the representation. To build a coherent representation, the reader has to connect elements of the text though semantic relations and construct an integrated whole (van den Broek & Espin, 2012). Unfortunately, we cannot measure mental representation directly, but we can get an idea about it by asking the reader what he/she remembers from the text. The tricky thing is, as we already noted, that there are different levels of comprehension and the elaboration of each of these model levels may vary from limited to complete and from shallow to deep. So for diagnostic purposes it would be beneficial if the instrument does not provide only an overall score but diverse scores giving information about partial skills.

In their theoretical model, van Dijk and Kintsch (1983) distinguish between three levels of basic comprehension. The surface level preserves exact wording and syntax. The text-based level refers to meaning based on explicit text content. The situation model corresponds to global and deeper comprehension and it is described as mental representation of what the text is perceived to be about. When constructing a situation model, the reader is required to integrate text-based information with information from their prior knowledge. Therefore, in order to map the ability to construct these mental models, test items addressing the reader's construction of meaning should include both literal (text-based) and inferential questions.

Inference making is considered to be a crucial ability to comprehend. While literal questions map information explicitly stated in the text, answering inferential questions requires the reader to go beyond explicit details. In their study Bowyer-Crane and Snowling (2005) classified several types of inferential questions. For example cohesive inferences that rely on linguistic cues in the text, knowledge-based inference or evaluative inference using real-world knowledge in interpreting the textual information relating to outcome of the event. Though it may be challenging to distinguish explicit and implicit comprehension, Eason, Goldberg, Young, Geist and Cutting (2012) warn against combining both types of questions in one comprehension construct. Findings show that performance on inferential questions is poorer and that for children, questions assessing literal comprehension are easier. When considering the developmental aspect of inference generating, inferences change not only in quantity but also quality. Older and more experienced readers identify more semantic

connections and wider variety of them (for review see e.g. Kendeou, van den Broek, Helder, & Karlsson, 2014).

When examining the construction of the meaning of a text we should bear in mind that it is a dynamic interaction among readers and texts (Wixson & Peters, 1987) and that the output is affected not only by the language and cognitive skills of the reader but also by a reading mode and the genre of the text. In addition, the findings about developmental patterns should be interpreted carefully in relation to the specific language and educational context.

Reading mode

At the beginning of instruction in reading it is a common practice that a child learns to read aloud. This enables teachers and parents to monitor the progress of their reading and to provide feedback regarding the accuracy and fluency of their decoding. When considering early reading comprehension, some of the research findings indicate the advantage of oral reading in comparison to silent reading (e.g. Fletcher & Pumfrey, 1988; Prior et al., 2011). In other words, beginner readers achieve better comprehension scores after oral than after silent reading. By the end of their elementary years, the superiority shifts from the oral mode to silent reading. As Prior et al. (2011) argue, findings also indicate that we should not regard silent reading as an identical copy of oral reading. In accordance with Vygotsky's theory of speech, an internalised phenomenon is reconstructed and modified – and once decoding skills are mastered, reading aloud can become disturbing for understanding the meaning of the text.

Since silent reading with comprehension is the goal of reading instruction, it is essential that a comprehension measure should be able to evaluate this aspect of reading. It can allow better understanding of the process of comprehension but also its underlying skills. For example, in contrast to previous results, some studies indicate no difference in comprehension mastery after oral versus silent reading but point out differences in efficiency (McCallum, Sharp, Bell, & George, 2004). In one-to-one administration conditions children from kindergarten and grade six read silently with the same comprehension scores but faster. This raises questions about the ecological validity of assessments. Another very important question we have to ask is about the merits of oral versus silent reading in the case of poor readers. Are educators and parents right when, in trying to help support comprehension, they let poor decoders read the text silently? Without including a measure of silent reading in reading comprehension assessments we cannot search for answers.

Text types

In a simplified view, we can distinguish two main stages in the process of the development of reading. There is a first phase in which children learn to read and a second in which they are expected to learn by reading. Although the transition is gradual and the phases overlap, these conditions have

an impact on the selection of the reading activities employed in the class. At the beginning of reading instruction, reading activities focus on progress in decoding accuracy and fluency. Besides wordlists, children read passages of narrative texts. Looking at these texts, Catts, Hogan and Adlof (2005) point out that they are linguistically quite simple and impose low demands on knowledge of vocabulary and grammar in order to understand the content. This is not the same in the case of the texts children encounter later. Usually, after third grade, reading activities rely heavily on reading expository texts, with the emphasis on comprehension (Best, Floyd, & McNamara, 2008).

Though little research has investigated reading comprehension of expository texts among elementary schoolchildren, there are some research findings available. It is well documented, for example, that young readers comprehend narrative texts more easily than expository texts and that different skills are needed for understanding narrative texts and expository ones (Best et al., 2008; Diakidoy, Stylianou, Karefillidou, & Papageorgiou, 2005). When comparing them, researchers name several major different features of each type of text. Narrative texts have a familiar structure for children and include causal relations. In contrast expository texts have abstract and logical structure that is unfamiliar to children. It has also been discussed that these characteristics have a significant influence on the strength of the relationship with listening comprehension (Diakidoy et al., 2005).

The Czech language and educational context

Before we move on to educational context, it might be useful to briefly provide some information about the Czech language system. In the writing system, orthography has a significant effect on the development of literacy (Perfetti & Dunlap, 2008). Alphabetical orthographies vary along a continuum of orthographic transparency (the consistency of grapheme-phoneme correspondences) and it is supposed that this influences the strategies adopted by readers (Perfetti & Dunlap, 2008; Ziegler & Goswami, 2005) and also the rate of acquisition of reading skills (Caravolas et al., 2005).

Czech is a Slavonic language with rather transparent orthography (for more details about Czech orthography see e.g. Caravolas et al., 2005). Czech is an inflectional language and because of this characteristic, when expressing the meaning the word order is very flexible in comparison to English. On the other hand, the bounds between words are formally declared by a complex system of declensions (Mathesius, 2001). As we noted earlier, despite the differences between the English and Czech languages the first cross-linguistic research findings show similar patterns and associations between the components of reading comprehension (Caravolas et al., 2005). However, the validity of many other research findings from English-speaking countries is yet to be tested in languages that possess transparent orthography, such as Czech.

In addition, literacy and its development are an important part of the primary curriculum (Wildová, 2012). Even if there are many reading instruction methods, the two prevailing ones are the analytic-synthetic method and genetic method. Both methods are classified as phonic methods, but there are major differences between them.

The analytic-synthetic method has been used in the Czech Republic for decades. 75 % of Czech schools use it as their main method for reading instruction. This method is based on grapheme-to-phoneme correspondence, acquisition of the alphabet, syllable composition, and syllable reading. Children become familiar with open and closed syllables and then they learn how to put them together, creating words. Persistent training in reading helps children to fix the reading patterns, as a result of which their reading becomes quicker and more fluent. After they have fixed the reading patterns, the children begin to perceive the meaning of the text.

The genetic method is used at 20 % of Czech schools. The syllable is not considered the main unit in this method. Children first learn what a phoneme and a grapheme are and they immediately start with reading whole words. At the beginning, the words are very simple. The more confident in reading a child becomes, the longer the words he or she reads. This method uses only capital letters at the beginning. However, from the start, the children focus on the meaning of the text. The texts include pictures and the teachers also work with context. Furthermore, the children are motivated to express their ideas in written form and by the end of the first grade, they have usually written a book of their own.

Even if the methods are very different, there are no major differences in the quality of reading achievements of young Czech students. Their results are very similar, especially after finishing the third grade. Several studies have been conducted on what the differences between the two methods are and the potential reading difficulties that the children with each method; however, none of the studies focused in depth on reading comprehension. This is why our research focuses solely on reading comprehension.

When assessing reading skills *Reading Assessment* (Matějček et al., 1987) is the most commonly used Czech testing set. This set includes 11 standardised texts and it allows several indicators of decoding skills to be measured: accuracy, fluency (speed), technical quality of reading (technical quality means what reading method the children are using), and reading comprehension. In addition, they analyse the mistakes that the children make.

Reading comprehension is assessed on the basis of how well the children can reproduce the text. First, the child is asked to read the text out loud. Second, the examiner asks the child to retell the story. If the child struggles, the examiner is allowed to ask questions and help the child. This form of assessment has many disadvantages (Kucharská, 2014). Is it really possible to deduce whether the child understood the text on the basis of how well he or she retells the story? Could the results be affected by the child's character traits (e.g. the child is shy or afraid of failure)? The child might not

be motivated enough to retell the story. Another issue is that the child might not have an adequate language level or the necessary oral expression skills. In such cases, the child might be able to understand the text; however, he or she may then be unable to reproduce it. Many practitioners and researchers point out other limitations of this set of tests: old reference norms (from the 1980s) and outdated language used in the stories.

The second most commonly used set is the *Set of diagnostic tests of literacy skills for students of the 2nd-5th grades* (Caravolas & Volín, 2005). The set includes an instrument especially created to assess reading comprehension (*Test of reading comprehension*), which we referred to in the introduction of the article. This cloze format test is the only standardised test of reading comprehension with valid reference norms in the Czech Republic. But as pointed out, using short passages of texts do not allow us to measure the ability to generate more elaborative inferences such as identify key ideas and integrating larger text units into a coherent whole (Sabatini et al., 2013; Valencia, Hiebert, & Afflerbach, 2014).

To complete the brief picture of the Czech educational context and approach to reading comprehension issues, we have to mention information about the reading performance that Czech students achieve in comparative international studies. The comparisons of the results of literacy levels in different countries show that the literacy level of Czech students is not at the highest level (e.g. the PISA and PIRLS research studies between the years 2001 and 2011). The level of reading comprehension has been pointed out as a critical point for Czech children (Vykoukalová & Wildová, 2013; Starý & Laufková, 2015).

SET OF NEW COMPREHENSION TASKS

When creating the tests, we took into consideration the requirements of classroom approaches to reading. Since high importance is attributed to reading comprehension in the Czech Republic, we wanted to base our tests on longer written accounts – stories. First, the children were asked to read the text out loud, and then they proceeded to answer questions about the text. Besides reading out loud, the children were assessed on their silent reading and listening comprehension. The text was read by a professional and the children listened to the recording once before they responded to the questions. The silent reading tests and listening comprehension tests have never been used in the Czech Republic before and are a key element of the original contribution of the research.

TABLE 1 Set of new comprehension tasks (Kucharská et al., 2014)[7]

Reading comprehension		Listening comprehension		Silent reading comprehension
Narrative	*Expository*	*Narrative*	*Expository*	*Narrative*
Going to the mountains (grade 1) Rabbits (grade 2-4)	Skating (grade 2-4)	Little Star (grade 1-4)	Mushrooms (grade 1-4)	Great friends (grade 1-2) Going on a trip (grade 3-4)

All the original tests were created in the same way. They contain the same number and type of questions (Table 2 and Table 3). Half of the questions focus on explicit, half on inferential comprehension. Explicit (also text-based or literal) comprehension assesses how much the children understood the information from the text and how much of the information they can recall. Inferential questions assess whether the children are able to "read in between the lines" and infer meaning. In addition, the scoring system is identical for all the questions – children can obtain 0, 1 or 2 points for each question.

The fact that the internal structure of the tests and questions is the same gives us the opportunity to observe the results in a comprehensive way. We are able to see the connections between individual tests. For example, we are able to tell whether a child has achieved the same scores for the inferential comprehension in the silent reading test, oral reading test, and listening comprehension test.

TABLE 2 Categories of literal questions – explicit (text-based) comprehension (Kucharská et al., 2014)

E 1	Straight question aiming at the content of defined category (text details recalling)	Answer: yes/no, word, enumeration of words, short sentence.
E2-E3:	Differentiating between present and absent details (with distractors)	Selection from a set of answers created by an administrator.
	E2: Phonological distractor	Word phonologically alike, rhyming, same number of syllables (vrcholek-hrbolek) (cherry-ferry).
	E3: Semantic distractor	Words phonologically and orthographically identical but with the same or different meaning (liška-liška) (band-band), or words with the same basis but shifted meaning (granule-granko) (hostel-hospital).
E4	Differentiation of the main and collateral characters and events	Answers: correct answer, ultimately choice from a prepared set of words, naming.
E5	Grasp of the storyline	Answers to questions which will specify the plot (story order). Answers: short sentences or a set of prepared answers.

---

7. In this article, we focus only on narrative comprehension tasks.

TABLE 3 Categories of inferential questions (Kucharská et al., 2014)

I6	Deciding the correctness of a claim in comparison with the text	Answers: from a set of prepared answers.
I7	Deciding the meaning of unknown words from context (semantic)	Answers: by definition, must be written down word-for-word by us.
I8	Inference based on a combination of information from the text and empathy with the character from the text (emotional, personal).	Answers: short sentence.
I9	Inference based on information from the text and understanding of the motivation of the hero	Answers: short sentence.
I10	Inference of the real meaning (instead of the literal meaning)	Answers: short sentence.

Aims

Numerous studies confirm that reading comprehension is not a unitary construct. In order to map the development of the reading comprehension of early-grade Czech students, we prepared and piloted a new set of comprehension tasks. In this paper, we aim to:

1. assess the reliability and validity of these new comprehension tests;

2. report performance data across our research sample.

Because of space limitations, we focus only on the analysis of a listening comprehension task and an oral and silent reading task based on a narrative text.

METHOD

Participants

The participants (N=515) were normally developing children from monolingual Czech environment which have not been diagnosed with sensory, neurological or cognitive disorders and haven't shown any signs of learning difficulties in early development. A guardian's consent was necessary and in compliance with ethical principles, parents were given the option to withdraw their consent at any point of the research.

Seventeen mainstream primary (elementary) schools, mostly from Prague, Central Bohemia and South Bohemia participated in the research. In accordance with the ethical requirements, information and results of each student was anonymous. Table 4 shows the age of children in each grade. When processing data analysis, we investigated differences in results achieved between teaching

(instructional) methods of reading[8]. But we didn't find fundamental differences (Kucharská &Wildová, 2015), that is why we included all students data in our analyses.

TABLE 4 Participants numbers, mean ages, and standard deviations, by grade

Grade	Phase	N	Age (months)		
			Mean	SD	Min-max
1	T1	131	81.8	5.44	71-104
	T2	131	88.8	5.37	78-111
2	T1	125	93.3	4.60	82-106
	T2	125	99.4	4.65	87-112
3	T1	125	104.7	4.52	96-116
	T2	125	110.8	4.62	102-123
4	T1	134	117.5	5.13	100-135
	T2	134	123.5	5.09	105-141
TOTAL		515			

MATERIALS

New comprehension tasks

We created three types of text comprehension tests regarding different comprehension modes – oral reading comprehension, silent reading comprehension and listening comprehension. Comprehension was evaluated through total score and two subscores (explicit and inferential). The maximum score was 20 points (10 points in each part of the test). All tests have similar type of questions, which are presented in Table 2. As mentioned earlier, in this article we focus on narrative comprehension tasks.

---

8. Approximately half of the children were taught by analytic-synthetic, half by genetic method of reading.

- *Narrative Listening comprehension (LC) Little Star* (fairy-tale kind of story). This test was administrated in all age group of students. Text had 163 words and we used a professional voice recording during the assessment. Listening comprehension is not dependent on reading skills. That is why only one version was used for all grades.

- *Narrative Oral reading comprehension (RC) Going to the mountains.* The test was designed for beginning readers (1st grade) and it contains 75 words. In addition to reading comprehension score we included two additional scores, the total time required to read the entire text, the score for number of words read correctly in time (1st minute, 2nd minute and third minute), and the number of errors.

- *Narrative Oral RC test Rabbits.* With a range of 159 words, the test was designed for more advanced readers (2nd to 4th grade). The total time of reading the text, the number of words read, and the errors were evaluated.

- *Narrative silent RC task Great Friends* is a version for 1st and 2nd grade students. Going on a trip is a version for 3rd and 4th grade student. Both tests were similar, differing in the length of the text and the difficulty level. Students read the text independently, without a time limit then they filled out the questionnaire in writing. Written text was present during whole testing period for students to revise or clarify details.

Reading comprehension test (Caravolas & Volín, 2005)

This standardized timed Reading comprehension test (RCT) provides a screening of a child's global reading level. The test is based on a text completion and consists of short passages (one to three sentences). Children are asked to fill in a missing word which they choose from a group of five words. Four words are distractors with various relationship with the target word (semantic, orthographic or phonology) or are completely unrelated.

Score 1 provides global index about reading skills and it is derived from the total number of correctly completed items. Score 2 informs about percentage accuracy and while omitting the factor of reading speed it refers more to the comprehension.

One Minute Reading Test (Caravolas & Volín, 2005).

This test was used to measure the speed of sight word recognition. The test consists of 140 real high-frequency regular words with increasing difficulty in phonological complexity arranged into columns. We record all correctly read words within the 60 seconds time limit.

Procedure

We collected data in two phases (T1 and T2) during the school year 2013/2014. The first phase (hereinafter referred to as T1) took place in the period from October to December 2013 and the second phase (hereinafter referred to as T2) in the period from March to May 2014. Within T1 there were three 60-minute sessions (depending on the pace of the child and his/her abilities) and during T2 we

executed two 60-minute sessions. The assessment was conducted by research assistants who were specially trained by research team members and worked under supervision.

TABLE 5 Data Collection

Phase	Session	Duration
T1	1	14/10 – 1/11/2013
	2	4/11 – 22/11/2013
	3	25/11 – 20/12/2013
T2	1	24/3 – 25/4/2014
	2	28/4 – 30/5/2014

RESULTS

Reliability

To examine the reliability of the measures we used two approaches.

Internal consistency

First, for each of the new measures, we explored internal consistency using Cronbach´s alpha. Due to the intention to map different levels of understanding (text based and inferential), we assessed reliability not only for the total scores but also for tests´ subscores (Table 6). As we can see coefficients alpha are higher for total scores (ranging from .42 to .63) than for explicit (.38-.42) and inferential subscores (.11-.58).

TABLE 6. Internal Reliability scores (coefficient alpha)

Comprehension test	N	Cronbach alpha		
		Total score	Explicit score	Inferential score
Narrative LC Little Star (Grade 1-4)	492	.63	.42	.56
Narrative Oral RC Rabbits (Grade 2-4)	372	.42	.40	.11
Narrative Silent RC Great friends (Grade 1-2)	237	.46	.38	.28
Narrative Silent RC Going on a Trip (Grade 3-4)	252	.61	.42	.58

When assessing total score reliability for each grade (Kucharska, Seidlová Málková, & Špačková, 2015), the reliability coefficients are lower (.31 to .60) For narrative LC test Little Star reliability scores showed increasing tendency with higher grades. It ranged from 0.45 at first grade to 0.6 at fourth grade. For narrative oral RC test Rabbits the reliability score ranged from 0.30 to 0.41 (with highest volume for the third grade). Both versions (for younger and older students) of narrative silent RC tests

produced low alpha levels ranging from 0.30 to 0.59 with the pattern: the higher grade, the lower alpha.

Test-retest reliability

Because only two comprehension measures were administered at Time 1 and Time 2, we can explore this type of reliability assessment only for narrative LC and oral RC task (Table 7) Correlation coefficients are higher for narrative LC task *Little Star* than for oral RC task *Rabbits*. Although all correlations are significant there seems to be an opposite pattern for listening and reading comprehension tasks. While test-retest correlation is lower for explicit score in the case of LC, for oral RC we found low correlation coefficient for total and both subscores.

TABLE 7 Correlation between measures at Time 1 and Time 2

Time 1 scores		Time 2 scores						
		Narrative LC Little Star (N=457)			Narrative oral RC Rabbits (N=358)			
		TS	ES	IS	TS	ES	IS	
Narrative LC Little Star	TS	.72***	0.52***	0.70***				
	ES	.50***	0.47***	0.53***				
	IS	.63***	0.41***	0.65***				
Narrative oral RC Rabbits	TS				0.48***	0.38***	0.40***	
	ES				0.44***	0.31***	0.40***	
	IS				0.32***	0.20***	0.33***	

Note: TS – total score, ES – explicit score, IS – inferential score. ***p < 0.001

Validity

We explored the concurrent criterion-related validity of all new comprehension measures by correlating total comprehensive scores with the criterion measures (Table 8). For validity assessment, we selected already standardized reading tests with good reliability and validity index. We explored correlations of total scores of new narrative comprehension tests with *Reading comprehension test* (Caravolas & Volín, 2005) and *One minute reading* (Caravolas & Volín, 2005).

TABLE 8 Correlation between new comprehension narrative tests and criterion measures at Time 1

Test	N	Reading comprehension test		One minute reading
		Score 1	Score 2	
Narrative LC Little Star (grade 1-4)	491	.48***	.36***	.44***
Narrative oral RC Going to the mountains (grade 1)	122	x	x	.37***
Narrative oral RC Rabbits (grade 2-4)	372	.38***	.27***	.27***

Note: LC – listening comprehension, RC – reading comprehension

Reading comprehension test was not administered to grade 1 children. For Narrative LC Little Star correlations with Reading comprehension test were computed only for grade 2-4 (N=369).

*p<.05 ***p<.001

TABLE 9 Correlations between new comprehension narrative tests and criterion measures at Time 2

Test	N	Reading comprehension test		One minute reading
		Score 1	Score 2	
Narrative silent RC Great friends (grade 1-2)	236	.18*	.21*	.36***
Narrative silent RC Going on a trip (grade 3-4)	252	.52***	.23***	.32***

Note. LC – listening comprehension, RC – reading comprehension

Reading comprehension test was not administered to grade 1 children. For Narrative silent RC Great friends correlations with Reading comprehension test computed only for grade 2 (N=119).

*p<.05, ***p<.001

   All new comprehension tests correlate significantly with selected criterion measures. Though the correlation coefficients are medium to low (.52-.18) all coefficients are statistically significant (in most cases p <.001). When comparing tests´ versions for younger and older children we can see slightly different pattern in correlations. Versions for older children (*Rabbits and Going on a trip*) correlate stronger to *RCT*, total comprehension score of test version for younger children (*Great friends*) correlate stronger to *One minute reading* (a decoding measure).

When assessing correlation for each grade separately, we found mixed results. All new narrative comprehension tests show low to medium, but statistically significant correlation to *RCT* score 1 (Pearson´s coefficients ranging from .18 to .53, p<.05). Correlations to *RCT* score 2 are lower but in most cases still statistically significant (p<.05). Within grades correlations between narrative comprehension tasks and *One minute decoding* tests are very low. Actually, there were statistically significant correlations only for silent (.28, p<.05) and oral (.37, p<.001) RC tests at grade 1 and silent RC *Going on a trip* (.35, p<.001) at grade 3.

Performance across research sample

All new comprehension tests shared identical structure and for a given grade level contained text of similar difficulty. The tests contained the same number of questions, the same types of questions and distractors (see Table 2). All of them assessed explicit and inferential comprehension. We were trying to investigate how demanding (difficult) each of the tests is in comparison to other tests and if the percentile performance is similar among all the observed grades.

The achieved results are contained in the Table 10 and Figure 1-4. They show the average percentile performance in the test. E.g. the figure 50% indicates that on an average, the children of a given grade reached half of the maximum number of points, i.e. 10 points.

Results of the 1st grade pupils: Repeated measures ANOVA showed significant effect of differences between difficulty of the tests, $F_{(2,228)}=255.5$, p<.001, $\eta2=.48$. Post-hoc paired t-test showed difference between all comparisons (all ps < 0.001 with applied Bonferroni correction). *Little Star* was the most difficult followed by *Going to the mountains* and *Great friends*.

Results of the 2nd grade pupils: Repeated measures ANOVA showed significant effect of differences between difficulty of the tests, $F_{(2,210)}=372.3$, p<.001, $\eta2=.65$. Post-hoc paired t-test showed difference between all comparisons (*Little Star* vs. *Rabbits* p < 0.06, *Little Star* vs. *Great friends* and *Rabbits* vs. *Great friends* ps < 0.001 with applied Bonferroni correction). *Little Star* was the most difficult followed by *Rabbits* and *Great friends*.

Results of the 3rd grade pupils: Repeated measures ANOVA showed significant effect of differences between difficulty of the tests, $F_{(2,220)}=59.1$, p<.001, $\eta2=.17$. Post-hoc paired t-test showed difference between all comparisons (*Little Star* vs. *Rabbits* p < 0.05, *Little Star* vs. *Going on a trip* and *Rabbits* vs. *Going on a trip* ps < 0.001 with applied Bonferroni correction). *Little Star* was the most difficult followed by *Rabbits* and *Going on a trip*.

Results of the 4th grade pupils: Repeated measures ANOVA showed significant effect of differences between difficulty of the tests, $F_{(2,252)}=112,6$, p<.001, $\eta2=.30$. Post-hoc paired t-test showed difference between all comparisons (*Little Star* vs. *Rabbits* p < 0.01, *Little Star* vs. *Going on a trip* and *Rab-*

bits vs. *Going on a trip* ps < 0.001 with applied Bonferroni correction). *Little Star* was the most difficult followed by *Rabbits* and *Going on a trip*.

To sum up, performance in each new test increase with higher grade level and new tests are not equally difficult across grades. We will reflect these findings in discussion.

TABLE 10 Performance on new comprehension tests – mean percentage (out of max. 20 points)

Test	N* (missing)	Mean percentage	SD	Median	Min.-max.
Grade 1					
Narrative LC Little Star (grade 1-4)	115	46.70	15.95	50	10-80
Narrative oral RC Going to the mountains (grade 1)	(16)	65.57	15.04	65	25-100
Narrative silent RC Great friends (grade 1-2)		82.78	14.66	90	40-100
Grade 2					
Narrative LC Little Star (grade 1-4)	106	51.98	15.96	55	15-80
Narrative oral RC Rabbits (grade 2-4)	(19)	57.45	13.53	60	25-90
Narrative silent RC Great friends (grade 1-2)		92.83	9.74	100	60-100
Grade 3					
Narrative LC Little Star (grade 1-4)	111	58.47	18.16	60	20-95
Narrative oral RC Rabbits (grade 2-4)	(14)	63.60	14.62	65	30-95
Narrative silent RC Going on a trip (grade 3-4)		75.5	15.14	80	25-100
Grade 4					
Narrative LC Little Star (grade 1-4)	127	62.8	17.46	65	10-100
Narrative oral RC Rabbits (grade 2-4)	(7)	67.32	12.32	65	40-95
Narrative silent RC Going on a trip (grade 3-4)		83.46	10.49	85	30-100

*Note: Children with missing data from one of the three tests we excluded.

Figure 1-4 Performances on the reading comprehension tests – achieved level of an average percentile performance*

*Note: axe x – comprehension tasks (LC, oral RC, Silent RC), axe y - average percentile performance

When exploring the explicit and inferential subscores we can see that for early grades explicit score is higher in all comprehension tasks. As reported in Kucharska et al. (2015) in reading comprehension tasks, the difference in performance on explicit and inferential subscore within first and second grade is statistically significant: *Mountains* (t(121)=11,3, p<.001, d = 1.02 for grade 1); *Great friends* (t(116) = 5,8, p < 0,001, d = 0,53 for grade 1; (t(119) =4,571, p < 0,001, d = 0,42 for grade 2).

When comparing the subscores between grades while explicit reading comprehension score is significantly growing each grade until the third grade (*Rabbits* F(1,359) = 36,37, p < 0,001), for inferential scores we found significant differences in performance achievement only between grade 3 and grade 4 *Rabbits* F(1,359) = 26,80, p < 0,001) (Kucharska et al., 2015).

In the case of listening comprehension task *Little star* we can see a different pattern. No statistically significant mean difference between literal and inferential subscore within any grade was found. Between grades the explicit score ($F_{(1,478)} = 66,31$, $p < 0,001$) was significantly higher only between grade 3 and grade 2 (p = .003). Inferential subscore ($F_{(1,478)} = 66,31$, $p < 0,001$) increased gradually till grade 3 (p = 0.008 between grade 2 and grade 1, p < 0,001 between grade 3 and grade 2) (Kucharska et al. (2015).

DISCUSSION

Presented reliability and validity assessments as well as analysis of tests performance gave us opportunity to reflect on the key features of preliminary versions of selected new comprehension tasks. Though far from optimal, we find the results encouraging.

Reliability

As expected, internal consistency reliability assessment showed low alpha values (ranging from .42 to .63 for total scores). Internal consistency reliability informs about the strength of inter-correlations between test items. The higher the inter-correlations among test items, the higher are Cronbach´s alpha values (Streiner, 2003b). Although we may doubt, we can achieve an acceptable value of .80 in the case of diverse construction-meaning questions depending on different weights on decoding, language and cognitive skills or prior reader´s knowledge. We agree that the alpha coefficients are far from optimal and can be increased in the process of adjustments of the preliminary versions of new comprehension tasks. It is a well known fact that we can increase the internal reliability by adding more items (questions) to the tests. 10 construction – meaning questions per test might just not be enough for reliable measure of text comprehension. Besides increasing the internal reliability by increasing the number of items, we have to consider modification or replacement of selected items based on examination of item – total correlations and item difficulty and discrimination analysis.

Test – retest reliability revealed acceptable results about the consistency of tests scores through two separate measurements over time, especially for narrative LC task *Little Star* (r = .72; p<.001). Although for oral RC task *Rabbits* the coefficient was lower (r = .48, p<.001) , when interpreting the findings we have to keep in mind the factor of six months interval between administrations of the tests. In case of young children and the reading construct, it is a rather long period and we have to be aware of changes in performance under the impact of intensive reading instruction.

Validity

All new comprehension tasks show positive statistically significant correlation to standardized reading comprehension measure *RCT* (Caravolas & Volín, 2005) even after controlling for grade status. When evaluating validity, we also conducted correlation coefficient between new tasks and word reading test – *One minute reading* (Caravolas & Volín, 2005) as widely accepted key component of reading comprehension. Correlations were much lower but still statistically significant for the whole research sample. These results are consistent with the findings about inter-correlations among reading comprehension tests including cloze format instrument (e.g. Kennan et al., 2008). In accordance with the authors, we may suggest that observed modest correlations can indicate that new assessments and cloze format RCT share the same underlying comprehension construct but that they each measure different comprehension skills too. Based on the findings, we can also suggest that in consistency with other research studies we may expect that oral reading comprehension test of format with global text and answering questions assessment will depend less on word reading skills than a cloze format test (e.g. Kennan et al., 2008; Cain & Oakhill, 2006).

Performance across research sample

If we are to evaluate how demanding each one of the test was, the LC test (*Little Star*) appeared to be the most difficult to comprehend. Performance between the 1st and 4th grades increases gradually, but the task remains still quite challenging even for pupils of the 4th grade (average percentile performance of 62.8 %). In contrast, we found silent RC task (for both versions *Great friends* and *Going on a trip*) to be the easiest one. We can speculate what caused these results. Although all tests were constructed identically, they are variously demanding, which can be caused by two reasons:

- We failed to create an equally challenging questions (questions after LC were easier in comparison to questions after oral or silent reading).

- There was an effect of the instructions itself – the comprehension of a story is higher after oral or silent reading than after a listening process. Listening might be an unusual task for pupils, they don´t have a visual support of the text while in oral and silent reading they do. Oral reading might facilitate text perception because of the need to articulate each of the word, while during text listening, pupils are passive and the concentration may vary significantly.

To verify the second possible interpretation, we would like to administer each text with different instructions – e.g. the text *Little Star* would be presented to one group of students in the form of listening comprehension task, to an another group in the form of oral reading and to an another group in the form of silent reading task.

The highest performance scores on silent RC tests are in contrast to research findings indicating the advantage of oral reading in comparison to silent reading (e.g. Fletcher & Pumfrey, 1988; Prior et al., 2011) and also in contrast with findings about superior comprehension performance in the case of young readers (e.g. Diakidoy et al., 2005). But we have to keep in mind that for silent RC tasks our instruction conditions were different. Comprehension was assessed by written answers and while answering, the children would still keep the text, which allowed them to apply diverse reading strategies.

Regarding reading comprehension subscore analysis, these results are in line with findings indicating that for children it is easier to answer literal questions (Eason et al., 2012). While in reading comprehension tasks the inferential subscore does not grow significantly until third grade and so children from grade 1 to grade 3 do not differ significantly in their inferential reading comprehension subscores. For listening comprehension task the performance inferential subscore grows gradually until third grade.

Implications

Based on the results from data analysis of our new narrative comprehension tasks we can conclude two implications for reading practise in elementary grades in Czech where systematic reading instruction and diagnostics of reading difficulties rely heavily on word recognition and decoding skills.

First, findings regarding analysis of concurrent criterion-related validity indicate, as in other research studies, that the correlation between word recognition/decoding and reading comprehension declines after the first year of schooling. When conducting coefficients for each grade separately, statistically significant correlations were found only for oral and silent RC in grade 1 and silent RC in grade 3. It should become therefore a natural praxis that teachers and counselling practitioners map and systematically promote not only the word recognition/decoding skills of children but also reading comprehension skills. Automatized word recognition/decoding does not itself lead to a satisfactory comprehension.

Second, new comprehension tasks seem to not be equally demanding for pupils within each grade. Results also indicate that, though these tasks measure similar construct as Reading comprehension test (RCT) with reference to high reliability and validity index, they also measure another aspects of text comprehension. Teachers and counselling practitioners should therefore carefully map the comprehension skills using tasks of various formats and reading mode.

Though it is clear we are just at the beginning of the process of developing a standardized set of comprehension assessment instruments, we find the first descriptive findings encouraging. We believe these tasks have a potential to become a helpful tool in respect to these changes in diagnostics as well as in intervention.

REFERENCES

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing, 19*(9), 933–958. doi:10.1007/s11145-006-9024-z

Ardoin, S. P., Eckert, T. L., Christ, T. J., White, M. J., Morena, L. S., January, S. A., & Hine, J. F. (2013). Examining variance in reading comprehension among developing readers: Words in context (curriculum-based measurement in reading) versus words out of context (word lists). *School Psychology Review, 42*(3), 243– 261.

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository *Reading* texts. *Psychology, 29*(2), 137–164. doi:10.1080/02702710801963951.

Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure?. *British Journal of Educational Psychology, 75*, 189–201. doi:10.1348/000709904X22674

Cain, K., & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal Of Educational Psychology,76*(4), 697–708. doi:10.1348/000709905X69807

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*(1), 31–42.

Caravolas, M., & Volín, J. (2005). *Baterie diagnostických testů gramotnostních dovedností pro žáky 2. až 5. ročníků ZŠ (Set of diagnostic tests of literacy skills for students of the 2nd-5th grades).* Praha: IPPP ČR.

Caravolas, M., Volín, J., & Hulme, C. (2005). Phoneme awareness is a key component of alphabetic literacy skills in consistent and inconsistent orthographies: Evidence from Czech and English children. *Journal of experimental child psychology, 92*(2), 107–139.

Catts H. W., Hogan T. P., & Adlof S. M. (2005). Developing changes in reading and reading disabilities. In: H. W. Catts, A. G. Kamhi (Eds.). *The connections between language and reading disabilities*. (pp. 50–71). Mahwah, NJ: Lawrence Erlbaum Associates.

Curtis, M. E. (1980). Development of components of reading skill. *Journal Of Educational Psychology, 72*(5), 656–669. doi: 10.1037/0022-0663.72.5.656

Coltheart, M. (2005). Modelling reading: The dual-route approach. For Snowling, M.J. & Hulme, C. (Eds). *The Science of Reading*. Oxford: Blackwells Publishing.

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. doi:10.1207/s1532799xssr1003_5

Diakidoy, I. N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2005). The Relationship Between

Listening and Reading Comprehension of Different Types of Text at Increasing Grade Levels. *Reading Psychology, 26*(1), 55-80. doi:10.1080/02702710590910584

Fletcher, J., & Pumfrey, P. D. (1988). Differences in text comprehension amongst 7–8-year-old children. *School Psychology International, 9*(2), 133–145. doi:10.1177/0143034388092008

Eason, S. H., Goldberg, L. F., Young, K. M., Geist, M. C., & Cutting, L. E. (2012). Reader–text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal Of Educational Psychology, 104*(3), 515-528. doi:10.1037/a0027182

Fletcher, J.& Pumfrey, P. D. (1988): Differences in text comprehension amongst 7-8-year-old children.listeninl School Psychology International, 9,133-145. *Lesen, Hören, Verstehen [Reading comprehension, listening comprehension, language comprehension]*. Available from: https://www.researchgate.net/publication/280254621_Lesen_Horen_Verstehen_Reading_comprehension_listening_comprehension_language_comprehension [accessed Nov 17, 2016].Gates, A. I. (1958). *Gates basic reading test, Type LC, F-3*. New York: Bureau of Publications, Columbia Univ.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal Of Educational Psychology, 93*(1), 103-128. doi:10.1037/0022-0663.93.1.103

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

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of Individual Differences in Reading Comprehension and Reading Fluency. *Journal Of Educational Psychology, 95*(4), 719–729. doi: 10.1037/0022-0663.95.4.719

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21*(2), 85–97. doi:10.1080/02702710050084428

Kamhi, A. G. (2005) Finding Beauty in the Ugly Facts About Reading Comprehension. In H. G. Catts, & A. K. Kamhi (Eds.) *The Connections Between Language and Reading Disabilities* (pp. 201–213). New York: Lawrence Erlbaum Associates.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*, 281–300.

Kendeou, P., van den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice, 29*(1), 10-16. doi:10.1111/ldrp.12025

Kucharská, A. (2014). *Riziko dyslexie (Risk of dyslexia)*. Praha: PedF UK.

Kucharská et al., (2014). *Porozumění čtenímu I. (Comprehension of reading I.)*. Praha: PedF UK.

Kucharská, A., & Wildová, R., (2015). Čtení z pohledu výukových metod: analyticko-syntetická metoda a metoda genetická (Reading from the perspective of teaching methods: the analytic-synthetic method and the genetic method). In A. Kucharská *et al., Porozumění čtenému III. Typický vývoj porozumění čtenému - metodologie, výsledky a interpretace výzkumu (Reading Comprehension III. A Typical Reading Comprehension Development - Methodology, Results and Interpretation of the Research)*. (pp. 13–27). Praha: PedF UK.

Kucharska, A., Seidlová Málková, G. & Špačková K. (2015). Proozumění čtenímu – vývojová dynamika a jeho přepdokady. (Reading comprehension – developmental dynamics and its precursors). In A. Kucharská, et al., *Porozumění čtenému III. Typický vývoj porozumění čtenému - metodologie, výsledky a interpretace výzkumu (Reading Comprehension III. A Typical Reading Comprehension Development - Methodology, Results and Interpretation of the Research)*. (pp. 13–27). Praha: PedF UK.

Matějček, Z. (1995). *Dyslexie. Specifické poruchy čtení (Dyslexia. Specific reading disabilities)*. Jinočany: H&H.

Matějček, Z. et al. (1987). *Zkouška čtení (Reading Assessment)*. Bratislava: Psychodiagnostika.

Mathesius, W. (2001) *Nebojte se angličtiny! (Do not be afraid of English!)*. Jinočany: H&H.

McCallum, R. S., Sharp, S., Bell, S. M., & George, T. (2004). Silent versus oral reading comprehension and efficiency. *Psychology In The Schools, 41*(2), 241–246. doi:10.1002/pits.10152

Neale, M. D. (1997). *Neale Analysis of Reading Ability – Revised*. Windsor: NFER–Neslon.

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.

Plaut, D. C. (2005). Connectionist Approaches to Reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 24–38). Oxford, England: Blackwell.

Perfetti, C. A., & Dunlap, S. (2008). Learning to read: General principles and writing system variations. In K. Koda & A. Zehler (Eds.). *Learning to read across languages* (pp. 13–38). Mahwah, NJ: Erlbaum.

Prior, S. M., Fenwick, K. D., Saunders, K. S., Ouellette, R., O'Quinn, C., & Harvey, S. (2011). Comprehension after oral and silent reading: Does grade level matter? *Literacy Research And Instruction, 50*(3), 183–194. doi:10.1080/19388071.2010.497202

Rathvon, N. (2004). *Early Reading Assessment: A Practitioner's Handbook*. New York: Guilford Press.

Sabatini, J., O'Reilly, T. & Deane, P. (2013) Preliminary reading literacy assessment framewort: foundation and rationale for assessment and system design. *Research Report Series*, 2013: i–50. doi:10.1002/j.2333-8504.2013.tb02337.x

Samuels, S. J. (1987). Factors that influence listening and reading comprehension. In S. J. Samuels (Ed.), *Comprehending oral and written language* (pp. 295–325). San Diego, CA: Academic Press, Inc.

Scarborough, H. S. (2005). Developmental relationships between language and reading: Reconciling a beautiful hypothese with some ugly facts. In H. W. Catts & A. G. Kamhi (Eds.). *The connections between language and reading disabilities* (pp. 3–24). Mahwah, NJ: Erlbaum.

Starý, K., & Laufková, V. (2015). Kde žáci v polovině základního vzdělávání selhávají při porozumění čtenému polemickému textu. (Where pupils in the middle of primary education fail at reading comprehension polemic text). *Pedagogika, 65*(4), 392–412.

Sticht, T. G., Beck, L., Hauke, R., Kleiman, G. M. & James, J. (1974). *Auding and Reading: A Developmental Model*. Alexandria, VA: Human Resources Research Organization.

Streiner, D. L. (2003b). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal Of Personality Assessment, 80*(1), 99-103. doi:10.1207/S15327752JPA8001_18

Sweet, A. P. (2005). Assessment of Reading Comprehension: The RAND Reading Study Group Vision. In S. G. Paris & S. A. Stahl, (Eds.), *Children's reading comprehension and assessment* (pp. 3-12). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Valencia, S. W., Hiebert, E. H., & Afflerbach, P. P. (2014). *Authentic Reading Assessment: Practices and possibilities*. (reprint of 1994 edition). Santa Cruz: TextProject, Inc.

van den Broek, P., & Espin, C. A. (2012). Connecting cognitive theory and assessment: Measuring individual differences in reading comprehension. *School Psychology Review, 41*(3), 315–325.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Vykoukalová, V. & Wildová, R. (2013). Čtenářská gramotnost žáků 1. *stupně a možnosti jejího rozvoje (Reading Literacy and the possibilities for its development)*. Praha: PedF UK v Praze.

Wiederholt, J. L., & Bryant, B. R. (2001). *Gray Oral Reading Tests-IV (GORT-4)*. Austin, TX: Pro-Ed.

Wildová, R. (2012). Rozvoj pregramotnosti a počáteční čtenářské gramotnosti v kurikulu evropských zemí. (Preliteracy development and early literacy in the curriculum of European countries). *Pedagogika, 62*(1–2), 10–21.

Wixson, K. K., & Peters, Ch. W. (1987). Comprehension Assessment: Implementing and Interactive View of Reading. *Educational Psychologist, 22*(3-4), 333–356. doi: 10.1080/00461520.1987.9653056

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III. Itasca, IL: Riverside Publishing.

Ziegler, J. C., & Goswami, U. (2005). Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychological Bulletin, 131*(1), 3-29. doi:10.1037/0033-2909.131.1.3