TRIINU KÄRBLA

Assessment of text comprehension and
teaching comprehension strategies
in Estonian basic school

DISSERTATIONES PEDAGOGICAE UNIVERSITATIS TARTUENSIS
35

TRIINU KÄRBLA

Assessment of text comprehension and teaching comprehension strategies in Estonian basic school

Institute of Education, Faculty of Social Sciences, University of Tartu, Estonia

Dissertation is accepted for the commencement of the Degree of Doctor of Philosophy in Education on May 26, 2020 by the joint Doctoral Committee of the Institute of Education and Institute of Ecology and Earth Sciences for awarding doctoral degrees in education, University of Tartu.

Supervisors:	Prof. Krista Uibu, PhD
	Institute of Education, University of Tartu, Estonia
	Research Fellow Mairi Männamaa, PhD
	Institute of Clinical Medicine, University of Tartu, Estonia
Opponent:	Prof. Marja-Kristiina Lerkkanen, PhD
	Department of Teacher Education, University of Jyväskylä, Finland
Commencement:	Digital environment, on August 21, 2020, at 11.00 a.m.

European Union
European Social Fund

Investing
in your future

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The doctoral study is based on the following original publications that are referenced in the text by their Roman numbers.

I Kärbla, T., Uibu, K., & Männamaa, M. (2017). Eesti keele riiklike taseme-tööde tekstimõistmisülesannete analüüs [Analysis of the comprehension tasks of national standardised tests of Estonian language]. *Eesti Rakenduslingvistika Ühingu aastaraamat. Estonian Papers in Applied Linguistics, 13,* 73–87. https://doi.org/10.5128/ERYa13.05

II Kärbla, T., Uibu, K., & Männamaa, M. (2018). Pikilõikeline sissevaade eesti keele taseme-ja eksamitööde tekstimõistmisülesannetesse [A longitudinal introspective into the text comprehension tasks of national tests and exams of the Estonian language]. *Eesti Haridusteaduste Ajakiri. Estonian Journal of Education*, 6(1), 157–178. https://doi.org/10.12697/eha.2018.6.1.07

III Kärbla, T., Uibu, K., & Männamaa, M. (2019). National Estonian-language tests: What is measured in text comprehension tasks? *New Trends and Issues Proceedings on Humanities and Social Sciences, 6*(5), 8–16.

IV Kärbla, T., Männamaa, M., & Uibu, K. (2019). Vocabulary and Text Comprehension Levels: What should be Considered in Assessments? *Educational Psychology*. https://doi.org/10.1080/01443410.2019.1703172

V Kärbla, T., Uibu, K. & Männamaa, M. (2020). Teaching strategies to improve students' vocabulary and text comprehension. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-020-00489-y

The author's contribution to the publications:

For Article I: composing the theoretical framework in cooperation with supervisors, formulating research questions, analysing the data and writing article in cooperation with supervisors.

For Article II: composing the theoretical framework in cooperation with supervisors, formulating research questions, employing qualitative analysis as the main author, analysing the data and writing article in cooperation with supervisors.

For Article III: writing the theoretical framework, formulating research questions, analysing the data, reporting the data and writing the article in cooperation with supervisors.

For Article IV: writing the theoretical framework in cooperation with supervisors, formulating research questions, developing the research design and methodology in cooperation with supervisors,

developing and improving the text comprehension test, collecting the data in the research group, analysing the data in cooperation with supervisors, reporting and interpreting results, and writing the article as a first author.

For Article V: writing the theoretical framework in cooperation with supervisors, formulating research questions, developing the research design and methodology, elaborating the text comprehension test, designing and improving the intervention and the materials for teachers and students, conducting the training program for teachers in cooperation with supervisor, collecting and analysing the data, reporting and interpreting results, and writing the article as a first author.

Related articles:

Kärbla, T., & Uibu, K. (2016). Assessment of Students' Higher-level Text Comprehension Skills in Basic School. *New Trends and Issues Proceedings on Humanities and Social Sciences*, *2*(5), 31–37.

1. INTRODUCTION

1.1 Overview of the research context

Over the years, several theories of and approaches to text comprehension, which is very complex, have been proposed (McNamara & Magliano, 2009). Contemporary studies and theories have been influenced mainly by the Construction–Integration (CI) comprehension model introduced by Walter Kintsch (1988, 1998). Although the final product of comprehension is considered to be one mental representation of a text, it involves a full range of interacting processes and components (McNamara & Kendeou, 2011) to understand texts at multiple levels (Kintsch, 1998). For example, lower-level components, such as decoding speed and accuracy, are needed at the surface level, vocabulary is crucial to the creation of a text base, and higher-level components, such as inference-making and strategic processes, are needed to create a situation model (Angosto et al., 2013; Oakhill et al., 2015).

In general, text comprehension is needed to ensure economic, cultural and social development (United Nations…, 2016) as well as to convey knowledge in school at every level and in most subjects (van den Broek & Espin, 2012). Even with the rapid changes in society and the labour market, comprehension remains a core work-related skill (World Economic Forum, 2016). Thus, comprehension should be purposefully and comprehensively promoted in schools. Despite efforts to determine the best practices to improve students' comprehension proficiency, limited comprehension has continued to be an ongoing global problem (National Assessment of Educational Progress [NAEP], 2019a; National Reading Panel [NRP], 2000; OECD, 2019). Poor comprehension is one of the reasons for academic failure, which is associated with decreased learning motivation and, in some cases, dropping out of school (Snow, 2002).

The reasons why comprehension is still problematic may be due to several factors, including teaching methods and assessments of text comprehension (Fletcher, 2006). Successful teaching comprises the understanding of student progress in the various comprehension components. For instructionally relevant feedback, multidimensional tests of student proficiency in the processes and components for comprehending texts at different levels are needed (Kendeou et al., 2007). These tests could identify the students who can recall explicitly stated information (literal comprehension level), interpret the meaning of implicit information (inferential comprehension level), and go beyond the text (evaluative comprehension level; Basabara et al., 2013). Equally important is to provide the information of which comprehension components and levels are assessed with the test (Kendeou & Papadopoulus, 2012; VanderVeen et al., 2007). However, schools tend to use unidimensional assessments, which provide very little information about student achievement in comprehension (Keenan et al., 2008). These tests include mostly practical information, e.g., the format and administration of the tests. The comprehension components and processes measured by

the tests are not explained (Keenan & Meenan, 2014). Hence, teachers cannot receive the necessary feedback to effectively promote text comprehension processes and components at various comprehension levels.

Because text comprehension is a developmental process, it cannot be assessed in the same way for every age group. The development of lower- and higher-level processes and components begins at a very young age (Kendeou et al., 2009); however, their importance in determining skilled and poor comprehenders varies. Specifically, the differences in the youngest readers are characterised more by the lower-level components, e.g., decoding and vocabulary range (Floyd et al., 2012). Later, when the basic skills are mostly mastered, they no longer provide a basis for distinguishing between poor and good readers. Proficiency differences in the higher-level processes and components, such as inference-making and strategy use, are important predictors of older students' comprehension (Floyd et al., 2012; Torgesen et al., 1997). However, to facilitate the monitoring of students' progress in comprehension, the higher-level components, to some extent, should also be included in the tests for younger students.

In Estonia, students' text comprehension in their native language is assessed through national standard-determining tests and examinations that are administered at the end of each school stage: Grades 3, 6, and 9 (Riigi Teataja, 2018). Because the national standard-determining tests have a strong influence on what and how teachers promote in their lessons (Vestheim & Lyngsnes, 2016), it is critical that they be instructionally relevant. However, many of these tests still provide information about students' proficiency in answering factual questions rather than their ability to create situation models (Allington, 2001). Accordingly, these tests might guide teachers to focus on lower-level comprehension skills.

In Estonia's national standard-determining tests, comprehension is just one component, indicated by a single score, among the other language skills that are measured (Riigi Teataja, 2018). No information is provided about the comprehension components that are measured. Hence, teachers cannot receive appropriate feedback about their students' comprehension proficiency. It is unclear, if the students understand only explicitly stated information in a text, or can they understand the text by making inferences and analysing and critically evaluating it.

Even though comprehension assessments are necessary for gaining an overview of what should be taught in language lessons, this is not sufficient for ensuring successful instruction. To purposefully promote students' comprehension skills for creating text-based representations or situation models (Kintsch, 1998), teachers should have a thorough understanding of the comprehension concept and how to promote it. The cognitive view of comprehension highlights the importance of the active construction of mental representations through conscious activities; thus, the teaching of multiple comprehension strategies is crucial (Graesser, 2007). These strategies are necessary for students to deliberately engage with texts and to recognise and to repair comprehension failures (Dole et al., 2014). Teachers often use multiple comprehension strategies

(e.g., asking questions, generation, summarising); however, they do not always teach them to their students (Duke & Pearson, 2009).

The approach to teaching comprehension, specifically the teaching of comprehension strategies, in Estonian schools is of concern. This concern has been heightened by comparisons of Estonian and Finnish students' progress in text comprehension. Comparisons of Estonian and Finnish students are worthwhile because of the similarities in the orthographies and school systems (see Soodla et al., 2019). According to the Soodla and her colleagues' study (2015), Estonian children enter school with significantly better decoding skills than their Finnish counterparts. However, beginning in the second grade, they exhibit lower comprehension proficiency (Must 1997; Soodla et al., 2019). The reason might be differences in the instructional processes in the two countries. Although, several text comprehension strategies have been introduced by Vardja (see 2011), it may assume that Estonian teachers still lack the awareness of the need for students to learn text comprehension strategies and the knowledge to teach them.

To conclude, text comprehension is one of the most important skills for school and everyday life. Thus, it has been of great interest to researchers for many years (NRP, 2000). Because of the complex nature of comprehension, students have experienced difficulties in mastering text comprehension techniques (NAEP, 2019a, 2009b). Therefore, the examination of the comprehension components will facilitate the determination of the best practices for supporting student progress.

1.2 Research focus

Failure to become a skilled comprehender could lead to learning difficulties (Snow, 2002). For teachers, being able to get information from the comprehension assessments about what text comprehension components should be promoted, and how to best do it would be beneficial. Regrettably, the Estonian national standard-determining tests do not provide information about what is measured by the comprehension tasks. Accordingly, it is difficult for teachers to determine the students' proficiency in specific comprehension components. Thus, it is important to determine the theoretical underpinnings, if any, of the comprehension tasks in the national standard-determining tests and the utility of these tests for the development of instructional practices. To develop an efficient instructional tool for teachers to improve students' text comprehension, the various influences, such as individual differences and the relationships between the comprehension components, should be thoroughly examined.

The goal of this doctoral study is to provide a comprehensive view of what is measured by the Estonian national standard-determining tests and to develop an instructional tool to promote basic school students' text comprehension.

Three aims, including the following research questions, were established:

1. Examine the comprehension levels of the text comprehension tasks in the Estonian-language (native language) national standard-determining tests in order to develop multidimensional text comprehension and vocabulary tests.
    1.1 How are the text comprehension levels considered in the national standard-determining tests? (Articles I, II, and III)
    1.2 To what extent does the distribution of text comprehension levels in the national standard-determining tests and examinations change by grade? (Articles II and III)
    1.3 How can the text comprehension levels be distinguished, and what is their association with vocabulary in the newly developed vocabulary and text comprehension tests? (Article IV)
2. Determine the unique underlying patterns in student vocabulary and text comprehension.
    2.1 What are the possible subgroups of students with individual differences in vocabulary and text comprehension at the literal, inferential, and evaluative comprehension levels? (Articles IV and V)
3. Develop an intervention program for improving text comprehension, and assess its effectiveness on students' vocabulary and comprehension.
    3.1 How does the teaching of text comprehension strategies improve students' vocabulary and text comprehension at each level? (Article V)
    3.2 To what extent does the teaching of text comprehension strategies influence vocabulary and text comprehension in various subgroups? (Article V)

On the basis of the aims, this doctoral study comprises three sections. First, the text comprehension tasks in the Estonian-language national standard-determining tests were examined to map the current situation and to highlight the need to develop models for developing a text comprehension (Articles I, II, and III). Next, new vocabulary test (VT) and text comprehension test (TCT) were developed to examine the students' comprehension skills and the associations between vocabulary and the text comprehension levels (Article IV). To achieve the second aim, the students' vocabulary and text comprehension subgroups were determined (Articles IV and V). The third aim focused on the development and assessment of an intervention program to enhance student vocabulary and text comprehension (Article V).

2. THEORETICAL FRAMEWORK

2.1 Text comprehension approaches

Text comprehension involves multiple cognitive processes that facilitate the understanding of written or oral language. Both, reading and listening comprehension are focused on accessing the meaning of a message by understanding explicit and implicit information, making inferences, and creating a mental representation (Aryadoust, 2017; Duke & Carlisle, 2011). However, there are also some differences in the processes of reading and listening comprehension. For example, vocabulary and sentence structure complexity, text types, and sentence word order. In general, studies on text comprehension have concentrated on examining either reading or listening comprehension, whereas some studies focus on these two processes together (Aryadoust, 2017). In this doctoral study, *text comprehension* refers to the understanding of written discourse.

Text comprehension is an incredible accomplishment because of the variety of components and processes that are involved (Graesser, 2007; Kong, 2019). The complexity and lack of a clear definition and boundaries regarding comprehension (Paris & Hamilton, 2014; Tennent, 2015) have led to a plethora of definitions and theories (McNamara & Magliano, 2009). Earlier researchers have outlined two diametrically opposed viewpoints: comprehension as a bottom-up process and comprehension as a top-down process (see Figure 1).
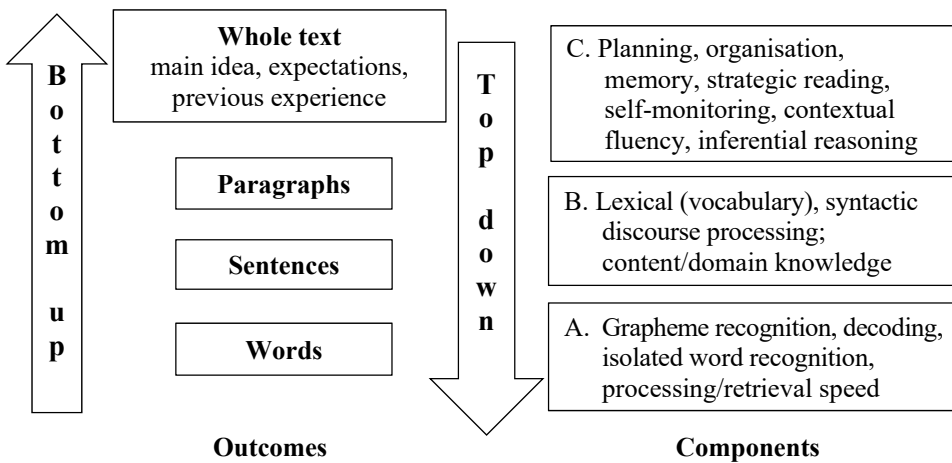

Figure 1. Components in and outcomes of bottom-up and top-down processes (based on Kintsch, 1988, 1998 and Kong, 2019)

In the bottom-up approach, comprehension begins with the decoding and understanding of words and a step-by-step progression towards the interpretation of the entire text. This approach highlights the importance of texts. The top-down

view starts with the whole text. The title, pre-knowledge, and expectations are used to make predictions about the text before moving to the smaller units, i.e., word decoding and recognition (Angosto et al., 2013). The focus is on the readers whose pre-knowledge and experience are brought to the comprehension process and, thus, interact with the text. Both approaches consider comprehension to be a strictly linear and sequential process. One stage is completed before another starts. Thus, before the reader moves to the next stage (e.g., from words to sentences or vice versa), the previous one has to be completed, and no correction is made to the earlier steps (Kong, 2019).

The perspective that comprehension is an interactive process between a text and a reader, between bottom-up and top-down processes forms a basis of the most influential current theories and comprehension studies (Angosto et al., 2013; Kong, 2019). Despite a wide variety of comprehension theories, the main view is that the product of comprehension is a mental representation that is generated by the interconnections between the textual information and the readers' pre-knowledge (Kendeou et al., 2014). The theories that have received the most attention, are categorised on the basis of their descriptions of the comprehension processes (McNamara & Magliano, 2009). First are the theories that describe the basic and overall comprehension processes. Examples are the Construction-Integration (CI) model (Kintsch, 1988, 1998) and the Landscape model (van den Broek et al., 2005). Second are the theories that detail the pre-knowledge retrieval and inferential processes. Examples are the Event-indexing model (Zwaan et al., 1995) and the Constructionist model (Graesser et al., 1994). Kintsch's (1988, 1998) CI model was a turning point in text and discourse research. It was the first to focus on comprehension processes and strategies in comprehension. Considered the most complete and well-formulated model, CI has provided a foundation for subsequent theories and models, including those discussed above (McNamara & Magliano, 2009).

The CI model is based on the idea that comprehension comprises three levels: *surface structure*, *propositional text base*, and *situation model* (Kintsch, 1998). The surface structure is related to the exact words and their syntactic relations in texts. Without meanings ascribed to words, there is little effect on comprehension. Therefore, it is generally not used in comprehension studies (McNamara & Magliano, 2009). In the propositional text base, words and sentences are combined into meaningful units (ideas and propositions) that in turn are integrated with the reader's pre-knowledge to construct a situation model (Kintsch, 1998).

The CI model, as its name indicates, combines the construction and integration phases (Kintsch, 1988). In the construction phase, the text base is created through the formulation of multiple parallel meanings of sentences. The inappropriate meanings are suppressed (deactivated) and excluded from the text representations during the integration phase. Only the appropriate constructions are organised to generate the *situation model*, which is the generation of a mental representation of what the text is about by integrating it with the reader's previous knowledge (McNamara et al., 2007).

The theoretical foundation of this doctoral study was the situation model and propositional text base levels in the CI model. The focus was student proficiency in the text comprehension components and processes, such as vocabulary and inference-making (see Figure 1, Boxes B and C), rather than basic skills (Figure 1, Box A). Although the basic skills are important in text comprehension, they are already developed and do not have a significant effect on text comprehension in Grades 4 and 6 (McNamara & Magliano, 2009), the sample in this study. Thus, literal, inferential, and evaluative comprehension, including the components and processes relevant to this age group, were examined.

2.2 Text comprehension components

Vocabulary, the knowledge of word meanings, is essential in the text comprehension process. It is a requirement for the formulation of text-based representations, the facilitation of the integration process, and the creation of situation models (Caemmerera et al., 2018; Cain & Oakhill, 2014). The distinction between the number of words (vocabulary breadth) and the detailed knowledge of words, i.e., the quality of word meanings (vocabulary depth) has been made (Ouellette, 2006). Both of these elements are necessary for text comprehension. Vocabulary breadth facilitates the use of the lower-level components, e.g., decoding (Figure 1, Box A). The addition of word forms to the students' vocabulary contributes to the growth of automatic word recognition, an important factor in decoding speed and accuracy (Ouellette, 2006; Paris & Hamilton, 2014). Vocabulary depth is particularly important in the processes (see Figure 1, Box B) that facilitate the creation of situation models (Oakhill et al., 2015). The organisation and expansion of a rich semantic system, i.e., vocabulary depth, allow for the quick and effective access to word meanings (Nation & Snowling, 1999). This then frees up enough cognitive load for the higher-level comprehension processes and deeper comprehension (Stahl, 1991).

The links between vocabulary and text comprehension are dynamic and multifaceted (Baumann, 2014) showing strong correlations between .6 to .8 from study to study (see Pearson et al., 2007). On the one hand, a larger vocabulary has been found to improve comprehension. A richer vocabulary supports inference-making by enabling the rapid activation of word meanings and the related concepts (Cain & Oakhill, 2014). A large vocabulary base facilitates access to additional information and thus the ability to create conceptual relationships (Oakhill & Cain, 2007). On the other hand, comprehension, especially inference-making, contributes to the extension of vocabulary (Verhoeven et al., 2011). When readers make inferences from a context, they ascribe meanings to unknown words and thereby expand their vocabularies (Silva & Cain, 2015). This expansion is easier and more precise if there is strong overlap between the information in the text and the meaning of a new word. Vocabulary building is also facilitated when the text contains only a few unknown words that are presented in a variety of contexts (Stahl, 1991).

Although vocabulary is important in text comprehension, it is not sufficient for ensuring understanding (Cain et al., 2004; Spencer et al., 2019). Successful text comprehension also requires the ability to discern the relationships between the elements in the text or sentence to create text-based representations that are integrated with the reader's pre-knowledge to build a situation model (Kintsch, 1998). This can be achieved by making several kinds of inferences (van den Broek, 1994). Inference-making is the ability to connect the information from or within the text with the readers' pre-knowledge (MacNamara & Kendeou, 2011) to fill gaps and to understand the implicit information in the text. Given the large repertoire of inferences, they can be generally categorised by dimension: *automatic* and *controlled/strategic*, the time: *during/on-line* and *after/off-line* the reading (Kintsch & Rawson, 2005), or the nature of it: *necessary/coherence/bridging* and *elaborative/interrogative* (Cain, 2010; Davoudi & Moghadam, 2015; Kintsch, 1998).

Automatic inferences, which are made quickly, consume a low cognitive load, whereas controlled inferences are highly resource-demanding (Kintsch & Rawson, 2005). Bridging inferences rely mostly on context, thereby ensuring a text-based representation or literal understanding of the text (McNamara & Magliano, 2009). In most instances, bridging inferences are automatic and are made during reading (Tennent, 2015). Elaborative inferences go beyond the given information. Text information is connected with pre-knowledge (Cook et al., 2001) for the creation of a situation model (Kintsch, 1998). These inferences, which enrich (Cain, 2010) or expand the mental representation of the text, are made in a controlled or strategic manner after reading (Tennent, 2015).

Gaining an overview of students' text comprehension proficiency, the processes that are involved (e.g., bridging or elaborative inferences), and the products that are generated (e.g., text-based representation, situation model) should be examined. These processes and products can be transferred to the three comprehension levels: literal, inferential, and evaluative. Literal comprehension is associated with the text-based representation that is achieved when readers understand the explicit information in texts. Bridging inferences connect textual information from multiple sentences (Kintsch, 1998). An example is "Mary is thirsty. She needs a glass of water." The bridging inference is that "she" in the second sentence refers to "Mary" in the first (Currie & Cain, 2015).

The inferential level is related to the ability to create a situation model by interpreting implicit meanings through elaborative inferences (Basabara et al., 2013). The reader's pre-knowledge and understanding of the existing relationships among the objects, characters, and events in the text is essential for the creation of the relevant inferences (Alonzo et al., 2009). The following sentence provides an example: "We wanted to go to Mexico on our holiday, but my wife could only take a vacation in July." This sentence would not make sense without the application of the pre-knowledge about the extreme heat in Mexico at that time (Hirsch, 2003).

The evaluative level goes somewhat further. It involves reasoning beyond the text (Kintsch & Rawson, 2005) by applying divergent thinking. Readers respond

cognitively and affectively to texts by making critical judgements on the basis of their own values and experiences and those of the characters in the story (Magliano et al., 1999). At this comprehension level, the previously created situation model is expanded. For example, a task to justify or to critique the characters' motives on the basis of common ethics would refer to the evaluative comprehension level. It should be noted that some evaluative questions reflect the inferences that readers might not normally construct (Magliano et al., 2007).

2.3 Individual differences in text comprehension

The interactions of the many lower- and higher-level components and processes, such as vocabulary, inference making, comprehension strategies, and conscious and controlled activities, determine comprehension quality (Angosto et al., 2013; Kintsch, 1998). Considerable individual differences can exist in the quality of these processes, the proficiency with which they are executed, and, thus, the comprehension product (Oakhill & Cain, 2007; van den Broek & Espin, 2012). The complexity of text comprehension defies categorisations such as good or poor comprehenders.

Generally, readers with poor lower-level (e.g., decoding) skills have limited higher-level (e.g., inference-making and comprehension strategy) skills (Cain & Oakhill, 2014). Low and inaccurate or non-automatic lower-level processes consume a great deal of the cognitive resources. This in turn reduces the cognitive processing capacity for the construction and integration phases, which are necessary for the creation of text-based representations and situation models (Kintsch, 1998; Perfetti, 2007). However, good lower-level skills do not necessarily ensure adequate text comprehension. Some students have good decoding skills and adequate vocabulary but limited proficiency in creating situation models (Cain et al., 2004; Spencer et al., 2019). One reason is the lack of a rich semantic system, i.e., the vocabulary depth for ascribing the appropriate meanings to words (Caemmerera et al., 2018; Perfetti, 2007). Higher-level skill deficits (when efficient lower-level skills are available) could also stem from the lack of knowledge of when and how to make necessary inferences, as well as the impaired use of various comprehension strategies (van den Broek & Espin, 2012). Specifically, the ability to make bridging and elaborative inferences and to monitor comprehension progress has consistently been found to facilitate the differentiations between poor and skilled readers (Oakhill & Cain, 2011; Perfetti, 2007).

Despite having limited lower-level skills, which should inhibit overall comprehension, some students may comprehend texts at higher levels (VanderVeen et al., 2007). One of the ways in which lower-level skills deficits can be neutralised is through pre-knowledge about the topic (Stahl, 1991). Most inferences are endorsed by pre-knowledge (Kintsch & Kintsch, 2005); thus, excellent pre-knowledge serves as a compensatory tool to improve the comprehension product (Hirsch, 2003).

Individual differences in text comprehension are closely related to the development of comprehension processes. There are many similarities in younger and older readers' comprehension processes. However, the differences in the quality and the depth of comprehension and the related processes in the various age groups are worth noting (Kendeou et al., 2007). For example, younger readers may make the same inferences as older readers; however, their inference-making processes are not so automatic and require more cognitive resources. Thus, the inferences tend to be more concrete and text-based (Cain & Oakhill, 2006; Currie & Cain, 2015). Over time, the quantity and the quality of inferences change. Students become more capable of generating the inferences that connect larger units of texts rather than those that relate solely to individual facts and events. Older students are also better able to connect abstract associations (van den Broek et al., 2005).

The differences in students' comprehension components and processes form certain text comprehension patterns or profiles. On the basis of these profiles, students can be placed into comprehender subgroups. Belonging to a comprehension subgroup tends to be persistent or even cumulative. Problems in specific aspects of comprehension tend to be ongoing (Uibu & Männamaa, 2014). Moreover, comprehension development can be characterised by the expression "the rich get richer." For example, the good comprehenders in the first grade know twice as many words as the poor comprehenders. By the end of the 12th grade, their vocabulary is four times that of the poor comprehenders (see Hirsch, 2003). This is related to the interactions of the text comprehension components and processes. Adequate lower- and higher-level comprehension processes form the basis for the acquisition of new words from the text, thereby expanding vocabulary (Cain et al., 2004).

2.4 Text comprehension assessments

Text comprehension is assessed in a variety of contexts for different purposes. For example, scientific assessments can be conducted to confirm or modify a theory or to design and implement reading interventions (Magliano et al., 2007). Scientific text comprehension assessments are mostly theory-based and influenced by the CI comprehension model (Kintsch, 1998), which incorporates the surface, text base, and situation models to explain the multidimensional understanding of texts (McNamara & Magliano, 2009). In educational settings, the critical reasons for measuring text comprehension are to monitor student progress, to detect possible text comprehension weaknesses, and to provide instructional feedback (Magliano et al., 2007). The assessments that are developed in school settings tend to be unidimensional. Thus, they are inadequate for determining the possible deficiencies in the essential skills or processes outlined in the text comprehension theories (Keenan et al., 2008). Accordingly, these tests are not instructionally relevant because of the lack of valuable feedback on the skills that should be targeted for improving students' text comprehension (VanderVeen et al., 2007).

One possibility for increasing the relevance of the information provided by text comprehension assessments is the differentiation of the tasks on the basis of the levels of cognitive processes that readers apply. The three-level taxonomy is widely used. It provides the information about students proficiency at the following levels: *literal* comprehension, i.e., the creation of text-based representations through the recall of explicit statements in the text; *inferential* comprehension, i.e., the interpretation of the author's meaning through the connection of implicitly stated information to create the situation model; and *evaluative* comprehension, i.e., the ability to go beyond the text to enrich the situation model (Alonzo et al., 2009; Basaraba et al., 2013; Tennent, 2015). Analogous dimensions have been used in several international comparison surveys, such as the National Assessment of Educational Progress (NAEP) and the Programme for International Student Assessment (PISA; NAEP, 2019b; OECD, 2009). Specifically, literal comprehension is similar to *information location or recall* in the NAEP and *access and retrieval* in PISA. The inferential comprehension level is similar to *integration and interpretation* in the PISA and NAEP. The evaluative level is comparable to *critique/evaluate* in the NAEP and *reflect and evaluate* in the PISA.

To gain a better understanding of student performance at the text comprehension levels, several types of tasks have been used, for example free-recall, short-answer, multiple-choice, problem-solving, and cloze tasks (Kikerpill & Türk, 2013; Kintsch & Rawson, 2005). Depending on the range of the text that is targeted in the task, multiple-choice tasks can provide information about proficiency in creating text-based representations (Keenan et al., 2008) and situation models (VanderVeen et al., 2007). To assess students' deeper understanding of the text, they can be asked to use the information from a text to solve a problem, or they can be asked to summarise the text with a short or long answer (McNamara & Kendeou, 2011).

Text comprehension is a developmental process with considerable variations between students in Grades 1 through 6; this must be considered in test design (Morsy et al., 2010). The assessment of lower-level skills, such as decoding and vocabulary breadth, is relevant for younger students; however, it is ineffective for detecting the poor comprehenders among the older students (Magliano et al., 2007). In older students, text comprehension is more closely related to higher-level processes and skills, e.g., vocabulary depth and inferences and strategy use (Oakhill & Cain, 2007; VanderVeen et al., 2007). Hence, the comprehension tests for older students should focus not on the measurement of text-based representations but, rather, the processes that contribute to the creation of situation models. In addition, the texts used in the assessments should be representative of those encountered by the age group in regular reading situations. Examples are more narrative texts for younger students and larger repertoires with different text types for older students. Thus, the texts and the comprehension levels and processes measured with tests should reflect the developmental stage of the targeted students (Magliano et al., 2007).

Vocabulary as an essential component should be included in comprehension tests. Vocabulary can be assessed in a contextual or decontextual manner, e.g.,

word lists (Pearson et al., 2007). Measuring vocabulary in a contextual manner is more common in the assessments of vocabulary depth (Ouellette, 2006). However, the results of contextual tests could be limited by the strategic processes that are necessary for deriving word meanings from context (Pearson et al., 2007). Although decontextualized single-word tests are used mostly in in the assessment of vocabulary breadth, they can be designed to reflect vocabulary depth. For example, word labelling, decoding, or recognition tasks are appropriate for measuring vocabulary breadth. The understanding of synonyms, antonyms, homonyms, figurative language, and oral definitions or multiple meanings of words is indicative of vocabulary depth (Ouellette, 2006).

Many countries, including Estonia, assess text comprehension with national standard-determining tests (Garbe et al., 2016; Riigi Teataja, 2018; Tengberg, 2017; Vestheim & Lyngsnes, 2016). In Estonia, students' accomplishments in the text comprehension is measured together with other language skills, such as grammar and writing skills (Innove, 2014) at the end of every school stage: Grades 3, 6, and 9 (Riigi Teataja, 2018). The Grade 3 and 6 national standard-determining tests, which are considered low-stakes tests, are not evaluative. They are designed to map student achievement and to provide teachers with the information to develop instructional practices for Estonian-language lessons (Haridus- ja Teadusministeerium, 2018; Innove, 2014). The national test for Grade 9 is considered a high-stakes test that determines students' learning opportunities (Pajupuu, 2007). Regardless of whether the tests are low- or high-stakes, the comprehension tasks should be in accordance with the students' development and the national curriculum. Thus, the Grade 3 tests should include more tasks that assess students' vocabulary and text-based representation processes. The focus in the Grade 6 and 9 tests should shift more to the higher-level processes, such as interpreting pre-knowledge to understand implicit information or to enrich situation models by critically evaluating the text (Estonian Government, 2011/2014; Magliano et al., 2007; Oakhill et al., 2015).

Assessments can yield useful information for teachers only if the measured skills, processes, and components are clarified (Kendeou & Papadopoulus, 2012). As there is no complete conception for assessing text comprehension in Estonia, there is a lack of clarity about the components and processes that are assessed by the comprehension tasks in the national standard-determining tests. Whether the tests are in accordance with the students' development is also unclear. Hence, the national standard-determining tests in Estonia give teachers little, if any, information about student progress and proficiency in the comprehension processes over time.

2.5 Text comprehension strategies

The cognitive view of text comprehension highlights the influence of the interactively working lower- and higher-level processes on the comprehension product (McNamara & Kendeou, 2011). Many of these processes rely on the various

comprehension strategies (Dole et al., 1991; van den Broek & Kremer, 2000) that enable readers to control, to monitor, and to foster the text comprehension processes (Kong, 2019). In general, a comprehension strategy is a set of conscious, intentional, and purposeful activities undertaken during the reading process to improve the comprehension product (Afflerbach et al., 2008; Graesser, 2007). To be more specific, the cognitive and metacognitive strategies are differentiated in a plethora of comprehension strategies.

Cognitive comprehension strategies can be specified as the mental or behavioural activities to increase the quality of text comprehension (Van den Broek & Kremer, 2000). Proficient readers implement effortless and time-consuming strategies, such as looking back at the preceding text to get relevant information, visualising the read information, and integrating pre-knowledge with new information, to better comprehend texts (Afflerbach & Cho, 2014; van den Broek & Espin, 2012). Cognitive comprehension strategies are especially important in the case of breakdowns at any level of text comprehension because they help readers to repair a comprehension failure (Peterson et al., 2001). For example, in the case of an unknown word in the text, the strategic cognitive activities could be to determine the meaning on the basis of the context or to find the definition in a dictionary, to reread the sentence, and to comprehend it as a whole (Graesser, 2007).

The understanding of the breakdowns requires the coordination and implementation of cognitive and metacognitive strategies (Afflerbach & Cho, 2014). These strategies can be defined as the self-monitoring and regulating activities that focus on the text comprehension process and product (Van den Broek & Kremer, 2000). Metacognitive strategies are particularly important in the detection of a breakdown and the selection of a relevant cognitive comprehension strategy to facilitate the adjustment of the reading process or product (Dole et al., 2014). Comprehension monitoring is one of the proficient readers' most essential metacognitive strategies. It involves the active control of text comprehension: adjustments of the reading speed in accordance with the difficulty level of the text, evaluations of what is or is not understood, and the efficiency of the applied strategies (Tennent, 2015).

Because different comprehension strategies are useful in various reading situations they should be used flexibly, according to the text and the processes that need to be fostered (Dole et al., 2014). Students should be able to choose, to apply, and to evaluate the effectiveness of text comprehension strategies for facilitating deeper comprehension (Duke & Pearson, 2009; Graesser, 2007). Although students might know different comprehension strategies, they do not always know when and how to apply them (Soodla et al., 2017). Hence, teaching multiple text comprehension strategies in reading lessons would be beneficial for improving reading outcomes.

A considerable number of strategies to support text comprehension at various cognitive levels have been proposed by researchers; however, they cannot all be included in a single intervention (Duke & Pearson, 2009; NRP, 2000). The focus of the selection criteria in the present doctoral study was on the strategies that

support students' metacognitive awareness and text comprehension at the inferential and evaluative levels. On the basis of previous studies (e.g., Boulware-Gooden et al., 2007; NRP, 2000; van Keer & Verhaeghe, 2005), the following six comprehension strategies were chosen: *skim reading*, *vocabulary building*, *monitoring*, *generating and answering questions*, *identifying the main idea*, and *summarising.*

2.6 Teaching text comprehension strategies

Over the years, several instructional programs have been designed and implemented for the purposeful teaching of comprehension strategies (see Slavin et al., 2008). Some programs are easier to implement because they involve just one strategy (Trabasso & Bouchard, 2002). Others are more complex with multiple comprehension strategies that have been implemented through various instructional methods, such as peer-tutoring, coaching, group work, and reciprocal teaching (NRP, 2000; Tennent, 2015; van Keer & Verhaeghe, 2005). Despite the differences in the instructional methods and comprehension strategies, most of the programs follow, to some extent, the Gradual Release of Responsibility model with explicit teaching, guided practice, and independent practice (Pearson & Gallagher, 1983). This model can result in students' knowing how, when, and why comprehension strategies should be used so that students can use them independently (Duke & Pearson, 2009). In this model, explicit teaching includes modelling through thinking aloud and explaining the mental reasoning involved in the strategy. It provides sufficient scaffolding for students to learn a new comprehension strategy through the visibility of the teacher's expert thinking (Peterson et al., 2001). Guided practice involves learning the strategies with classmates and benefiting from the teacher's and mates' feedback and adaptions if needed. The mastery of the strategies requires sufficient guided practice followed by independent practice. Over time, the teacher's support gradually decreases as the students learn to use the strategy (Dole et al., 1991).

The explicit teaching of text comprehension strategies has been found to be effective for enhancing comprehension (Duke & Pearson, 2009; NRP, 2000). The benefits have been debated. Some studies have found that it enhances proficient readers' comprehension (Van Keer & Verhaeghe, 2005). Others have found that skilful readers' comprehension is improved by their ability to use their preferred strategies (Griffith, & Ruan, 2005). The consensus on poor comprehenders seems to be that text comprehension strategies must be taught (Elleman, 2017; Van Keer & Verhaeghe, 2005). It seems that the effects of an intervention depend also on the comprehension strategies that are used and the students who are taught. For example, better readers have found to gain more from metacognitive strategy instruction (Griffith, & Ruan, 2005). Poor comprehenders benefit from almost any strategy, especially when it is explicitly taught (Applegate et al., 2006).

The teaching of multiple comprehension strategies has been found to be more effective for older students. It is related to the cognitive load that is associated with strategy learning. Learning a new strategy can be a strain. It places a burden on the cognitive resources by demanding the reader's full attention for successful execution (Afflerbach & Cho, 2014). If lower-level skills, such as decoding speed and accuracy, are not automatised, the learning and execution of strategic processes might be compromised (van den Broek & Espin, 2012). However, when proficiency has been achieved in the lower-level skills, younger students improve their text comprehension after being taught multiple strategies, e.g., making predictions, using pre-knowledge, and sequencing story events (Eliers & Pinkley, 2006).

3. RESEARCH METHODS

3.1 Research design

The focus of this doctoral study was threefold: (1) to provide a comprehensive view of how text comprehension is measured in the Estonian national standard-determining tests; (2) to develop text comprehension and vocabulary tests, to increase the understanding of the associations among vocabulary and text comprehension levels, and to examine students' proficiency in these components; and (3) to determine effective ways for enhancing students' vocabulary and text comprehension at various levels. The analysis of the tasks and the students' text comprehension skills was based on the three-level taxonomy: literal, inferential, and evaluative comprehension (see Tennent, 2015). Qualitative and quantitative methods were used in this study. Qualitative analysis was used to provide an overview of how comprehension is measured in the national standard-determining tests and to categorise the comprehension tasks by text comprehension level. With regards to the quantitative methods, a variable-oriented approach was implemented to generate group-level comparisons (Muijs, 2004), and a person-oriented approach was used to determine the patterns within the groups (Bergman et al., 2003).

The first aim of the study was achieved in two stages: analysis of the comprehension tasks in the national standard-determining tests (Articles I, II, and III) and the design of new vocabulary test (VT) and text comprehension test (TCT) (Article IV). First, the comprehension tasks in the national standard-determining tests were analysed to understand how the multidimensionality of text comprehension is considered in the Estonian-language national standard-determining tests. National standard-determining tests have been found to play an essential role in the teaching of text comprehension. Therefore, the tests should be theory-based, and the students' development should be considered. The new comprehension test was compiled with consideration for the multidimensionality of text comprehension (Kintsch, 1998; Magliano et al., 2007). In addition, it was guided by the three-level taxonomy: literal, inferential, and evaluative comprehension (Tennent, 2015). Because of the important influence of vocabulary in text comprehension, a VT was also designed.

To achieve the second aim of the study, individual differences in vocabulary and text comprehension were investigated (Articles IV and V). Text comprehension involves several interactively working components that could develop at different timepoints and thus contribute to the comprehension process differently (Cain & Oakhill, 2006). Furthermore, effect of the instructional interventions could be influenced by the students' abilities (NRP, 2000; van Keer & Verhaeghe, 2005). A thorough understanding of students' individual differences and developmental trajectories is essential to improving text comprehension.

The third aim of the study was the design and implementation of an intervention to enhance students' text comprehension at different levels (Article V). The intervention developed for this doctoral study employed a quasi-experimental

design. The groups were not randomly selected. Instead, intact classes were selected to be either the control or experimental group. The quasi-experimental design was chosen because unlike laboratory experiments, it enables the implementation of an intervention in authentic conditions (Cohen et al., 2007): in this case, in whole class settings led by teachers. It is important that the effectiveness of educational interventions be evaluated in school settings because these programs are eventually meant to be classroom tools for teachers.

The developed intervention considered the different text comprehension levels (Basaraba et al., 2013; Kintsch, 1988, 1998), the role of vocabulary (Currie & Cain, 2015), and the strategies that enhance vocabulary and text comprehension (Boulware-Gooden et al., 2007; van Keer & Verhaeghe, 2005). A pre-test–post-test control–experimental group design was used. Information from Article IV about the students' vocabulary and text comprehension skills guided the intervention and the assignment of the students to the control and experimental groups. The students in the experimental group were taught multiple text comprehension strategies during a three-month period. Those in the control group continued their regular reading lessons without a focus on comprehension strategies. The students in both groups were pre- and post-tested with the same VTs and TCTs. The pre-tests provided baseline data on the students' vocabulary and text comprehension levels, and the post-tests measured the changes. Table 1 provides an overview of the methods in the articles.

Table 1. Overview of the methodology related to the research questions

Research questions	Data collection timeline	Samples	Instruments	Data analyses	Articles
RQ1: How are text comprehension levels considered in the national standard-determining tests?	2013–2015 2013–2016	*Sample I* 67 tasks Grade 6 *Sample II* 226 tasks 78: Grade 3 67: Grade 6 81: Grade 9	National Estonian-language standard-determining tests	Content Analysis Descriptive Statistics	I, II, III
RQ2: To what extent does the distribution of text comprehension levels in the national standard-determining tests and examinations change by grade?	2013–2016	*Sample II* 226 tasks 78: Grade 3 67: Grade 6 81: Grade 9	National Estonian-language standard-determining tests	Content Analysis Descriptive Statistics	II, III
RQ3: How can the text comprehension levels be distinguished, and what is their association with vocabulary in the developed vocabulary and text comprehension tests?	2017	*Sample III* 301 students Grade 4	Text comprehension test Vocabulary test	Confirmatory Factor Analysis (CFA) Structural Equation Modelling (SEM)	IV
RQ4: What are the possible subgroups of students with individual differences in vocabulary and text comprehension at literal, inferential, and evaluative levels?	2017 2019	*Sample III* 301 students Grade 4 *Sample IV* 257 students: 153: experimental group 104: control group Grade 6	Text comprehension test Vocabulary test	Latent Profile Analysis (LPA)	IV, V
RQ5: How does the teaching of text comprehension strategies improve students' vocabulary and text comprehension at each level? **RQ6:** To what extent does the teaching of text comprehension strategies influence vocabulary and text comprehension in various subgroups?	2019	*Sample IV* 257 students 153: experimental group 104: control group Grade 6	Text comprehension test Vocabulary test	Repeated Measures MANOVA	V

3.2 Samples

Text comprehension tasks

Samples I and II consisted of text comprehension tasks from the national Estonian-language standard-determining tests (see Table 1). The tasks for all the grades were based on various types of texts, e.g., fictional, informational, and scientific. In Sample I, the 2013–2015 Grade 6 text comprehension tasks were analysed to determine the types of tasks and levels of comprehension. The sample consisted of 67 tasks: 21 in the 2013 and 2015 tests and 25 in 2014. Sample II comprised the 2013–2016 tests for Grades 3, 6, and 9. Altogether, 226 comprehension tasks were analysed: 78 comprehension tasks for Grade 3, 67 for Grade 6, and 81 for Grade 9. The Sample II tasks were compared to determine the similarities in the tests for the same grade over time. To examine whether the students' development was considered in the test design, the changes in the tests were analysed. The Grade 3 and 6 tasks included oral and written texts. The Grade 9 tasks involved written texts only.

Students

Samples III and IV involved students from different Estonian schools: municipal school students who studied in regular classes according to the Estonian National Curriculum for Basic Schools (Estonian Government, 2011/2014). Only students who had received parental approval were included in the study. Sample III consisted of Grade 4 students from 12 Estonian schools (N = 301; 53.5% girls and 46.5% boys). The average age was 10.66 years (SD = .50). Sample IV comprised Grade 6 students from 10 Estonian schools. Altogether, 257 students (57.6% girls and 42.4% boys) participated (see Table 1). The average age was 12.3 years (SD = .47). The students were assigned to experimental and control groups. The experimental group comprised 153 students: 56.9% girls and 43.1% boys. The control group had 104 students: 58.7% girls and 41.3% boys.

3.3 Instruments

No specific instrument was used to gather data for Articles I, II, and III. The comprehension tasks in the 2013–2016 Grade 3, 6, and 9 national Estonian-language standard-determining tests were chosen for analysis. For the analysis of Articles IV and V, new VT and multidimensional TCT were developed for Grades 4 and 6. The same VT and TCT were used to measure the Grade 6 students' skills before and after the intervention period in Grade 6 (Article V).

The design of the TCT for Grade 4 (Article IV) and Grade 6 (Article V) was based on the three-level taxonomy (Basaraba et al., 2013; OECD, 2009), the analysis of the Estonian-language national standard-determining tests (Kärbla et al., 2017, 2018, 2019), and the Estonian National Curriculum for Basic Schools

requirements (Estonian Government, 2011/2014). The Grade 4 TCT (Article IV) comprised multiple tasks ($N = 30$), such as picture–sentence matching, multiple-choice, open-ended, picture ordering, and closed items, which were based on different types of reading and listening texts (i.e., fictional and informational). The Grade 4 TCT was adapted to create an age-appropriate test for Grade 6 (Article V). Some relatively easy texts and tasks were excluded, and more complex ones were included. The number of tasks in the Grade 6 TCT was 31. The proportions of literal, inferential, and evaluative tasks were 13, 10, and 7, respectively, for Grade 4, and 13, 9, and 9, respectively, for Grade 6. The students' answers were coded dichotomously (right: 1; wrong or unanswered: 0), and the total scores from the text comprehension levels were used in the analysis. Examples of tasks for each text comprehension level are presented in Table 2.

Table 2. Examples of tasks in the developed vocabulary and text comprehension tests (a modification of Article IV)
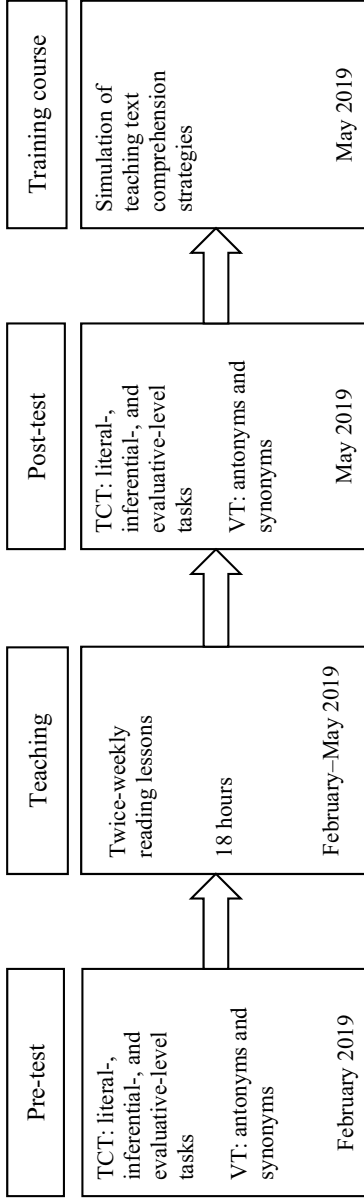
TEXT COMPREHENSION TEST

Comprehension levels	Examples	Response formats	Additional information
Literal level	Complete the table with information from the text. (Grades 4, 6)	Short-answer task Students had to read the text and complete the table with information about the nature and number of lessons presented by the character every day.	Information was explicitly stated in the text. Students had to recognise the subjects mentioned in the text and to understand when and how many of them the character had each day.
Inferential level	Order the pictures on the basis of the stories heard. (Grade 4)	Ordering task Students were asked to order six pictures related to the listening text.	Students had to understand the activity represented by the pictures, which they were required to place in the right order according to heard text.
Evaluative level	Summarise the story with a proverb. Justify your choice of proverb. (Grade 6)	Open-ended task	Students had to go beyond the text to give meaning to the text with a proverb. They had to thoroughly understand the text and to apply their pre-knowledge of proverbs.

VOCABULARY TEST

Parts of speech	Examples	Multiple choice tasks	Additional information
Noun (Antonym)	Dialogue (Grade 6)	a) chatting b) communication c) monologue d) diagram (pronunciation similarity in Estonian)	Students had to choose the response option that best represented the listed word.
Verb (Synonym)	Shine (Grades 4, 6)	a) ray b) glow c) hurry (pronunciation similarity in Estonian) d) beam	
Adjectives (Antonym)	Foggy (Grade 6)	a) clear b) indoor (pronunciation similarity in Estonia) c) unclear d) light	

The design of the Grade 4 and 6 VTs was informed by of the Estonian frequency dictionary (Kaalep & Muischnek, 2002) and the words used in the primary school textbooks (Kitsnik & Metslang, 2011). The Grade 4 VT contained 35 synonyms, and the Grade 6 VT included 38 synonyms and 38 antonyms. The students were presented a list of words (e.g., nouns, verbs, and adjectives in equal percentages), and they had to choose the correct answer from four alternatives: (a) pronunciation similarity, (b) substantive similarity (the correct answer in the case of a synonym), (c) situational similarity, and (d) the opposite word (the correct answer in the case of an antonym). Examples of the tasks are presented in Table 2. The students' answers to the VTs were coded dichotomously, and the total vocabulary score was used in the analysis.

3.4 Description of the intervention

Article V describes the design and implementation of the pre-test–post-test control–experimental intervention. The students were assigned to the experimental and control groups on the basis of the previous results of the VT and TCT (Article IV). The experimental group participated in the intervention twice-weekly 45-minute sessions for 12 weeks (18 hours total). The control group continued their usually twice-weekly reading classes, which lacked a focus on teaching comprehension strategies (see Figure 2).

Control group

Pre-test		Teaching		Post-test		Training course
TCT: literal-, inferential-, and evaluative-level tasks		Twice-weekly reading lessons		TCT: literal-, inferential-, and evaluative-level tasks		Simulation of teaching text comprehension strategies
VT: antonyms and synonyms		18 hours		VT: antonyms and synonyms		
February 2019		February–May 2019		May 2019		May 2019

Experimental group

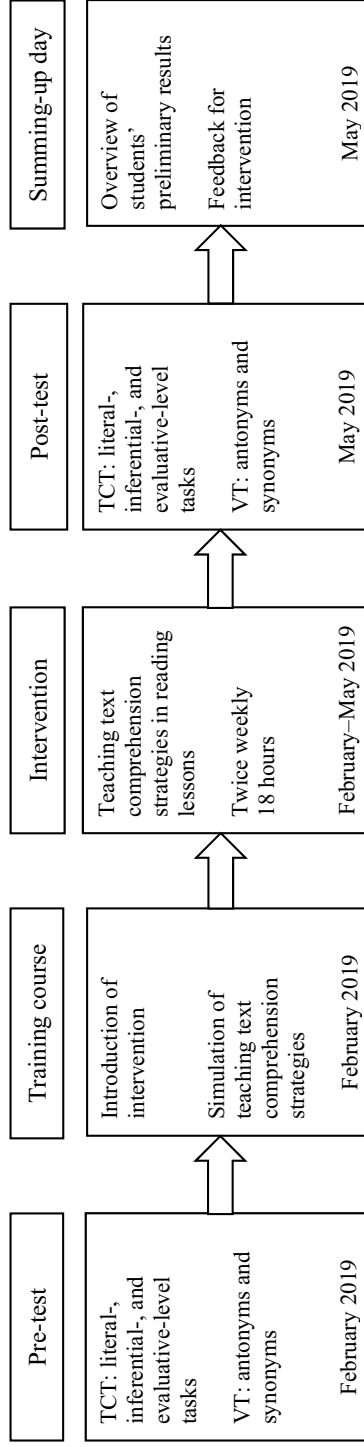| Pre-test		Training course		Intervention		Post-test		Summing-up day
TCT: literal-, inferential-, and evaluative-level tasks		Introduction of intervention		Teaching text comprehension strategies in reading lessons		TCT: literal-, inferential-, and evaluative-level tasks		Overview of students' preliminary results
VT: antonyms and synonyms		Simulation of teaching text comprehension strategies		Twice weekly 18 hours		VT: antonyms and synonyms		Feedback for intervention
February 2019		February 2019		February–May 2019		May 2019		May 2019

Figure 2. Intervention design

31

The intervention included two major sections: *explicitly teaching* comprehension strategies and *practicing the strategies* with different texts. The explicit teaching followed the Gradual Release of Responsibility model with three phases (Pearson & Gallagher, 1983): modelling, co-practicing, and independent practice. In the modelling phase, the teachers used the think-aloud method to explain and to model the strategy in normal reading situations. In the co-practicing phase, the teachers instructed the students to apply the strategy in whole-class, group, or pair work settings. During the independent practice phase, the students used the strategies individually. In the second section, practicing the strategies, the skills were transferred to other texts. The students had to implement learned strategies to comprehend three new texts with different complexity levels.

Six comprehension strategies were chosen for the intervention program. Except for the last two (identifying the main idea and summarising), each strategy was taught separately. For strategy learning, each student was given a strategy card that contained the information about the strategy and instructions for its application (see Appendix 1). An overview of the approach to the teaching of each strategy is presented below.

The *skim reading strategy* included the previewing and scanning of the text by reading it and finding information about the characters, actions, time, and place. For the determination of the difficulty level of the text, the students were required to detect and to underline unfamiliar words. This strategy is useful in adapting the reading according to the difficulty of texts thus promoting students' metacognitive skills (Samuels et al., 2005).

Vocabulary building involved finding the meanings of unfamiliar words by relying on the context; using the dictionary to search for different meanings, synonyms, and antonyms; and using these words in context. The students had to justify their word selections on the basis of the context and to explain the effects of word choices on the meaning of the text. Teaching this strategy increases the students' ability to find the meanings to unknown words, promotes a deeper understanding of the words, and improves the students' conceptual understanding of vocabulary (Boulware-Gooden et al., 2007). It also enables them to comprehend the text at higher cognitive levels (NRP, 2000).

With the *monitoring strategy*, the students were encouraged to stop after each paragraph and to use self-questioning to consider their understanding of the text. They also had to use their own sign systems to detect the important information and the surprising or interesting and confusing or incomprehensible parts of the text. If there were breakdowns in understanding, the students were taught to locate (position in the text) the problem, to determine its nature, and to rephrase the incomprehensible section of the text. They re-read the text or read forward in search of helpful information. Finally, they expressed and justified their solutions. Teaching the monitoring strategy can improve the self-regulation skills needed to understand texts at every comprehension level (Griffith & Ruan, 2005).

Forming and answering questions began with introducing questions that reflected literal, inferential, and evaluative comprehension. After the introduction, the students were given a worksheet with questions that had to be categorised at

the literal, inferential, and evaluative levels. Finally, the students created their own inferential and evaluative questions and discussed the quality of the questions and answers with classmates. Promoting the students' knowledge of question types, especially deeper questions, can improve their text comprehension skills at each comprehension level.

Identifying the main idea and *summarising* were taught together because of their interrelationship: one is a prerequisite for the other. The teaching of these strategies was divided into two parts: working with paragraphs and working with the whole text. For each paragraph, the students were asked to underline the main character, the most important action, the information that implied the time and location of the action, and information that provided clues about the reason for the activity. The students then summarised the important information in two sentences, which they wrote next to the paragraph. Finally, they wrote headlines for each paragraph.

Working with the whole text comprised the identification of the main idea and the summary of the whole text. The students were encouraged to read through the conclusions and headlines they had written for each paragraph, to find the repetitions, to detect the interrelationships and commonalities among the paragraphs, and to generate the main message of the text. In addition, they had to describe the text with a proverb and to explain the relationship between their decision about the main idea and the important information in the text. Teaching these strategies supports the creation of an organised, relevant, and coherent memory representation of what has been read (NRP, 2000). It promotes greater interaction with the text and, thus, increases precision in the creation of situation models (Griffith & Ruan, 2005).

3.5 Reliability and validity of the tests and the intervention

To achieve content validity in the VTs and TCTs, discussions were held with several experts (e.g., teacher educators, psychologists, speech therapists, and in-service teachers). To ensure construct validity, the subject-related theoretical material was analysed, and previous studies were considered (Sullivan, 2011). The VTs and TCTs were piloted in Grades 4, 5, and 6 (total of 52 students). The relatively similar or easy items (item difficulty: 90–100%) were excluded from the final tests. Some questions in the pilot version of the TCT were open-ended, and the typical errors informed the formulation of distractors for the multiple-choice questions in the final test. The selection of response options for each item was analysed, and only the items with good distractors (i.e., every response option was chosen by at least one student) were included in the final test. In addition, item response theory (IRT), which enables the estimation of the difficulty level of the items regardless of the individual's abilities (Dimitrov, 2014), was used to distinguish and to exclude the items with low separation reliability.

The students' answers in both tests were coded dichotomously: right (1) or wrong or unanswered (0). To increase the credibility of the study, two researchers

independently coded the open-ended answers. In cases of disagreement, they discussed their decisions until consensus was achieved. To confirm the reliability of the VTs and TCTs, i.e., the internal consistency of the test items, Cronbach's alpha coefficients were calculated. For the Grade 4 VT, the Cronbach's alpha was 0.91, and for the TCT, the internal consistency for the literal, inferential, and evaluative levels was 0.82, 0.72, and 0.82, respectively (Article IV). For the Grade 6 VT, the internal consistency was .97, and for the TCT, the literal, inferential, and evaluative levels yielded Cronbach's alpha values of .81, .67, and .71, respectively (Article V).

To increase the validity of the experiment, the intervention procedure and tasks were piloted, and the experts were involved in the training course. On the basis of the comments and suggestions from the experts and the teachers in the pilot study and training course, the intervention program was adjusted. Furthermore, internal validity (Cohen et al., 2007) was enhanced by the use of approved instruments and control and experimental groups. However, validity is undermined if the intervention is not faithfully implemented, i.e., the intervention is not executed as intended (Gresham, 2009).

To limit the variability in treatment fidelity and to increase reliability, the teachers in the experimental group received one full-day training course before the intervention (see Figure 2). The course involved the introduction of the intervention and a simulation of the implementation of the lesson scenarios and the explicit teaching of the comprehension strategies in the classroom. Furthermore, all the teachers in the experimental group were provided with elaborated lesson scenarios for teaching comprehension strategies, conclusive lesson plans for practicing the strategies with different texts, and materials for the students (see Appendices 1 and 2). The lesson scenarios included precise descriptions of the classroom activities during which the strategies could be explicitly taught. Conclusive lesson plans contained a summary of activities that needed to be completed while working with texts. The student materials included strategy cards that contained the aim and information on how and when to use the strategy. The teachers were given manuals, in printed and web-based formats, with the instructions and necessary materials for implementing the intervention. The intervention was continually monitored, and teachers were supported by the researchers via a web-based platform, e-mail, and telephone. After the intervention, the teachers participated in the summing-up day. They reflected on the experience, received feedback, and were given the students' preliminary test results.

3.6 Data collection

Text comprehension tasks

For Articles I, II, and III the comprehension tasks in the national standard-determining tests designed by the education competence centre Innove, the foundation that coordinates and promotes general and vocational education in

Estonia (About Innove, n.d.), were analysed. The 2013–2016 Grade 3, 6, and 9 national Estonian-language standard-determining tests were downloaded from the Innove website. All the analysed tests were in a printed format. Because the 2016 Grade 6 test was a pilot in an electronic format, it was excluded from the analysis in this study.

Students

The data for Articles IV and V were collected through the VTs and TCTs that were developed specifically for this doctoral study. The Article IV data were gathered from the students at the end of Grade 4. For the Grade 6 intervention (Article V), the Article IV results were used to assign the students to the control and experimental groups. During approximately the same period, the students in the experimental and control groups completed the adapted VT and TCT before and after the intervention period. The teachers administered the Grade 4 and 6 assessments in the classrooms. Printed tests with instructions were mailed to the schools. The students completed two tests, the VT and TCT, during Estonian-language lessons (2×45 minutes).

3.7 Ethical benchmarks in the study

The participants in a scientific study should be aware of the aim and content of the work, and they should be assured of the confidentiality of their data (Eesti Teaduste Akadeemia, 2002; Kline, 2011). In addition, they must agree to the study's conditions and give their consent to participate in it (Hammersley & Traianou, 2012). For this doctoral study, the participants voluntarily provided the data, and they received explanations about the research aim, content, and confidentiality policy. Information about the study was sent to the schools, and only the schools whose teachers had agreed to participate were involved. Next, the parents were informed about the study aim and the students' role. Parental consent was obtained, and the data of the students whose parents did not provide consent were excluded from the analyses.

The intervention (Article V) included experimental and control groups. The teachers for both groups were provided information about the purpose of their participation. The teachers who led the experimental group knew that the program was part of an intervention. Those in the control group were informed that they were in the control condition. After the intervention period, the teachers in the control group were asked to participate in the training course on teaching comprehension strategies, and they were provided with all the necessary materials.

The data collected from this doctoral study is in the possession of the University of Tartu. To ensure the participants' confidentiality, the data were encoded, and no personal information was shared.

3.8 Data analysis

3.8.1 Content analysis of text comprehension tasks

Qualitative and quantitative analyses of the text comprehension tasks in Articles I, II, and III were performed. The qualitative phase consisted of content analysis (Ryan & Bernard, 2003) to identify implicit and explicit information within the data (Guest et al., 2012). The text comprehension tasks were categorised on the basis of the three-level taxonomy: literal, inferential, and evaluative comprehension (see Basaraba et al., 2013; NAEP, 2019b; OECD, 2009). An example of the creation of the categories is presented in Figure 3.
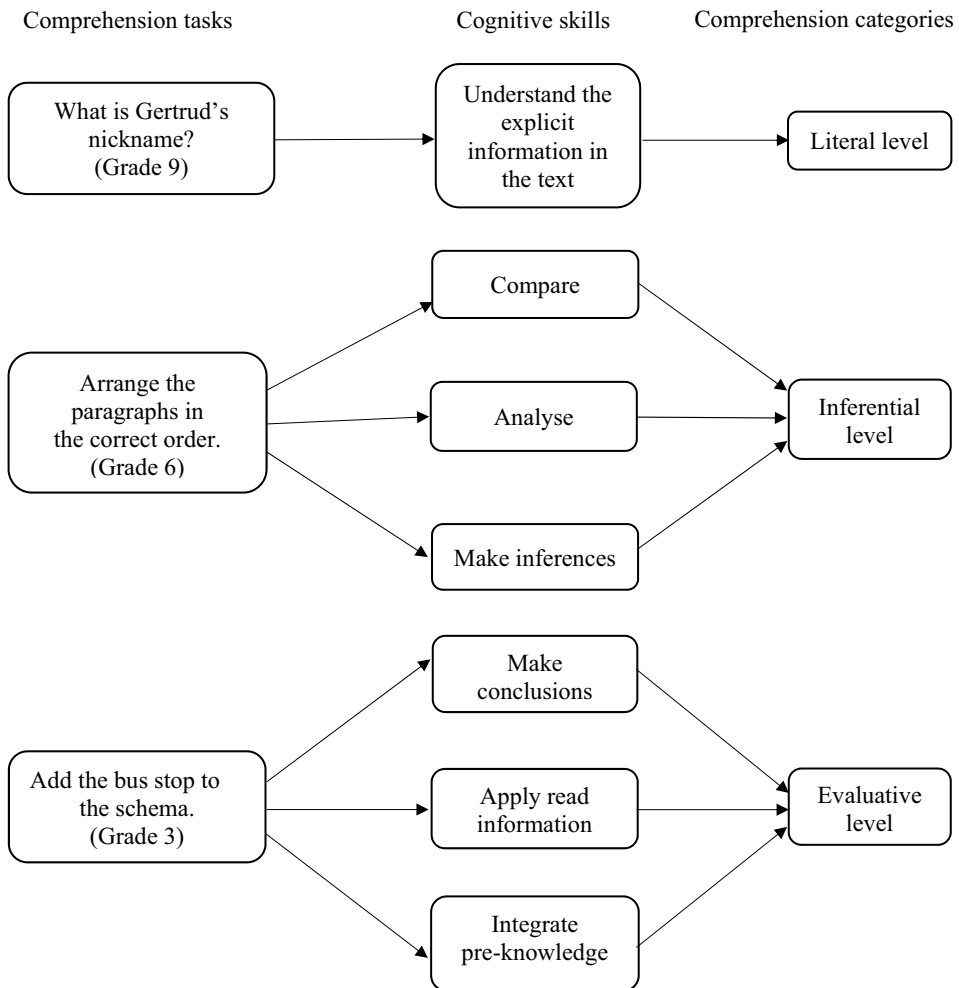
Comprehension tasks   Cognitive skills   Comprehension categories

What is Gertrud's nickname? (Grade 9)	→ Understand the explicit information in the text → Literal level
Arrange the paragraphs in the correct order. (Grade 6)	→ Compare / Analyse / Make inferences → Inferential level
Add the bus stop to the schema. (Grade 3)	→ Make conclusions / Apply read information / Integrate pre-knowledge → Evaluative level

Figure 3. Example of the creation of comprehension categories (a modified version from Article III)

To ensure the reliability of the three-level categorisation of the test items, all the Grade 6 tasks in the tests were read and categorised independently by two researchers. After the first categorisation, the researchers compared their results. In the case of a disagreement, they justified their decisions. As a result of discussions and the revision of the three-level text comprehension taxonomy, a consensus was achieved, and the items were reconsidered. All the tasks from the Grade 3 and 9 tests were categorised by the author of this doctoral study. Randomly selected tasks were categorised by two researchers and their agreements were evaluated. No discrepancies were found. In addition, the author of this study categorised all tasks twice. In case of discrepancies between the first and second categorising, a discussion between the two researchers led to the final decisions.

3.8.2 Group-level analysis of vocabulary and text comprehension

Quantitative methods were used to analyse text comprehension tasks in Articles I, II, and III and students' vocabulary and text comprehension in Articles IV and V. In the quantitative analysis of the comprehension tasks, the total scores for each comprehension category were calculated to examine the theoretical underpinnings among the text comprehension tasks in the national standard-determining tests. The students' vocabulary and text comprehension were analysed through variable- and person-oriented methods (Bergman et al., 2003; Muijs, 2004) in Mplus version 8.1 software (Muthén & Muthén, 1998–2012) and IBM SPSS Statistics, Version 25.0.

Group-level analyses were conducted to assess the newly developed tests and to compare the students' vocabulary and text comprehension results. To confirm the construct validity of the newly developed tests, the three-factorial model for the TCT and the one-factorial model for the VT were examined with confirmatory factor analysis (CFA). In the text comprehension model, each item was allowed to load on one factor. Model appropriateness was evaluated with the chi-square (χ^2) test, the comparative fit index (CFI), the Tucker–Lewis index (TLI), and the root mean square error of approximation (RMSEA). The data fit of a model was considered good if the CFI and TLI were ≥.95 (Hu & Bentler, 1999), and a value of .90 indicated an adequate fit (Kline, 2011). For the RMSEA, the acceptable values could be as high as .08 (Hair et al., 2010). Structural equation modelling (SEM) was used to estimate the relationship between vocabulary and the comprehension levels (Article IV).

For Article V, the experimental and control group results were compared with one-way multivariate analysis of variance (MANOVA). The repeated measures MANOVA was used to analyse the changes in the experimental and control groups' vocabulary and text comprehension pre- and post-test results. In addition to p-values, Cohen's proposed guidelines for interpreting the effect size of partial η^2 were used: 0.01, small effect; 0.06, moderate effect; and 0.14, large effect (Cohen et al., 2007).

3.8.3 Individual-level analysis of vocabulary and text comprehension

The exclusive use of group-level analysis could lead to superficial conclusions because samples are rarely homogeneous. Text comprehension comprises multiple lower- and higher-level processes and components; thus, there could be variations in student performance in several text comprehension skills. The study sample could comprise multiple groups of students with similar text comprehension patterns. To develop a thorough understanding of the sample, individual-level analyses that reveal unobserved or hidden patterns within cases were employed (Bergman et al., 2003; Williams & Kibowski, 2016).

Latent profile analysis (LPA) as person-oriented mixture modelling was used (Williams & Kibowski, 2016) for Articles IV and V to distinguish the student subgroups by the differences in vocabulary and text comprehension levels. First, because of the difference in the number of vocabulary and comprehension items, Z-scores were calculated for the total vocabulary and text comprehension scores in order to achieve comparability. Next, the models with various numbers of latent subgroups were fitted. The appropriate number of subgroups was evaluated by using three criteria: model fit, distinguishability of the latent groups, and usability and interpretability of the latent subgroups. On the basis of the results of the LPA, the students were grouped by the Z-scores for vocabulary and text comprehension at the literal, inferential, and evaluative levels.

4. FINDINGS

The results of this doctoral study are presented in three sections in accordance with the aims of the study. The first section contains the results for the analyses of the text comprehension tasks in the national standard-determining tests and the VT and multidimensional TCT that were developed specifically for this study. The second section focuses on the differences in vocabulary and text comprehension at the literal, inferential, and evaluative levels. The third section includes the results of the instructional intervention that was developed and implemented to enhance students' vocabulary and comprehension at each level. Table 3 provides an overview of the most important findings for each research question.

Table 3. Overview of the main research results for the research questions

Research questions	Articles	Data analyses	Main results
RQ1: How are the text comprehension levels considered in the national standard-determining tests?	I, II, III	Content Analysis Descriptive Statistics	**Characteristics of text comprehension tasks in the national standard-determining tests:** 1) Overreliance on literal level tasks in the Grade 3, 6, and 9 tests; 2) Very few evaluative tasks in the Grade 3, 6, and 9 tests; 3) Lack of variability, i.e., mostly multiple-choice tasks, in the Grade 6 tests; 4) Lack of consistency in the tests for the same grade in different years (2013–2016).
RQ2: To what extent does the distribution of text comprehension levels in the national standard-determining tests and examinations change by grade?	II, III	Content Analysis Descriptive Statistics	**Cognitive development considerations in national standard-determining tests:** 1) No consideration in the 2013, 2014, and 2016 tests: more evaluative and fewer literal level tasks in the Grade 3 tests than in the Grade 6 and 9 tests; 2) Consideration in only the 2015 Grade 3 and 9 tests.
RQ3: How can the text comprehension levels be distinguished, and what is their association with vocabulary in the developed vocabulary and comprehension tests?	IV	Confirmatory Factor Analysis (CFA) Structural Equation Modelling (SEM)	**Characteristics of the vocabulary and text comprehension tests:** 1) Literal, inferential, and evaluative comprehension levels were distinguished in the text comprehension test; 2) Single-factor model for vocabulary test was established; 3) Associations between vocabulary and text comprehension were found at every level, the strongest being between vocabulary and inferential text comprehension, and the weakest being between vocabulary and evaluative text comprehension.

Table 3. (continued)

Research questions	Articles	Data analyses	Main results
RQ4: What are the possible subgroups of students with individual differences in vocabulary and text comprehension at literal, inferential and evaluative levels?	IV, V	Latent Profile Analysis (LPA)	**Individual differences in vocabulary and text comprehension:** 1) No proficient comprehenders in either grade; 2) Most of the students in both grades categorised in the subgroup with average vocabulary and text comprehension scores; 3) Small groups with variations in vocabulary and text comprehension.
RQ5: How does the teaching of text comprehension strategies improve students' vocabulary and text comprehension?	V	Repeated Measures Multivariate Analysis of Variance (MANOVA)	**Changes in vocabulary and text comprehension:** 1) Improvements in experimental groups' vocabulary and literal, inferential, and evaluative comprehension after three months; 2) Improvements in only literal comprehension in control group.
RQ6: To what extent does the teaching of text comprehension strategies influence the vocabulary and text comprehension in the student subgroups?	V	Repeated Measures MANOVA	**Changes in vocabulary and text comprehension by subgroup:** **Experimental group** 1) Improvements in literal, inferential, and evaluative comprehension in students with low average vocabulary and text comprehension; 2) Improvements in literal comprehension in students with high average vocabulary and text comprehension; 3) Improvements in vocabulary and literal comprehension in students with low scores. **Control group** 1) Improvements in literal comprehension in students with low average vocabulary and text comprehension scores.

The following sections detail the important findings regarding the study aims.

4.1 Text comprehension tests

4.1.1 Text comprehension tasks in national standard-determining tests

The text comprehension tasks in the 2013–2016 Grade 3, 6, and 9 national standard-determining tests were analysed to determine the theoretical foundation of tests (Articles I, II, and III). The text comprehension tasks were categorised on the basis of the comprehension levels in the three-level taxonomy. The focus of the tasks in most of the tests tended to be on the evaluation of literal comprehension, i.e., the understanding of explicitly stated information. Two tests for Grade 6 and one for Grade 9 did not include evaluative comprehension tasks. In addition, the similarities in the task distribution for the three comprehension levels in the same-grade tests for different years were examined (Articles II and III). In consecutive years, there seemed to be shifts in the targeted comprehension levels. For example, in the 2015 Grade 9 tests, 20% of the tasks were evaluative; however, there were no such tasks the following year. In the Grade 3 test, 31.8% of the tasks in 2013 and 63.6% in 2014 were at the literal level. More information is provided in Figure 4.
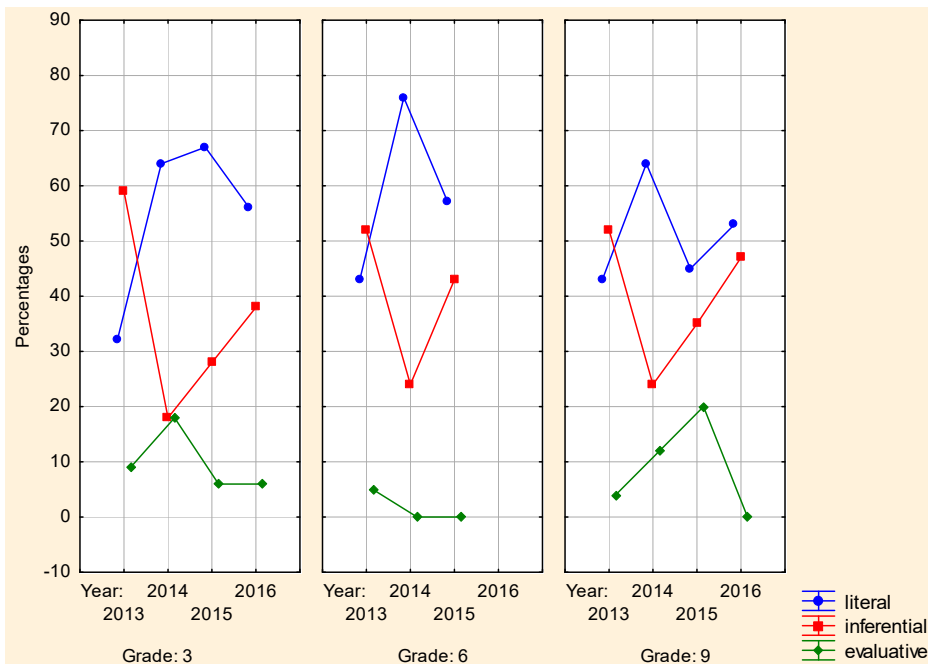

Figure 4. Distribution of text comprehension levels in Grade 3, 6, and 9 tests (from Article II)

Next, the changes in the task distribution at the various text comprehension levels in the tests for the various grades were examined (Articles II and III). The analysis revealed that in most of the years under review, the changes in task distribution

were not in accordance with the students' cognitive development. Thus, the Grade 3 tests generally contained more inferential and evaluative tasks and fewer literal comprehension tasks than the Grade 6 and 9 tests. Only the 2015 Grade 3 and 9 tests considered the students' development. Thus, the Grade 3 tests contained more literal and fewer evaluative tasks than the Grade 9 tests.

The task types in the Grade 6 national standard-determining tests in Article I were examined. Text comprehension was found to be assessed mostly with multiple-choice tasks (e.g., 80% of the 2014 test) or gap-fill tasks (e.g., the remainder of the 2014 tasks). Only the 2013 test included ordering and short-answer tasks. The types of tasks that measured Grade 6 students' text comprehension are presented in Figure 5.
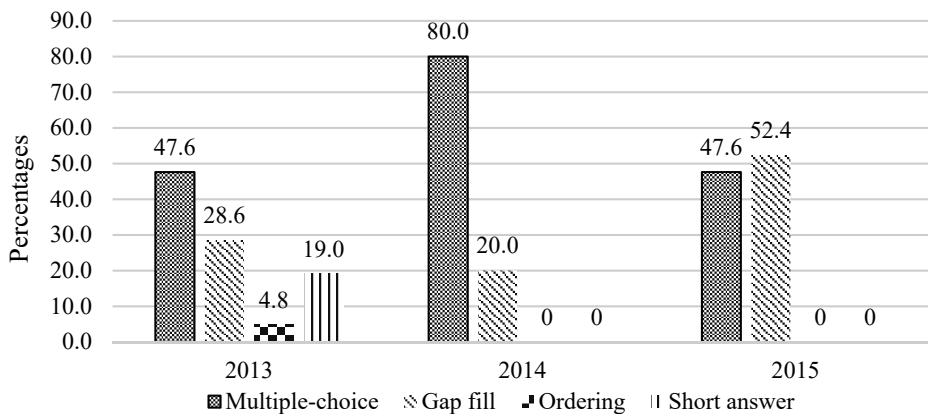

Figure 5. The tasks in the Grade 6 national standard-determining tests (from Article I)

To conclude, the text comprehension tasks in the Estonian-language national standard-determining tests were developed without consideration of the contemporary comprehension theories that have defined the levels of understanding of a text (Kintsch, 1998; McNamara & Kendeou, 2011). The students' text comprehension development was not considered. Therefore, there is need for a theoretically relevant and age-appropriate TCT with a task variety.

4.1.2 Vocabulary and text comprehension tests

To gain more in-depth information about the students' text comprehension, a VT and a TCT with three comprehension levels were developed (see also Basaraba et al., 2013; OECD, 2009; Tennent, 2015). The VT was used to examine the associations between vocabulary and text comprehension at the literal, inferential, and evaluative levels. Confirmatory factor analysis (CFA) was used to control the factor structures of the VT and TCT. The fit of the three-factorial model for the TCTs and the one-factorial model for the VTs was evaluated. The fit indices for the final TCTs were acceptable and very good for the VTs (see Table 4).

Table 4. Fit indices for vocabulary and text comprehension tests

Tests	χ^2	CFI	TLI	RMSEA
TCT for Grade 4	465.980	.94	.93	.02
VT for Grade 4	599.382	.98	.97	.02
TCT for Grade 6	518.608	.96	.96	.02
VT for Grade 6	3245.920	.98	.98	.02

Note: TCT = text comprehension test; VT = vocabulary test; χ^2 = chi-square test of model fit; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation

In Article IV, the associations among the students' vocabulary and text comprehension levels were analysed with structural equation modelling (SEM). The model fit indices for the SEM model were very good: CFI = .96, TLI = .96, RMSEA = .01. There were significant associations between vocabulary and all text comprehension levels, the strongest being between vocabulary and inferential and literal comprehension and the weakest being between vocabulary and evaluative comprehension (see Figure 6).
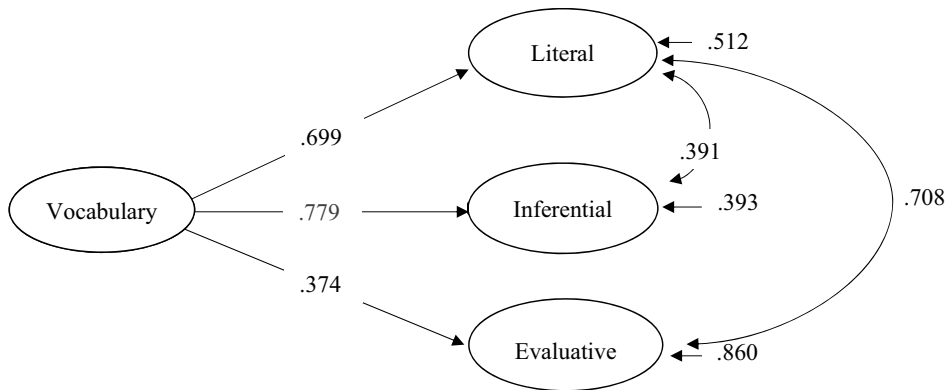

Figure 6. Model of vocabulary and text comprehension levels (from Article IV)

Note: The double-headed arrows represent the noncausal associations between the factors. The nonsignificant path coefficients (inferential–evaluative) have been excluded.
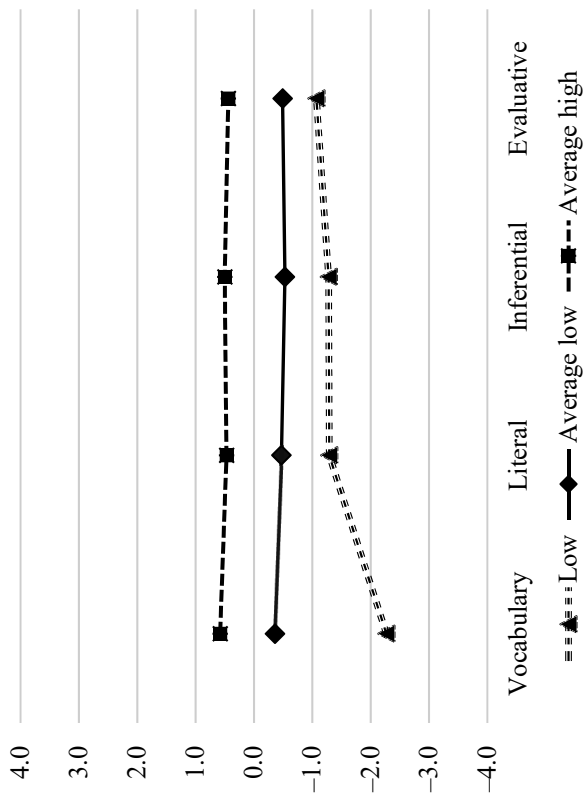
The analysis revealed that high vocabulary scores were predictive of higher comprehension scores: inferential ($\beta = .53$, $p < .001$), literal ($\beta = .45$, $p < .001$), and evaluative ($\beta = .28$, $p < .001$).

4.2 Subgroups of students with differences in vocabulary and text comprehension

In Articles IV and V, the VTs and TCTs were used to examine the individual differences in the Grade 4 and 6 students' vocabulary and text comprehension. Prior to the application of LPA, the possible outliers were removed. Some similarities were observed in the LPA results for Grades 4 and 6. Specifically, three subgroups with differences in vocabulary and literal, inferential, and evaluative text comprehension levels were identified in both grades. The largest subgroups in both grades were identified by the average vocabulary and text comprehension scores. A small subgroup with variations in vocabulary and text comprehension were found in both grades. These students had very poor results in all the measured skills, especially the lower-level skills. No subgroups with high vocabulary or text comprehension scores were found in Grades 4 or 6. More information about the Grade 4 and 6 subgroups is presented in Figures 7 and 8.
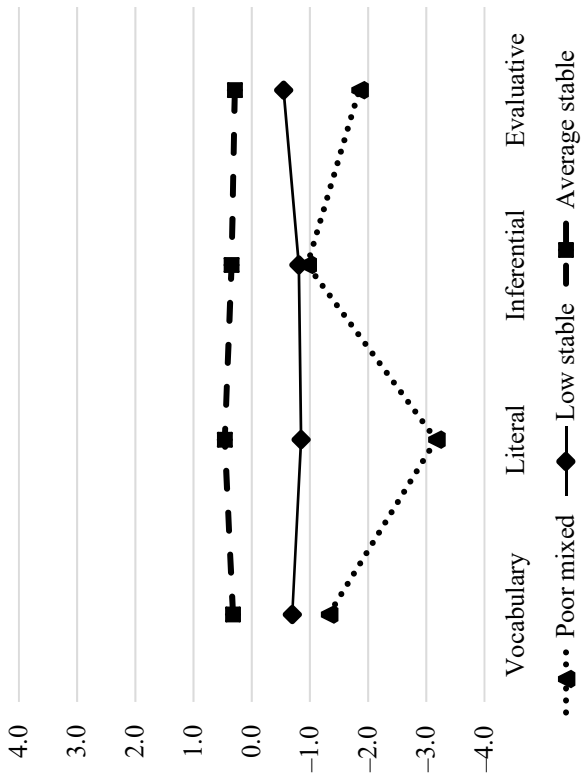
Figure 7 shows that two of the three Grade 4 subgroups were characterised by stable vocabulary and text comprehension scores. Thus, within the subgroup, the achievement levels for the measured skills were similar (Article IV). The largest Grade 4 subgroup (70.4%) was referred to as *Average Stable* because the students had consistently average vocabulary and text comprehension scores. The other stable Grade 4 subgroup (26.2%), *Low Stable*, had low vocabulary and text comprehension scores. The smallest Grade 4 subgroup (3.3%), *Poor Mixed,* was characterised by very low vocabulary and inferential and evaluative text comprehension scores and extremely low literal comprehension scores.

Figure 8 indicates that the two subgroups with average and stable scores were found in Grade 6 (Article V). The largest subgroup, *Average High* (60.0%), had high average vocabulary and text comprehension scores. The second Grade 6 subgroup, *Average Low* (30.0%), had low average vocabulary and text comprehension scores. The smallest Grade 6 subgroup, *Low* (10.0%), exhibited low scores at all the text comprehension levels and extremely low vocabulary scores.

Grade 4

Figure 7. Standardised vocabulary and literal, inferential, and evaluative text comprehension scores for the three Grade 4 subgroups (a modified version from Article IV)

Grade 6

Figure 8. Standardised vocabulary and literal, inferential, and evaluative text comprehension scores for the three Grade 6 subgroups (a modified version from Article V)

4.3 Changes in vocabulary and text comprehension

The third aim of the doctoral study was to develop an intervention to improve students' text comprehension and to assess its effectiveness in enhancing vocabulary and text comprehension (Article V). First, to control for the similarities in the starting points of the experimental and control groups, the students' pre-test results were compared by one-way MANOVA. The analysis revealed that there was no difference in the pre-test vocabulary and literal and inferential text comprehension scores of the experimental and the control groups. However, differences were found in the evaluative comprehension scores. The students in the control group had statistically better results than the students from the experimental group ($p = .03$).

To examine the changes in the students' vocabulary and comprehension after three months, the pre- and post-test data were compared through repeated measures MANOVA. The students who participated in the intervention program exhibited significantly improved vocabulary and literal, inferential, and evaluative comprehension ($p = .002$, $p < .001$, $p < .001$, $p = .003$, respectively). The students who attended the regular reading lessons without focused attention on teaching comprehension strategies exhibited only enhanced literal comprehension ($p = .01$).

To gain a greater understanding of the changes in vocabulary and text comprehension of students with different abilities, the results for the subgroups were analysed. Table 5 provides an overview of the changes in vocabulary and text comprehension.

Table 5. Changes in vocabulary and text comprehension in the subgroups (from Article V)

Sub-groups	Skills	Pre-test		Post-test		F	p	Partial η^2
		M	SD	M	SD			
Low	*Control group (n = 8.7%)*							
	Literal	8.78	3.1	9.00	2.6	.04	.85	.01
	Inferential	3.78	1.3	4.22	1.5	1.00	.35	.11
	Evaluative	3.56	2.5	4.11	2.3	.61	.46	.07
	Vocabulary	25.33	6.0	30.78	13.9	1.76	.22	.18
	Experimental group (n =11.1%)							
	Literal	8.53	2.2	9.76	2.0	4.46	.05*	.22
	Inferential	3.29	.9	3.53	1.7	.35	.56	.02
	Evaluative	3.18	2.0	3.47	1.9	.50	.49	.03
	Vocabulary	21.35	5.9	32.35	15.0	11.33	.004*	.42

Table 5. (continued)

Sub-groups	Skills	Pre-test		Post-test		F	p	Partial η²
		M	SD	M	SD			
Average low	*Control group (n = 31.7%)*							
	Literal	10.58	1.5	11.52	1.3	9.72	.004*	.23
	Inferential	4.82	1.4	5.15	1.9	.82	.37	.03
	Evaluative	4.73	1.4	5.00	2.1	.56	.46	.02
	Vocabulary	48.12	6.9	49.42	11.4	.67	.42	.02
	Experimental group (n = 28.8%)							
	Literal	9.64	1.7	10.39	1.9	7.08	.01*	.14
	Inferential	4.66	1.4	5.61	1.8	10.03	.003*	.19
	Evaluative	4.05	1.8	4.91	2.0	8.44	.01*	.16
	Vocabulary	47.16	7.3	48.75	9.8	1.55	.22	.04
Average high	*Control group (n = 59.6%)*							
	Literal	11.94	.96	12.13	1.2	1.69	.20	.03
	Inferential	6.66	1.5	6.56	1.4	.22	.64	.00
	Evaluative	6.42	1.5	6.32	1.6	.16	.69	.00
	Vocabulary	61.74	4.9	62.73	5.9	1.48	.23	.02
	Experimental group (n = 60.1%)							
	Literal	11.87	1.1	12.21	.9	5.51	.02*	.06
	Inferential	6.72	1.3	7.00	1.4	2.83	.09	.03
	Evaluative	5.97	1.2	6.21	1.4	2.17	.14	.02
	Vocabulary	59.82	5.6	60.90	8.0	2.00	.16	.02

Note: % in parenthesis: subgroup distribution of control and experimental group students; *significance at $p < .05$.

The analysis revealed that the explicit teaching of text comprehension strategies was effective for all students. The intervention was most beneficial for the students who had low average vocabulary and text comprehension scores (*Average Low*). Their post-test scores were significantly higher than their pre-test literal, inferential, and evaluative comprehension scores (p = .01; .003 and .01, respectively). The intervention was slightly less effective for the students with low vocabulary and text comprehension scores (*Low*). They statistically improved their vocabulary and literal comprehension (p = .004 and .05, respectively). The intervention was least beneficial for the students who were characterised as having high average vocabulary and text comprehension scores (*Average High*). Specifically, they exhibited statistically significant changes in literal text comprehension only (p = .02).

The improvements in vocabulary and text comprehension in the absence of the explicit teaching of comprehension strategies were minimal. Only the control group students with low average vocabulary and text comprehension scores (*Average Low*) exhibited significant improvements in literal comprehension ($p = .004$).

5. DISCUSSION

Text comprehension involves the interactions of lower- and higher-level components and processes to form text-based representations and situation models from read information (Kintsch, 1998; McNamara & Magliano, 2009). To effectively promote text comprehension, it is important to have a comprehensive view of the students' progress in the various aspects of text comprehension. This doctoral study is based on the three-level taxonomy: literal, inferential, and evaluative comprehension. Thus, the discussion is undergirded by the taxonomy. The discussion chapter comprises four sections. The first section focuses on the instruments that assess students' text comprehension. The second addresses individual differences in comprehension. The third section reviews the text comprehension intervention. The final section discusses the strengths and limitations of the study.

5.1 Text comprehension assessment

5.1.1 Text comprehension levels in the national standard-determining tests

The text comprehension tasks in the 2013–2016 Grade 3, 6, and 9 national standard-determining tests were examined to determine the extent to which the multidimensionality of text comprehension was considered. The analysis of the literal, inferential, and evaluative text comprehension tasks revealed that the focus tended to be on the assessment of literal comprehension (Articles I, II, and III). These national standard-determining tests can provide a comprehensive view of students' ability to form text-based representations; however, information about inferential and evaluative comprehension was insufficient or even unavailable. These findings confirm those of previous studies. Specifically, reading tests tend to measure how well students can understand the facts in a text but not necessarily how well they actually comprehend texts (Keenan et al., 2008; Sabatini et al., 2013).

The reason for the predominance of literal-level tasks might be the relative ease with which they can be designed. In addition, the evaluation of answers to factual questions is straightforward. The creation of tasks to address the complexity of text comprehension requires the test developers to have a nuanced understanding of the subject (Alonzo et al., 2009). According to the cognitive view of text comprehension, the understanding of texts at multiple levels is the result of the interactions of lower- and higher-level processes (Kintsch, 1998; McNamara & Kendeou, 2011). However, the predominance of tasks that assess the ability to create text-based representations could lead teachers to conclude that students with adequate literal comprehension skills are proficient comprehenders. Teachers could therefore miss deficiencies in the inferential and evaluative comprehension skills required for the creation and expansion of situation models (Basaraba et al.,

2013; Kintsch, 1998). To detect possible deficiencies and to provide teachers with adequate feedback to design reading instruction, the national standard-determining tests should include a balanced range of literal, inferential, and evaluative comprehension tasks (VanderVeen et al., 2007).

The distribution of literal, inferential, and evaluative comprehension tasks for the same grade in different years was analysed to determine the consistency of the tests (Articles II and III). The task distribution in the national standard-determining tests appeared to be random. For example, in the years under study, 18.2% to 59.1% of the Grade 3 tasks tested inferential comprehension, and 42.9% to 76.0% of the Grade 6 tasks focused on literal comprehension. The reason for this variability within age groups could be the lack of a framework for assessing text comprehension in Estonia. A comprehensive framework with the necessary information, as it is compiled for assessing text comprehension with several international assessments, and science subjects in Estonia (NAEP, 2019b; OECD, 2009; Pedaste et al., 2017), would direct the designing of text comprehension tasks in the national standard-determining tests. This would facilitate the development of tests that continuously measure the different comprehension components and processes with age-appropriate tasks.

The national standard-determining tests are designed to evaluate learning outcomes and to provide teachers with the relevant information for incorporating the appropriate instructional practices in their Estonian-language classes (Haridus- ja Teadusministeerium, 2018). However, the dissimilarities in the tests within age groups would provide inaccurate information about the appropriate focus of reading classes (Vestheim & Lyngsnes, 2016). For example, on the basis of the text comprehension tasks in the national standard-determining tests, teachers could be encouraged to emphasise inferential comprehension during one year but literal comprehension during the next. Moreover, the variability in the distribution of the components measured in tests for the same grade would affect the year-over-year comparisons of student performance. Such comparisons are important for making relevant conclusions about students' development in text comprehension and developing instructional practices for reading classes (Cutting & Scarborough, 2006).

The changes in the distribution of the text comprehension levels that were assessed in different grades were also analysed (Articles II and III). The distribution of literal, inferential, and evaluative comprehension tasks was inconsistent and not necessarily in accordance with the students' development. Specifically, in most years, the tests for the younger students (i.e., Grade 3) included more inferential and evaluative tasks than the tests for the older students (Grades 6 and 9). This is not in accordance with developmental progression in text comprehension: namely, the ability to make more precise inferences and to critically evaluate texts increases over time (Kendeou et al., 2014). Although younger students are capable of inferential and evaluative comprehension, the quality and quantity of their inferences are considerably different from those of older students. First, older students' lower-level processes (e.g., decoding and text-based representations) are sufficiently automatised so that more resources can be available for

higher-level processes (Kibui, 2012). Second, they have more pre-knowledge, which facilitates inference-making, and reading experience, which allows for the application of multiple strategies to achieve deeper comprehension. Accordingly, older students' comprehension tends to be characterised by the analysis and interpretation of implicit information to create enriched situation models (Oakhill & Cain, 2007). Assessments of these students should contain a greater percentage of inferential and evaluative tasks. On the positive side, the national standard-determining tests for Grade 3 included literal, inferential, and evaluative tasks. It is crucial that every comprehension level be included in assessments of younger students so that their progress can be monitored. For example, the students' drastic decrease in text comprehension, referred to as late-emergent comprehension disability, may be due to the absence of higher-level comprehension tasks in the tests for younger students (Keenan, 2016).

The results also indicated that the Grade 6 students were assessed with tasks that mostly provided response options and targeted concrete aspects of the text (Article I). This is consistent with previous studies that found that multiple-choice tasks are often preferred because they are familiar to students and easy to administer (Keenan, 2016; Morsy et al., 2010). However, variations in task types could provide more comprehensive information about students' text comprehension. For example, multiple-choice, ordering, cloze, and open-ended tasks require different levels of proficiency in students' skills and cognitive processes (Kikerpill & Türk, 2013; Pearson & Hamm, 2005). Accordingly, the inclusion of a variety of tasks in assessments could lead to variability in student outcomes and facilitate the collection of in-depth information about students' text comprehension (Applegate et al., 2002). An over-reliance on multiple-choice tasks in national standard-determining tests would provide information about lower-level skills, especially when the targeted information is limited and the answer is supported by the similarities between the correct response option and the information in the text (NAEP, 2019b; Ozuru et al., 2013). The tasks that do not have one obvious answer and those in which the students need to create their own responses (e.g., open-ended and short-answer tasks) better reflect the students' ability to apply the components and processes that are needed to create situation models of the read information (McNamara & Kendeou, 2011).

In conclusion, for the effective teaching of text comprehension, tests should be designed to assess the students' proficiency in the components and processes necessary for the understanding of texts at different levels (Kendeou et al., 2007) bearing in mind the students' age. Equally important is the sharing of the information regarding the comprehension components that are measured by the test (VanderVeen et al., 2007). The analysed national standard-determining tests did not include information on the comprehension aspects that were measured nor were the tests based on text comprehension theories. Therefore, the tests could not fulfil their purpose of providing teachers with information for developing instructional practices in Estonian-language classes.

5.1.2 Vocabulary and text comprehension tests

The VTs and TCTs that were developed for this doctoral study could determine vocabulary and literal, inferential, and evaluative comprehension proficiency and deficiencies (Articles IV and V). The TCT contained items for assessing the ability to understand explicitly stated information, to create text-based representations, and the proficiency in interpreting implicit meanings to create and to expand situation models (Kintsch, 1998). The VT reliably measured vocabulary depth, which is essential for creating situation models (Oakhill et al., 2015), i.e., in the inferential and evaluative processes that are important predictors of older students' text comprehension proficiency (Floyd et al., 2012). Accordingly, these tests could enable teachers to monitor student progress, to detect possible deficiencies in the various text comprehension components, and, thus, to provide instructionally relevant feedback (Magliano et al., 2007).

Next, the associations between vocabulary depth and the three text comprehension levels were analysed (Article IV). The findings confirmed those of previous studies. The strongest association was between vocabulary and inferential text comprehension (Cain & Oakhill, 2014; Currie & Cain, 2015). A richer vocabulary facilitates the ability to draw more inferences because it allows for the faster selection of the appropriate meanings and enables the recognition of relationships between concepts (Cain & Oakhill, 2006; Calvo, 2005; Oakhill et al., 2015). An essential association between vocabulary and literal comprehension was also found. Although, it has been claimed that vocabulary breadth is more important for creating text-based representations (Ouellette, 2006), the current study found that vocabulary depth also played an important role in this process. A reason could be that literal comprehension requires the establishment of links between multiple propositions in texts in order to create text-based representations (Kintsch, 1998). To make these links without straining the cognitive resources, a broad knowledge of word meanings is needed to facilitate the quick selection of context-appropriate meanings (Cain & Oakhill, 2014; Perfetti et al., 2008).

A small but still statistically important association was additionally found between vocabulary and evaluative comprehension. The inferences that are necessary for evaluative understanding are constructed after the text has been read (Tennent, 2015) and are influenced by pre-knowledge and metacognitive skills more than by vocabulary (Duke & Carlisle, 2011). However, literal and inferential comprehension is a prerequisite to the critical evaluation of texts (Basaraba et al., 2013; Veeravagu et al., 2010). The quicker and more automatised the processes at the literal and inferential levels, the lower is the targeted cognitive load. Thus, enough resources are available to expand and enrich situation models at the evaluative level (Stahl, 1991). As vocabulary depth contributes to the automatization of lower-level processes, it may be that the evaluative comprehension is related to vocabulary indirectly through the literal and inferential levels. Thus, the significant relationships between vocabulary and literal, inferential, and evaluative comprehension highlight the importance of increasing students'

vocabulary by teaching them comprehension strategies that improve their ability to find the meanings for unknown words in texts.

5.2 Individual differences in vocabulary and text comprehension

There are differences in readers' implementations of text comprehension processes (van den Broek & Espin, 2012). It is therefore essential that individual differences be examined to enhance text comprehension. The analysis of the individual differences in vocabulary and text comprehension revealed three subgroups of Grade 4 and 6 students (Articles IV and V). As expected, most of the students in both grades belonged to the subgroups with average and stable vocabulary and text comprehension scores. On the one hand, the results indicated that a majority of the students had age-appropriate mastery of text comprehension; only a few students had poor scores. On the other hand, this finding suggests that the focus of reading classes in Estonia is the promotion of average students without providing enough challenging tasks to achieve the highest comprehension proficiency.

This speculation is confirmed by the findings related to the first aim. The present study found that the focus of the national tests was the tasks that a majority of the students could complete, i.e., the tasks that assessed lower-level skills. However, the effects of these tests on instruction and, thus, the comprehension skill development should not be underestimated. The national standard-determining tests, which had a predominance of literal comprehension tasks, encourage the promotion of lower-level skills that do not contribute to the development of text comprehension proficiency (Vestheim & Lyngsnes, 2016). Moreover, this assumption is supported by the finding that there were no students with high vocabulary and text comprehension scores. Comparisons of Estonian and Finnish students have indicated that fewer Estonian students achieved very good results in text comprehension (Soodla et al., 2019). On the contrary, the latest PISA international survey results (OECD, 2019) indicated that there was a slight increase in the number of students (13.9%) with top scores compared to the results from previous years. This might be related to the large number of students with average scores. Some students could have improved their skills enough to achieve a higher-level understanding of texts. An increase in teachers' awareness of the need for students to progress from being average achievers to high achievers and for more explicit support to be provided to talented students could lead to a greater percentage of students with high scores. For example, more top readers can be produced by enhancing metacognitive and critically thinking skills by evaluating texts through discussions, argumentation, and open-ended questions (Babic & Baucal 2011; Mercer, 2013).

Because of the stability of the vocabulary and text comprehension scores of most of the students in both grades, the existence of commonalities in the inter-action and reciprocity of vocabulary and literal, inferential, and evaluative

processes could be assumed (Tennent, 2015; Verhoeven et al., 2011). Improvements in vocabulary lead to improvements in text comprehension and vice versa. When readers have mastered and automatised lower-level skills (e.g., vocabulary and literal comprehension), they have enough cognitive resources for higher-level processing (Cain & Oakhill, 2014; Perfetti, 2007). This likely has a positive influence on higher-level comprehension.

Although most of the students were in the subgroups with stable vocabulary and text comprehension in both grades, a few students were in the subgroups with very low and unstable results. Unexpectedly, these students had better scores in inferential and evaluative comprehension than in vocabulary and literal comprehension. This may indicate that the literal understanding of a text is not always necessary for deeper comprehension. For some students, answering inferential and evaluative questions can be easier than extracting facts from texts (Basaraba et al., 2013). Also, the shortcomings in lower-level skills can be compensated by thorough pre-knowledge of the topic (Stahl, 1991). A good pre-knowledge of read topics can enhance inference-making from and critical judgements of texts and thus neutralise vocabulary or literal comprehension deficiencies (Hirsch, 2003; VanderVeen et al., 2007). These students can use their pre-knowledge to create situation models of the texts; however, they might not be able to find specific text-related facts.

5.3 Effectiveness of the intervention on vocabulary and text comprehension

The students who participated in the intervention exhibited significant increases in vocabulary and literal, inferential, and evaluative text comprehension (Article V). This suggests that the explicit teaching of text comprehension strategies can lead to greater proficiency at even higher comprehension levels. The knowledge of when and how to apply text comprehension strategies allows readers to control, to monitor, and to increase their understanding of texts (Kong, 2019). The strategies are needed in understanding of when and where comprehension fails and in repairing these breakdowns (Afflerbach & Cho, 2009; Peterson et al., 2001). This is especially important in creating and expanding situation models, i.e., the higher-level understanding of texts (Kintsch, 1998). Graesser (2007) asserted that the knowledge and application of multiple comprehension strategies can help to crack the illusion of comprehension in students for whom text-based representations suffice as reading outcomes. Teaching these students several comprehension strategies could help them to progress from being poor to average and even high achievers in text comprehension.

The students who did not receive explicit teaching in text comprehension strategies exhibited significant improvements in literal comprehension only. This may suggest that literal comprehension is automatised by Grade 6 and can be improved without the explicit teaching of comprehension strategies. However, a

knowledge of multiple comprehension strategies is needed to perform controlled processes that are needed in inferential and evaluative comprehension (Dole et al., 1991; Kintsch & Rawson, 2005). To successfully teach comprehension strategies, teachers should be aware of their own reading strategies (Duffy, 2003). A study conducted in Estonia indicated that teachers lacked knowledge about reading strategies (Soodla et al., 2017). Although teachers often use comprehension strategies, such as asking questions or summarising texts, they do not always explicitly teach students how and when these strategies should be used (Duke & Pearson, 2009). Without the provision of multiple comprehension strategies in reading classes, the focus might continue to be literal comprehension. This would result in superficial reading that could lead to a decline in reading motivation (Applegate et al., 2002).

Previous studies have found that the benefits of strategy teaching can be moderated by differences in text comprehension proficiency (Griffith & Ruan, 2005; Van Keer & Verhaeghe, 2005). Therefore, the changes in the student subgroups were examined. There were changes in vocabulary and text comprehension in all the subgroups that received explicit teaching in comprehension strategies. The intervention was most valuable for students with low average vocabulary and text comprehension proficiency. These students improved their literal, inferential, and evaluative comprehension proficiency. The students with low vocabulary and text comprehension scores improved their lower-level skills, i.e., vocabulary and literal comprehension. This could be indicative of enhancements in the zone of proximal development.

According to McNamara and Kendeou (2011), better readers tend to improve their higher-level skills (e.g., inferential and evaluative comprehension), and poor comprehenders improve their ability to create more precise text-based representations. However, this explanation does not clarify the results for the students with high average vocabulary and text comprehension scores. Specifically, in these subgroups, significant improvements were observed only in literal comprehension; nevertheless, their inferential comprehension was slightly enhanced. The expectation of improvements in evaluative text comprehension in light of the zone of proximal development was not realised. The reason could be that these students might have already acquired the comprehension strategies needed for deeper reading (Duke & Pearson, 2009). It has been previously noted that skilful readers comprehend texts better if they can use their preferred strategies (Griffith & Ruan, 2005). Teaching these students new strategies could distract their attention from or even disrupt their own functioning systems (Elleman, 2017). Elaborating their knowledge and increasing their metacognitive awareness of their preferred strategies might be more effective.

The conclusion that poor readers are the greatest beneficiaries of strategy teaching (Applegate et al., 2006; NRP, 2000) was not confirmed by this study. It may be that the intervention time was too short for the poor comprehenders. Learning new strategies requires a great deal of effort and conscious activity that can tax cognitive resources (Graesser, 2007). The students who had very low vocabulary and text comprehension scores might have already been cognitively

loaded because of substantial deficits. They might have needed to significantly improve their lower-level skills in order to have enough resources to increase their higher-level comprehension proficiency. A longer intervention period would have given these students more time to practice new strategies. This, in turn, would have contributed to the automatised, unconscious, and effortful use of the strategies (Dole et al., 1991). It would have made more resources available for improving the necessary skills for achieving higher levels of understanding.

The changes among the control group students were found only in those with low average scores in vocabulary and text comprehension. These students exhibited improvements in literal comprehension only. This suggests that teachers might prefer to work with students with average ability. The result is that less attention is given to those with below or above average students (see Paragraph 5.2). That the improvements were observed only in literal comprehension suggests that teachers tend to focus on promoting students' lower-level skills. The reason might be deficiencies in the preparation of reading teachers. Comprehension instruction should include the explicit teaching of multiple comprehension strategies (Duke & Pearson, 2009; NRP, 2000). However, whether teachers have sufficient knowledge for teaching higher-level comprehension strategies is in question (Soodla et al., 2017). Deficiencies in this area could lead to situations in which teachers assess rather than teach text comprehension. Assessment should not replace instruction, especially when the questions lead to understanding the texts at literal level.

5.4 Strengths and limitations of the study

This doctoral study has some limitations related to the research design and methodology. First, the newly developed VTs and TCTs could not distinguish students with very good results in vocabulary and literal, inferential, or evaluative comprehension. Comparisons of the development of skilled readers and average and poor comprehenders are needed for a more in-depth examination of the critical differences between successful and unsuccessful performances in text comprehension. Therefore, to verify these results, the tests developed for this study should be improved. In any modifications of the tests, attention should be paid to the inferential comprehension tasks because of the relatively low internal reliability for this comprehension level at Grade 6. A goal for new test designs should be the achievement of a more balanced distribution of the tasks to measure the various comprehension levels.

Second, the VTs and TCTs were implemented by Estonian-language teachers who received detailed explanations about the administration of the tests. However, the utility of this information and its role in the students' performance are unknown. To ensure more precise and comparable results, a researcher should be involved in the test-taking process.

Third, the text comprehension tasks in the national standard-determining tests were categorised on three levels and analysed. The use of these three comprehension levels in future analyses of Estonian students' comprehension scores in the national standard-determining tests would be worthwhile. This would provide information for students' proficiency at these comprehension levels.

Limitations are associated also with the intervention design. The intervention was very intensive because it was administered within a short period and included the teaching of six text comprehension strategies. More time with a dispersed amount of activities in the intervention would have given students enough time to internalise and to automatise the comprehension strategies learned during the intervention. In addition, retention tests to measure the long-term effects of the intervention would have been valuable. However, this is already being considered. The results of these tests are being analysed in other research projects.

The study was limited by the selection of students in the experimental and control groups. Although the formulation of the experimental and control groups was based on the Grade 4 assessment, the evaluative comprehension results for these two Grade 6 groups were not similar. For greater precision in the conclusions and implications, the samples should have been selected more carefully. One possibility would have been to include additional selection criteria, such as IQ scores and average academic achievement. Another option would have been to engage specific groups, such as students with learning disabilities and those who are high-achieving. This would have allowed for a more accurate definition of the students' abilities. Greater homogeneity in the samples would have provided clearer indications of the effects of the intervention.

Yet another limitation was that the possible progression of students between the subgroups was not examined on the basis of their pre- and post-test results. This information would have provided some indication of the effects of the intervention (experimental group) or general teaching (control group) on the students' developmental paths. Additionally, the intervention, which was implemented under regular classroom conditions, was executed by the teachers only. The teachers were trained before the intervention and provided with detailed scenarios and comprehensive support from the researchers. However, their capacity to convey this information and to use the scenarios in the classroom might have been affected by their own experience and teaching styles. On the one hand, this might have added to the variance in the amount of improvement. On the other hand, conducting educational interventions as a part of regular classroom instruction is crucial because these interventions are meant to be used by teachers in reading classes.

Despite these limitations, the doctoral study has several strengths. Thus, it can provide guidance for teachers, test developers, and researchers who are planning text comprehension studies and interventions. This study represents one of the first attempts to provide an in-depth examination and interpretation of the text comprehension tasks in the national Estonian-language standard-determining tests from the cognitive view of comprehension. Accordingly, it provides detailed information about the aspects of comprehension that are measured and the

effectiveness of these tests for identifying deficiencies and monitoring student progress in comprehension. Further, this study presents a model for designing theoretically relevant TCTs that could achieve the goal of the national standard-determining tests: to provide instructionally relevant feedback to teachers. These findings could be of interest to test developers and teachers.

The results of this doctoral study emphasise the importance of person-oriented approaches to the identification of individual differences. Such approaches are crucial to the design of interventions to examine the effectiveness of various types of instruction for students with individual peculiarities (Bergman & Wångby, 2014). The results indicated that students do not develop or benefit from interventions in the same way. Therefore, students' individual differences should be considered in instructional practices.

The results confirmed the suitability of this intervention for enhancing text comprehension under normal classroom conditions. Detailed information about the intervention design and several examples of lesson scenarios, student materials, and strategy descriptions would be beneficial in the design of future interventions. The aforementioned examples could also assist teachers by showing them how to purposefully teach text comprehension strategies to improve comprehension.

6. CONCLUSIONS AND IMPLICATIONS

6.1 Conclusions

Text comprehension is one of the most important skills for school and everyday life. Although the comprehension product is seen as a mental representation, it is achieved by the interaction of lower- and higher-level processes at multiple levels of understanding (Kintsch, 1998; McNamara & Kendeou, 2011). This doctoral study provided an in-depth examination of text comprehension. The students' performance in the various text comprehension components and processes was considered. The study aimed to shed light on how text comprehension is measured with the national standard-determining tests, and what kind of instructional feedback could it provide to teachers. To expand the knowledge of how to better enhance literal, inferential, and evaluative comprehension, an intervention to explicitly teach text comprehension strategies was developed.

First, the analysis revealed that the text comprehension tasks in the national standard-determining tests were not based on viable comprehension theories. Specifically, the tests measured text comprehension unidimensionally and did not include enough tasks at different text comprehension levels, especially at evaluative levels. In addition, there were inconsistencies in the distribution of tasks for the same grades over the years. These considerations indicated the need to develop theory-based TCTs. Additionally, the VTs were designed to better understand the associations between vocabulary and text comprehension. Newly developed TCTs could satisfactorily distinguish the three text comprehension levels. The essential associations between vocabulary and every text comprehension level indicate the need to enhance the students' vocabulary to improve comprehension.

Second, individual differences were found in the Grade 4 and 6 students' vocabulary and text comprehension levels. More than half of the students belonged to the subgroup with average and stable vocabulary and text comprehension per-formance. This stability in the results showed the interactions of the lower- and higher-level text comprehension components and processes. That a majority of the students had average scores and no students had high scores was an indication of teachers' tendency to focus on average students at the expense of challenging the high achievers. In both grades, fewer than one-tenth of the students had low and inconsistent vocabulary and text comprehension scores. These students need special attention in reading classes and in research.

Third, an intervention to enhance vocabulary and text comprehension was successfully implemented. The explicit teaching of text comprehension strategies appeared to be most beneficial for the students with low average vocabulary and text comprehension scores. These students exhibited improvements in literal, inferential, and evaluative comprehension. Those with low scores improved their overall vocabulary and literal comprehension. Least effective was the intervention for students with high average vocabulary and text comprehension scores. Only their literal comprehension was enhanced. The students who attended the regular

reading classes without the explicit teaching of text comprehension strategies exhibited only minor improvements. Only the students with low average vocabulary and text comprehension scores exhibited improvements in literal comprehension.

In sum, this doctoral study highlights the need for the design of multidimensional TCTs. Information about the interpretation of the results should also be provided. The national standard-determining tests that measure text comprehension at different levels could provide teachers with instructionally appropriate feedback for planning reading lessons. The differences in the students' developmental paths and the variety of students who benefitted from the intervention indicate the importance of understanding students' individual peculiarities in order to better enhance comprehension at different levels. Moreover, the finding that the students who did not participate in the intervention did not exhibit improvements in inferential and evaluative comprehension could be an indication of deficiencies in the reading classes. This might be the result of the preparation of the reading teachers and deficiencies in the national standard-determining tests, which guide teachers' work. The results of this study could contribute to the design of new reading assessments and a different in- and pre-service teacher training system.

6.2 Implications and recommendations

The findings of this study have several theoretical and methodological implications for text comprehension research and test development. The results also have practical implications and provide recommendations for teachers and teacher educators.

The following are the theoretical and methodological implications:
1. The text comprehension tasks in the Estonian-language national standard-determining tests were analysed on the basis of a three-level taxonomy: literal, inferential, and evaluative comprehension. The analyses revealed the over-reliance on literal tasks in the national standard-determining tests. There was variation in the task distribution for the same age group. The tests for the younger students had a higher proportion of evaluative tasks (Articles I, II, and III). Accordingly, there is a need for new tests that are based on a cognitive view of text comprehension and include a balanced range of age-appropriate tasks at multiple comprehension levels. To provide more comprehensive tests, the cooperation of scientists from various fields (psychologists, educational scientists, philologists, and speech therapists) is needed. For example, psychologists can provide their expertise in the cognitive processes that are involved in text comprehension (e.g., memory, metacognition, attention, thinking, and reasoning), and educational scientists could provide advice on age-appropriate tasks and texts.

2. In this doctoral study, unidimensional VTs and three-dimensional TCTs were developed (Articles IV and V). To gain a comprehensive view of the students' text comprehension, the tests should include tasks at multiple cognitive levels. Information on the three-level text comprehension taxonomy (Basaraba et al., 2013; NAEP, 2019b; OECD, 2009; Tennent, 2015) and the tests designed for this study can inform the development of new TCTs for scientific purposes (e.g., to investigate text comprehension) or educational purposes (e.g., to design national standard-determining tests).
3. The individual-level analyses clearly indicated the differences in the text comprehension components and processes, as well as the variance in the effects of the intervention (Articles IV and V). Thus, to comprehensively investigate text comprehension and to determine the best interventions to improve student achievement, individual-level analyses should be applied in future studies.
4. The study revealed that the implementation of a complex intervention in classroom conditions, even for a short period, can be effective (Article V). To understand the effects of teaching, it is important to increase the number of interventions that are implemented in normal classroom conditions. Therefore, instead of implementing instructional interventions under specific conditions, researchers should include more teachers in the administration of such programs. The guidelines from this study could be used to develop and to implement new well-designed and organized interventions in schools.

Next, the practical implications and recommendations for teachers, teacher educators, and educational politicians to improve the quality of educational practices in reading classes are considered.
1. In the analysed national standard-determining tests, text comprehension tasks were just one component among several other language skills (Articles I, II, and II). However, text comprehension is considered a highly important basic skill that is assessed with single-purpose TCTs (Tengberg, 2017; Vestheim & Lyngsnes, 2016). To provide more precise information, the development of single-purpose comprehension tests should be considered. This would facilitate the identification of possible deficiencies and thus enable teachers to improve their instructional practices.
2. The analyses revealed that the text comprehension tasks in the national Estonian-language standard-determining tests were developed without consideration of text comprehension theories (Articles I, II, and III). These tests can help to measure the students' ability to find and to remember the facts from texts but not necessarily their ability to discern the meaning of the text. Accordingly, Estonian schools should not be classified on the basis of the results of these tests because of their failure of the tests to adequately measure text comprehension. Furthermore, there is a need for the development of new multidimensional national assessments that include tasks at different cognitive levels and information about the components and processes being assessed. This design would provide teachers with the necessary feedback on student

proficiency and deficiencies and thus facilitate improvements in instructional practices.

3. The study found variations in the text comprehension components and processes. In addition, there were differences in the students' responses to the teaching (Articles IV and V). Thus, training programs for in- and pre-service teachers should provide models that promote the comprehensive teaching of text comprehension. For example, these programs could give teachers a thorough understanding of the text comprehension construct, demonstrate the relations between text comprehension and various cognitive processes, development in these processes and examine the variance in students' cognitive abilities as well as how to use this knowledge for teaching and assessing text comprehension.

4. The effectiveness of the intervention shows the importance of the explicit teaching of text comprehension strategies (Article V). Given the essential role of these strategies in inferential and evaluative comprehension, strategy teaching should be a valued component in the curriculum and, specifically, reading classes. Accordingly, strategy teaching should be highlighted in teacher training programs to allow teachers to develop an awareness of their own reading strategies and then become metacognitively skilful at deploying the explicit teaching of comprehension strategies in reading classes.

APPENDICES

Appendix 1. Example of the strategy card for students from intervention (Article V)

SKIM READING – PREVIEW AND SCAN OF THE TEXT

The purpose of the preview and scan of the text is to have a general overview of the context and the difficulty level of the text. If you have an overview of the events in the text, it is easier for you to understand new words. Having a grasp of how difficult or easy the text is, you can choose an appropriate reading speed.

Based on the text answer to: WHO? WHAT? WHEN? WHERE?
1. Find from the text the *characters, events, time and place*.
2. Underline the unknown or confusing words.

Appendix 2. A fragment from an intervention lesson scenario for teachers (Article V)

Monitoring strategy (lesson no 5)

Materials needed:
- Text *Raven and crow* for every student
- Monitoring strategy card for each student

Introduction
- CLASS DISCUSSION. Ask students to remember the strategies that they have learned in previous reading classes.
- LECTURE.
  - Explain to the students that today they will learn a new activity that helps them understand how they comprehend the text.
  - Tell them that this activity is called monitoring.
- INDEPENDENT WORK. Ask students to read the first box from the strategy card (*Tracking own understanding*).
- CLASS DISCUSSION based on the read information from the strategy card
  - Ask students to explain why it is important to track your own understanding.
  - Ask students, what helps you in tracking your understanding (e.g., stopping, sign system, side-notes, asking yourself questions).
  - Discuss about similar activities that students have used before. Ask them to provide examples of their system (e.g., underlying unknown words).
- INDEPENDENT WORK. Ask students to read the other box from the strategy card (*Working with confusing part of the text*).
- CLASS DISCUSSION based on the read information from the strategy card. Ask students to explain what should be done if they do not understand any part of a text.

Modelling the strategy (10 min)
- ACTIVE LECTURE.
  - Ask students to name the first activity in Working with a confusing part of a text (Map the location of a confusing part of a text). Give an example from text by **thinking aloud**: *I cannot understand the first sentence from the second paragraph. I will stop here. I will underline the sentence.*
  - Ask students to name the second activity in Working with a confusing part of a text (Identify the problem/why you do not understand this part of the text). Provide an example from the text by **thinking aloud**: *I cannot understand what the author means by the sentence: The grey partridges will be weak from hunger and can be easily caught.*
  - Ask students to name the third activity in Working with a confusing part of a text (Rephrase the confusing part of a text). Provide an example from the text by **thinking aloud**: *I will try to rephrase the sentence. If grey*

partridges cannot get food, they will be weak, and crows and ravens can easily hunt them.

o Ask students to name the fourth activity in <u>Working with a confusing part of a text</u> (Look back). Provide an example from the text by **thinking aloud**: *But why would thaw and freeze leave grey partridges without food? I will read the previous paragraphs where we can find information about grey partridges. Oh ok, it says here that grey partridges can get their food from the ground. So, they have to get to the ground somehow. But if it thaws in the daytime and freezes at night then a snow crust will be formed, and grey partridges cannot reach the ground. So, they cannot get the food and they will starve.*

o Ask students to name the fifth activity in <u>Working with a confusing part of a text</u> (Look forward). Explain that as in the previous steps, sometimes it is necessary to look forward to find information that would explain the confusing part of a text.

REFERENCES

About Innove. (n.d.). https://www.innove.ee/en/about-innove/

Afflerbach, P., & Cho, B. Y. (2014). Identifying and describing constructively responsive comprehension strategies in new and traditional forms of reading. In Susan E. Israel & Gerald G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 69–90). New York and London: Routledge.

Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The reading teacher, 61*(5), 364–373.

Allington, R. L. (2001). *What really matters for struggling readers: Designing research-based programs*. New York: Addison Wesley Longman.

Alonzo, J., Basaraba, D., Tindal, G., & Carriveau, R. S. (2009). They read, but how well do they understand? An empirical look at the nuances of measuring reading comprehension. *Assessment for Effective Intervention, 35*(1), 34–44.

Angosto, A., Sánchez, P., Álvarez, M., Cuevas, I., & León, J. A. (2013). Evidence for top-down processing in reading comprehension of children. *Psicologia Educativa, 19*, 83–88.

Applegate, M. D., Quinn, K. B., & Applegate, A. J. (2002). Levels of thinking required by comprehension questions in informal reading inventories. *The Reading Teacher, 56*(2), 174–180.

Applegate, M. D., Quinn, K. B., & Applegate, A. J. (2006). Profiles in comprehension. *The Reading Teacher*, *60*(1), 48–57.

Aryadoust, V. (2019). An integrated cognitive theory of comprehension. *International Journal of Listening*, *33*(2), 71–100.

Babic, D. P. & Baucal, A. (2011). The big improvement in PISA 2009 reading achievements in Serbia: Improvement of the quality of education or something else? *CEPS Journal: Center for Educational Policy Studies Journal, 1* (3), 53–74.

Basaraba, D., Yovanoff, P., Alonzo, J., & Tindal, G. (2013). Examining the structure of reading comprehension: Do literal, inferential, and evaluative comprehension truly exist? *Reading and Writing, 26*(3), 349–379. https://doi.org/10.1007/s11145-012-9372-9

Baumann, J. F. (2014). Vocabulary and reading comprehension: The nexus of meaning. In Susan E. Israel & Gerald G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 347–370). New York and London: Routledge.

Bergman, L. R., Magnusson, D., & El Khouri, B. M. (2003). *Studying Individual Development in an Interindividual Context*. London: Lawrence Erlbaum Associates.

Bergman, L. R., & Wångby, M. (2014). The person-oriented approach: a short theoretical and practical guide. *Eesti Haridusteaduste Ajakiri. Estonian Journal of Education*, *2*(1), 29–49.

Boulware-Gooden, R., Carreker, S., Thornhill, A., & Joshi, R. M. (2007). Instruction of metacognitive strategies enhances reading comprehension and vocabulary achievement of third-grade students. *The reading teacher*, *61*(1), 70–77.

Caemmerer, J. M., Maddocks, D. L., Keith, T. Z., & Reynolds, M. R. (2018). Effects of cognitive abilities on child and youth academic achievement: Evidence from the WISC-III and WIAT-III. *Intelligence*, *68*, 6–20.

Cain, K. (2010). *Reading Development and Difficulties*. Oxford: Wiley–Blackwell.

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology, 76*(4), 683–696. https://doi.org/10.1348/000709905X67610

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année psychologique*, *114*(4), 647–662. https://doi.org/10.4074/S0003503314004035

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*(4), 671–681. https://doi.org/10.1037/0022-0663.96.4.671

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). London: Routledge.

Cook, A. E.*,* Limber, J. E.*,* & O'Brien, E. J. (2001). Situation-based context and the availability of predictive inferences*. Journal of Memory & Language, 44,* 220–234*.*

Currie, N. K., & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology, 137*, 57–75. https://doi.org/10.1016/j.jecp.2015.03.005

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

Davoudi, M., & Moghadam, H. R. H. (2015). Critical Review of the Models of Reading Comprehension with a Focus on Situation Models. *International Journal of Linguistics*, *7*(5), 172–187.

Dimitrov, D. M. (2014). *Statistical methods for validation of assessment scale data in counseling and related fields*. Alexandria: John Wiley & Sons.

Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research, 61,* 239–264.

Dole, J. A., Nokes, J. D., & Drits, D. (2014). Cognitive Strategy Instruction. In Susan E. Israel & Gerald G. Duffy (Eds.), *Handbook of research on reading comprehension*, 347–372. New York and London: Routledge.

Duffy, G. G. (Ed.). (2003). *Improving comprehension: 10 research-based principles*. Washington, DC: National Education Association.

Duke, N. K., & Carlisle, J. (2011). The development of comprehension. In M. L. Kamil, P. D. Pearson, E. Birr Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research 4* (pp. 199–229). New York, London: Routledge.

Duke, N. K., & Pearson, P. D. (2009). Effective practices for developing reading comprehension. *Journal of education, 189*(1–2), 107–122.

Eesti Teaduste Akadeemia. (2002). *Eesti teadlaste eetikakoodeks* [The Estonian Code of Conduct for Research Integrity; in Estonian]. https://www.etag.ee/wp-content/uploads/2013/09/Eetikakoodeks2002.pdf

Elleman, A. M. (2017). Examining the impact of inference instruction on the literal and inferential comprehension of skilled and less skilled readers: A meta-analytic review. *Journal of Educational Psychology*, *109*(6), 761–781.

Estonian Government. (2011/2014). *National Curriculum for Basic Schools*. Riigi Teataja. https://www.hm.ee/sites/default/files/est_basic_school_nat_cur_2014_general_part_1.pdf.

Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific studies of reading*, *10*(3), 323–330.

Floyd, R., Meisinger, E., Gregg, N., & Keith, T. (2012). An explanation of reading comprehension across development using models from Cattell–Horn–Carroll theory: Support for integrative models of reading. *Psychology in the Schools*, *49*(8), 725–743.

Garbe, C., Kerge, K., Lafontaine, D., Puksand, H., Shiel, G., Sulkunen, S., … Valtin, R. (2016). *Literacy in Estonia. Country report. Children and adolescents.* 1−63.

Graesser, A. C. (2007). An introduction to strategic reading comprehension. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 137–172). Mahwah, NJ: Erlbaum.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371–395.

Gresham, F. M. (2009). Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review*, *38*(4), 533–540.

Griffith, P. L., & Ruan, J. (2005). What is metacognition and what should be its role in literacy instruction? In Susan E. Israel, Cathy C. Block, Kathry. L. Bauserman, & Kathryn Kinnucan-Welsch (Eds.), *Metacognition in Literacy Learning. Theory, Assessment, Instruction, and Professional Development* (pp. 25–40). London: Lawrence Erlbaum Associates.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. Thousand Oaks, CA: SAGE Publications, Inc. https://doi.org/10.4135/9781483384436

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective* (7th ed.). Upper Saddle River, NJ: Pearson.

Hammersley, M., & Traianou, A. (2012). *Ethics and educational research*. London: British Educational Research Association. https://www.learnersfirst.net/private/wpcontent/uploads/Ethics-and-Educational-Research.pdf

Haridus- ja Teadusministeerium. (2018). *Tasemetööd* [National standard-determining tests]. https://www.hm.ee/et/tasemetood

Hirsch, E. D. (2003). Reading comprehension requires knowledge of words and the world. *American Educator, 27*(1), 10–13.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55. https://doi.org/10.1080/10705519909540118

Innove. (2014). *Üleriigiline tasemetöö, eesti keel, 6. klass* [National standard-determining test in Estonian language for Grade 6]. https://haridusinfo.innove.ee/et/yldharidus/tasemetood/tasemetoode-materjalid/tasemetood-2014

Kaalep, H.-J., & Muischnek, K. (2002). *Eesti kirjakeele sagedussõnastik* [The Estonian frequency dictionary]. Tartu: TÜ kirjastus.

Kärbla, T., Uibu, K., & Männamaa, M. (2017). Eesti keele riiklike tasemetööde teksti-mõistmisülesannete analüüs [Analysis of the comprehension tasks of national standardized tests of the Estonian language]. *Eesti Rakenduslingvistika Ühingu aastaraamat, Estonian Papers in Applied Linguistics, 13*, 73–87. https://doi.org/10.5128/ERYa13.05

Kärbla, T., Uibu, K., & Männamaa, M. (2018). Pikilõikeline sissevaade eesti keele taseme-ja eksamitööde tekstimõistmisülesannetesse [A longitudinal introspective into the text comprehension tasks of national tests and exams of the Estonian language]. *Eesti Haridusteaduste Ajakiri. Estonian Journal of Education*, *6*(1), 157–178. https://doi.org/10.12697/eha.2018.6.1.07

Kärbla, T., Uibu, K., & Männamaa, M. (2019). National Estonian-language tests: What is measured in text comprehension tasks? *New Trends and Issues Proceedings on Humanities and Social Sciences, 6*(5), 8–16.

Keenan, J. (2016). Assessing the assessments reading comprehension tests. *Perspectives on Language and Literacy*, *42*(2), 17–21.

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*(3), 281–300.

Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities, 47*(2), 125–135.

Kendeou, P., & Papadopoulos, T. C. (2012). The use of CBM-Maze in Greek: A closer look at what it measures. *A measure of success: The influence of Curriculum-Based Measurement on education*. Minneapolis, MN: University of Minnesota Press.

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.

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2007). Preschool and early elementary comprehension: Skill development and strategy interventions. In D.S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 27–45). Mahwah, NJ: Erlbaum.

Kendeou, P., P. van den Broek, M. J. White, & J. S. Lynch. (2009). Predicting Reading Comprehension in Early Elementary School: The Independent Contributions of Oral Language and Decoding Skills. *Journal of Educational Psychology, 101*, 765–778.

Kikerpill, T., & Türk, Ü. (2013). Tekstisisesed seosed eesti keele kui teise keele lugemistestis: mõõtmisvahendi valiidsus [Testing understanding text cohesion and coherence in Estonian as a second language: The validity of different task types]. *Lähivõrdlusi. Lähivertailuja*, (23), 121–147.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95*, 163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kintsch, W., & Kintsch, E. (2005). Comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 71–92). Mahwah, NJ: Erlbaum.

Kintsch, W., & Rawson, K. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 211–226). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch12

Kitsnik, M. & Metslang, M. (2011). *Põhikooli ainetundide sõnavara. Abimaterjal õppekeelest erineva emakeelega õpilaste keeleõppeks [Vocabulary lessons in primary school. Additional learning material for students with a different native language]*. Tallinn: Integratsiooni ja Migratsiooni Sihtasutus.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Kong, J. (2019). *Investigating the Role of Test Methods in Testing Reading Comprehension: A Process-Focused Perspective*. Singapore: Springer.

Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. *Reading comprehension strategies: Theories, interventions, and technologies*, 107–136.

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91*, 615–629.

McNamara, D. S., & Kendeou, P. (2011). Translating advances in reading comprehension research to educational practice. *International Electronic Journal of Elementary Education*, *4*(1), 33–46.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of learning and motivation*, *51*, 297–384.

McNamara, D. S., Ozuru, Y., Best, R., & O'Reilly, T. (2007). The 4-pronged comprehension strategy framework. *Reading comprehension strategies: Theories, interventions, and technologies*, 465–496.

Mercer, N. (2013). The social brain, language, and goal-directed collective thinking: A social conception of cognition and its implications for understanding how we think, teach, and learn. *Educational Psychologist, 48* (3), 148–168. https://doi.org/10.1080/00461520.2013.804394

Morsy, L., Kieffer, M., & Snow, C. (2010). *Measure for measure: A critical consumers' guide to reading comprehension assessments for adolescents*. Final Report from Carnegie Corporation of New York's Council on Advancing Adolescent Literacy. Carnegie Corporation of New York.

Muijs, D. 2004. *Doing Quantitative Research in Education with SPSS*. London: Sage.

Must, O. (1997). Literacy: The case of Estonia. *Trames: Journal of the Humanities and Social Sciences, 1*, 40−65.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus User's Guide: Statistical Analysis with Latent Variables* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition*, *70*(1), B1–B13.

National Assessment of Educational Progress (NAEP). (2019a). *Explore Results from the 2019 NAEP Reading Assessment*. https://www.nationsreportcard.gov/reading/?grade=4

National Assessment of Educational Progress (NAEP). (2019b). *Reading Framework for the 2019 National Assessment of Educational Progress*. https://www. nagb.gov/naep-frameworks/reading.html

National Reading Panel (NRP). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute for Child Health and Human Development.

Oakhill, J. V., Berenhaus, M. S., & Cain, K. (2015). Children's reading comprehension and comprehension difficulties. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 344–360). New York: Oxford University Press.

Oakhill, J., & Cain, K. (2007). Issues of Causality in Children's Comprehension Development. In D. S. McNamara (Ed.), *Reading Comprehension Strategies: Theory, Interventions, and Technologies* (pp. 47–71). Mahwah, NJ: Lawrence Erlbaum Associates.

Oakhill, J., & Cain, K. (2011). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16,* 91–121.

OECD. (2009). *PISA 2009 assessment framework. Key competencies in reading, mathematics and science*. https://www.oecd.org/pisa/pisaproducts/ 44455820.pdf

OECD. (2019). *PISA 2018 Results (Volume I): What Students Know and Can Do*. Paris: OECD Publishing. https://doi.org/10.1787/5f07c754-en.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554–566. https://doi.org/10.1037/0022-0663.98.3.554

Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 67(3), 215.*

Pajupuu, H. (2007). Kuidas hinnata suure panusega testide hindajaid [*How to assess the raters of high-stakes tests*]. *Eesti Rakenduslingvistika Ühingu aastaraamat*, 3, 221−233.

Paris, S. G., & Hamilton, E. E. (2014). The development of children's reading comprehension. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 56–77). New York and London: Routledge.

Pearson, P. D., & Gallagher, M. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology, 8*(3), 317–344.

Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices – Past, present, and future. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13–69). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, *42*(2), 282–296. https://doi.org/10.1598/RRQ.42.2.4

Pedaste, M., Brikker, M., Rannikmäe, M., Soobard, R., Mäeots, M., & Reiska, P. (2017). *Loodusvaldkonna õpitulemuste hindamine [Assessment of achievements in Science Education].* https://www.innove.ee/UserFiles/Organisatsioonist/ESF%20tegevused/Loodusvaldkonna%20e-hindamine.pdf

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, *11*(4), 357–383.

Perfetti, C. A., Yang, C.–L., & Schmalhofer, F. (2008). Comprehension skill and word-to-text integration processes. *Applied Cognitive Psychology, 22*, 303–318. https://doi.org/10.1002/acp.1419

Peterson, C. L., Caverly, D. C., Nicholson, S. A., O'Neal, S., & Cusenbary, S. (2001). *Building Reading Proficiency at the Secondary Level: A Guide to Resources. Introduction*.

Riigi Teataja. (2018). *Riigieksamite vormid ja ajad, ühtsete põhikooli lõpueksamite ning tasemetööde õppeained, vormid ja ajad 2018/2019. õppeaastal* [The subjects, forms and time of harmonized final examinations and standard-determining tests of basic school in 2018/2019]. https://www.riigiteataja.ee/akt/118052018004

Ryan, G. W. & Bernard, H. R. (2003). Techniques to Identify Themes. *Field Methods 15*(1), 85–109.

Sabatini, J., O'Reilly, T., & Deane, P. (2013). Preliminary reading literacy assessment framework: Foundation and rationale for assessment and system design. *ETS Research Report Series, 2013*(2), i–50.

Samuels, S. J., Ediger, K. M., Willcutt, J. R., & Palumbo, T. J. (2005). Role of automaticity in metacognition and literacy instruction. In S. E. Israel, C. C. Block, K. L. Bauserman, & K Kinnucan-Welsch (Eds.), *Metacognition in literacy learning: Theory, assessment, instruction, and professional development* (pp. 41–59). Mahwah, NJ: Erlbaum.

Silva, M., & Cain, K. (2015). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology*, *107*(2), 321.

Slavin, R. E., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best-evidence synthesis. *Reading Research Quarterly, 43*, 290–322.

Snow, C. E. (2002). *Reading for Understanding: Toward a Research and Development Program in Reading Comprehension* (pp. 19–28). Santa Monica, CA, USA: RAND Corporation.

Soodla, P., Jõgi, A. L., & Kikas, E. (2017). Relationships between teachers' metacognitive knowledge and students' metacognitive knowledge and reading achievement. *European journal of psychology of education*, *32*(2), 201–218.

Soodla, P., Lerkkanen, M.–K., Niemi, P., Kikas, E., Silinskas, G., & Nurmi, J.-E. (2015). Does Early Reading Instruction Promote the Rate of Acquisition? A Comparison of Two Transparent Orthographies. *Learning and Instruction 38*, 14–23.

Soodla, P., Torppa, M., Kikas, E., Lerkkanen, M.–K., & Nurmi, J. E. (2019). Reading comprehension from grade 1 to 6 in two shallow orthographies: comparison of Estonian and Finnish students. *Compare: A Journal of Comparative and International Education*, *49*(5), 681–699. https://doi.org/10.1080/03057925.2018.1445963

Spencer, M., Wagner, R. K., & Petscher, Y. (2019). The reading comprehension and vocabulary knowledge of children with poor reading comprehension despite adequate decoding: Evidence from a regression-based matching approach. *Journal of Educational Psychology, 111*(1), 1–14. https://doi.org/10.1037/edu0000274

Stahl, S. A. (1991). Beyond the instrumentalist hypothesis: Some relationships between word meanings and comprehension. In. P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 157–186). Hillsdale, NJ: Erlbaum.

Sullivan, G. M. (2011). A Primer on the Validity of Assessment Instruments. *Journal of Graduate Medical Education*, *3*, 119–120.

Tengberg, M. (2017). National reading tests in Denmark, Norway, and Sweden: A comparison of construct definitions, cognitive targets, and response formats. *Language Testing, 34*(1), 83–100. https://doi.org/10.1177/0265532215609392

Tennent, W. (2015). *Understanding reading comprehension: Processes and practices.* Thousand Oaks: SAGE.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second- to fifth-grade children. *Scientific Studies of Reading*, *1*, 161–185.

Trabasso, T., & Bouchard, E. (2002). Teaching readers how to comprehend text strategically. In C. C. Block, & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 176–200). New York: Guilford.

Uibu, K., & Männamaa, M. (2014). Õpetamistegevused ja õpilaste tekstimõistmine üleminekul esimesest kooliastmest teise astmesse [Teaching practices and text comprehension in students during the transition from the first to second stage of school]. *Eesti Haridusteaduste Ajakiri, Estonian Journal of Education, 2*(1), 96–131.

United Nations Educational, Scientific and Cultural Organization. (2016). *Education 2030: incheon declaration and framework for action for the implementation of sustainable development goal 4*. https://unesdoc.unesco.org/images/0024/002456/245656e.pdf.

van den Broek, P. (1994) Comprehension and memory of narrative texts: inferences and coherence. In M.A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 539–88). San Diego, CA: Academic Press.

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 den Broek, P., & Kremer, K. (2000). The mind in action: What it means to compre-hend during reading. In B. M. Taylor, M. F. Graves, & P. van den Broek (Eds.), *Reading for meaning: Fostering comprehension in the middle grades* (pp. 1–31). Newark, DE: International Reading Association.

van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist approaches in accounts of reading comprehension. *Discourse Processes, 39*, 299–316.

van Keer, H., & Verhaeghe, J. P. (2005). Effects of explicit reading strategies instruction and peer tutoring on second and fifth graders' reading comprehension and self-efficacy perceptions. *The Journal of Experimental Education, 73*(4), 291–329.

VanderVeen, A., Huff, K., Gierl, M., McNamara, D. S., Louwerse, M., & Graesser, A. (2007). Developing and validating instructionally relevant reading competency profiles measured by the critical reading section of the SAT Reasoning Test™. In D. McNamara (Ed.) *Reading Comprehension Strategies. Theories, Interventions, and Technologies* (pp. 137–172). New York, London: Taylor & Francis group.

Vardja, M. (2011). *Lugemine. Lugemisstrateegiad ja nende kujundamine* [Reading. Reading strategies and formation of these]. https://translate.google.com/#view=home &op=translate&sl=et&tl=en&text=kujundamine

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, *15*(1), 8–25. https://doi.org/10.1080/10888438.2011.536125

Vestheim, O. P., & Lyngsnes, K. M. (2016). A study of Norwegian schools achieving good national test results. *International Journal of Educational Research, 79*, 1–9. https://doi.org/10.1016/j.ijer.2016.06.005

Williams, G. A., & Kibowski, F. (2016). Latent class analysis and latent profile analysis. In L. A. Jason, & D. S. Glenwick (Eds.) *Handbook of methodological approaches to community-based research: Qualitative, quantitative, and mixed methods*, (pp. 143–151). Oxford: Oxford University Press.

World Economic Forum. (2016). The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution. In *Global Challenge Insight Report, World Economic Forum, Geneva*. https://www3.weforum.org/docs/WEF_Future_ of_Jobs.pdf.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*, 292–297.

ACKNOWLEDGEMENTS

These four years of doctoral studies have provided me with many new, stimulating and expanding experiences and self-development as well as brought me together with people to whom I am very grateful. Thanks to my supervisors Krista Uibu and Mairi Männamaa, I explored the path of scientific life, and their continuous and critical feedback with valuable advice and guidance led to my improvement in conducting researches. Thank you!

Further, I am grateful to my employer and colleagues from Võru Kesklinna Kool who were by my side being supportive and helpful whenever I needed it. I also thank the students from my school who participated in pilot studies, and the Estonian language teachers who carried out the pilot tests and intervention. Likewise, my sincere gratitude goes to all the students who attended and the teachers who carried out the main tests and intervention. In addition, I really appreciate the contribution of those people from the University of Tartu who contributed to the improvements of tests and intervention. Special thanks to Jaan Mikk and Ilona Võik for giving me feedback and guiding me in training the teachers for the intervention. Similarly, I am very grateful for Krista Kerge and Miia Rannikmäe for their valuable time and constructive comments and recommendations that helped me to improve my work after pre-defense. I also thank Anneli Vainumäe and David Bill for language editing the different parts of my doctoral study. Thank you for devoting your time.

The doctoral study is dedicated to my family, my husband and my three daughters. Thank you Veiko, Kristin, Liis and Sandra for being there and growing with me during this intense journey. Although in the last four years I spent less time with you than all of us would have wanted, your acceptance and tolerance helped me to go through this process more smoothly.

SUMMARY IN ESTONIAN

Tekstimõistmise hindamine ja tekstimõistmist toetavate strateegiate õpetamine Eesti põhikoolis

Tekstist arusaamine on oluline oskus, mis võimaldab uusi teadmisi omandada ja igapäevaelus hakkama saada (van den Broek & Espin, 2012). Vähene tekstimõistmisoskus on üks peamisi põhjuseid, miks õpilastel tekivad õpiraskused. Õpiraskused võivad omakorda vähendada õpimotivatsiooni ja põhjustada koolist väljalangemise (Fiester, 2010; Keenan, 2016; Snow, 2002). Tekstimõistmisraskused on tingitud mitmesugustest asjaoludest, kaasa arvatud sellest, kuidas teksti mõistmist koolis õpetatakse ja hinnatakse (Fletcher, 2006). Seejuures rõhutavad nüüdisaegsed tekstimõistmisteooriad, et tekstimõistmisel on mitu tasandit ning see hõlmab mitme omavahel seotud tekstimõistmise komponendi ja protsessi samaaegset kasutamist (McNamara & Kendeou, 2011). Tekstiga töötamisel moodustatakse loetust kõigepealt tekstibaas. Selle käigus antakse sõnadele tähendus. Seejärel seotakse sõnade tähendused omavahel ning ühendatakse need eelteadmistega. Sel viisil luuakse loetud tekstist situatsioonimudel (Kendeou et al., 2014; Kintsch, 1998; McNamara & Magliano, 2009).

Selleks, et jälgida, kuidas õpilane tekstibaasi ja situatsioonimudeli loomiseks vajalikes protsessides areneb, võib tekstimõistmist käsitleda kolmel tasandil. Tekstimõistmise *sõnasõnalisel* tasandil luuakse tekstibaas ning mõistetakse tekstis sisalduvat selgesõnalist infot. Tekstimõistmise *järeldaval* tasandil tõlgendatakse kaudset infot, täidetakse tekstis esinevad tühimikud, seotakse laused ja lõigud ühtseks tervikuks ning luuakse situatsioonimudel. Tekstimõistmise *hindaval* tasandil laiendatakse ja täiustatakse situatsioonimudelit. Seejuures hinnatakse kriitiliselt teksti sisu, tuginedes oma kogemustele ja moraalinormidele (Basaraba et al., 2013; Magliano et al., 1999).

Õpetamiseks vajaliku tagasiside saamiseks peavad tekstimõistmistestid sisaldama ülesandeid, millega hinnatakse õpilaste võimekust tekste eri tasanditel töödelda (Kendeou et al., 2007). Ülesanded tekstimõistmise tasandite protsesside hindamiseks peaksid olema eri liiki, et saada õpilaste tulemustest parem ülevaade (Kikerpill & Türk, 2013). Samas on leitud, et õpilaste teadmisi hinnatakse tihti just üht liiki ülesannete abil (Fletcher, 2006) ja et tekstimõistmistestid on sagedasti ühetasandilised ja suunatud õpilaste sõnasõnalise tekstimõistmise hindamisele (Keenan et al., 2008). Lisaks on tähtis, et tekstimõistmistestid sisaldaksid infot selle kohta, kuidas testiga saadud tulemusi tõlgendada ehk missuguseid tekstimõistmise tasandeid või protsesse saab nende ülesannetega mõõta (Kendeou & Papadopoulus, 2012). Eestis kasutusel olevad taseme- ja eksamitööd ei sisalda aga õpetajale vajalikku infot selle kohta, millisele tekstimõistmisteooriale testid tuginevad ja mida õpilaste tekstimõistmisel täpselt hinnatakse (Kärbla et al., 2018). Et saada ülevaade riiklike taseme- ja eksamitööde olemusest, on tarvis uurida, kuivõrd on nende tekstimõistmisülesannetes arvestatud tekstimõistmise mitmetasandilisust ja õpilaste tekstimõistmise arengut.

Tekstimõistmine kõrgemal ehk järeldaval ja hindaval tasandil nõuab lugejalt sihipärast, aktiivset ja eesmärgipärast tööd tekstiga. See omakorda eeldab, et lugeja on tekstimõistmisstrateegiatest teadlik (Graesser, 2007). Oskuslikud tekstimõistjad kasutavad ja õpivad tekstimõistmisstrateegiaid spontaanselt, kuid nõrgemad lugejad vajavad õpetust, kuidas, millal ja millist strateegiat kasutada (Soodla et al., 2017; van Keer & Verhaeghe, 2005). Kuigi õpetajad kasutavad oma ainetundides tihti mitmesuguseid tekstimõistmisstrateegiaid, nt küsimuste esitamist ja kokkuvõtete tegemist, ei pruugi nad strateegiaid õpilastele otseselt õpetada (Duke & Pearson, 2009). Tekstimõistmise eri tasandite protsesside arendamiseks on tarvis põhjalikke teadmisi nii tekstimõistmise olemusest kui ka metatasandi teadmisi tekstimõistmisstrateegiatest ja nende õpetamisest (Griffith & Ruan, 2005).

Eesti ja Soome õpilaste võrdlevast uuringust on selgunud, et Eesti õpilaste tekstimõistmine on alates 2. klassist oluliselt kehvem, kuigi nende lugemisoskus oli 1. klassi astudes Soome laste omast märkimisväärselt parem (Soodla et al., 2015; Soodla et al., 2019). Kuigi Eesti õpetajatele on tutvustatud erinevaid tekstimõistmisstrateegiaid (Vardja, 2011), võib siiski oletada, et nad vajavad rohkem teadmisi selle kohta, mil viisil eesmärgipäraselt õpilaste tekstimõistmisarengut toetada. Seepärast tuleb välja töötada tekstimõistmisstrateegiate õpetamise programm, mis võimaldaks õpetajatel arendada õpilaste eri tasanditel toimuvat tekstimõistmist.

Doktoritöö eesmärk on analüüsida, kuivõrd eesti keele tasemetöödes sisalduvad tekstimõistmisülesanded mõõdavad õpilaste tekstimõistmist eri tasanditel, ning töötada välja õpilaste tekstimõistmist arendav programm. Töö eesmärgi saavutamiseks on sõnastatud kolm alaeesmärki ja kuus uurimisküsimust:

1. Selgitada välja, kuidas hinnatakse tekstimõistmist eesti keele tasemetöödega, ning arendada välja ühetasandiline sõnavaratest ja mitmetasandiline tekstimõistmistest.

    1.1 Milline on tekstimõistmise tasandite osakaal eesti keele tasemetöödes? (Artiklid I, II, III)

    1.2 Mil määral muutub tekstimõistmise tasandite osakaal eri klasside tasemetöödes? (Artiklid II, III)

    1.3 Kuidas eristuvad sõnavara ja tekstimõistmise tasandid doktoritöö uuringuteks välja töötatud sõnavara- ja tekstimõistmistestides ning kuidas on tekstimõistmise eri tasandid seotud sõnavaraga? (Artikkel IV)

2. Tuvastada õpilaste individuaalsed erinevused sõnavaras ja tekstimõistmises eri tasanditel.

    2.1 Millised õpilasrühmad eristuvad sõnavara ning sõnasõnalise, järeldava ja hindava tekstimõistmise alusel? (Artiklid IV, V)

3. Koostada sekkumisprogramm õpilaste sõnasõnalise, järeldava ja hindava tekstimõistmise arendamiseks ning kontrollida programmi tõhusust.

    3.1 Kuidas mõjutab tekstimõistmisstrateegiate õpetamine õpilaste sõnavara ja tekstimõistmise arengut eri tasanditel? (Artikkel V)

    3.2 Mil määral mõjutab tekstimõistmisstrateegiate õpetamine erineva võimekusega õpilaste sõnavara ja tekstimõistmise arengut? (Artikkel V)

Doktoritöös analüüsiti esmalt seda, kuidas eesti keele taseme- ja eksamitööde tekstimõistmisülesanded jaotuvad eri tasandite vahel (I, II ja III artikkel), s.t missuguseid tekstimõistmise tasandeid nende ülesannetega saab hinnata. Seejärel töötati välja mitmetasandiline tekstimõistmistest ja ühetasandiline sõnavaratest (IV artikkel).

Analüüsiks valiti 2013.–2016. aasta 3., 6. ja 9. klassi eesti keele taseme- ja eksamitöödes sisalduvad tekstimõistmisülesanded (3. klassist 78 ülesannet, 6. klassist 67 ülesannet ja 9. klassist 87 ülesannet). Need ülesanded liigitati järgmistesse kategooriatesse: sõnasõnalised, järeldavad ja hindavad ülesanded. Analüüsist selgus, et väga suur osa ülesandeid kontrollib seda, kui hästi õpilased selgesõnalisest infost aru saavad. Väga vähe oli aga niisuguseid ülesandeid, mis kontrollivad, kui hästi saavad õpilased tekstist aru kõrgeimal ehk hindaval tasandil. Lisaks oli sama klassi taseme- ja eksamitöödes sisaldunud eri tasandeid mõõtvate ülesannete jaotus aastati erinev. Samuti selgus, et üldjuhul sisaldasid 3. klassile mõeldud testid rohkem kõrgema tasandi tekstimõistmisülesandeid kui 6. ja 9. klassi tööd. Õpilaste tekstimõistmist hinnati enamasti vastusevariantidega ülesannete abil.

Doktoritöö raames koostati järgmisena teooriale tuginedes uued sõnavara- ja tekstimõistmistestid, analüüsiti õpilaste tekstimõistmist eri tasanditel ning tekstimõistmise tasandite ja sõnavara omavahelisi seoseid (IV artikkel). Uuringu valimisse kuulus 301 neljanda klassi õpilast koolidest üle Eesti. Kinnitava faktoranalüüsi abil uuriti, kuivõrd eristuvad tekstimõistmise tasandid tekstimõistmistestis ja kuivõrd homogeensed on sõnavaratestid. Analüüsi põhjal selgus, et testid võimaldasid hinnata õpilaste sõnavara ühel tasandil ja tekstimõistmist kolmel tasandil. Sõnavara ja tekstimõistmise tasandite omavaheliste seoste leidmiseks kasutati struktuurivõrranditega mudeldamist. Sellest analüüsist kerkis esile sõnavara olulisus kõikidel tekstimõistmise tasanditel – eriti tähtsaks osutus sõnavara tundmine tekstimõistmise sõnasõnalisel ja järeldaval tasandil. Sõnavara mõjutas märkimisväärselt ka tekstimõistmist hindaval tasandil.

Teiseks analüüsiti doktoritöös õpilaste individuaalseid erinevusi sõnavara ja tekstimõistmise eri tasandite tulemuste alusel. Uurimisandmed koguti väljatöötatud sõnavara- ja tekstimõistmistestidega 4. klassis (IV artikkel) ning samade testide täiendatud versiooniga 6. klassis (V artikkel). 4. klassi valimisse kuulus 301 ja 6. klassi valimisse 257 õpilast üle Eesti. Õpilaste individuaalseid erinevusi ja jaotumist profiilirühmadesse hinnati latentse profiili analüüsi (LPA) abil. Vastavalt LPA tulemustele jagati õpilased rühmadesse nende sõnavara ja tekstimõistmise tasandite z-skooridest lähtudes.

4. klassi valimis eristus kolm profiilirühma, mille puhul õpilaste tulemused sõnavaras ja eri tasandite tekstimõistmises erinesid statistiliselt kaasõpilaste omadest. Esimesse profiilirühma kuulusid õpilased, kelle tulemused sõnavaras ja eri tasandite tekstimõistmises olid keskmisel tasemel ja stabiilsed. Selliseid õpilasi oli ootuspäraselt kõige rohkem. Ülejäänud kaks profiilirühma moodustasid kehvade tulemustega õpilased, kes eristusid üksteisest mitmes aspektis. Nimelt olid teise profiilirühma õpilaste tulemused kõikide oskuste lõikes sarnasel tasemel. Nende sõnavara ja tekstimõistmine sõnasõnalisel ja hindaval tasandil olid küll

tagasihoidlikud, kuid kolmanda rühma tulemustest oluliselt paremad. Kolmanda profiilirühma õpilased paistsid silma aga selle poolest, et nende sõnavara ja tekstimõistmise tulemused olid madalal tasemel ja varieerusid eri oskustes: nende tulemused olid kehvad sõnavaras ja järeldavas tekstimõistmises, väga kehvad hindavas tekstimõistmises ja äärmiselt kehvad sõnasõnalises tekstimõistmises. Ka 6. klassis eristus kolm profiilirühma, kuid kaks neist olid keskmise tasemega ja üks kehva tasemega. Keskmiste tulemustega õpilased, keda oli vaatlusaluses valimis kõige rohkem, jagunesid sooritustaseme poolest kahte profiilirühma: mõlema rühma tulemused sõnavaras ja eri tasandite tekstimõistmises olid stabiilsed, kuid esimese rühma tulemused olid oluliselt paremad kui teisel õpilasrühmal. Kolmandasse ehk kehva tasemega rühma paigutus kümnendik 6. klassi õpilastest, kelle tulemused eri tasandite tekstimõistmises olid kasinad ja sõnavaras väga kasinad.

Kolmandaks töötati doktoritöö raames välja sekkumisprogramm, mis võimaldab õpilaste tekstimõistmist parandada. Programmi tõhusust kontrolliti 6. klassi õpilastel (V artikkel). Enne ja pärast sekkumist hinnati õpilaste sõnavara ja tekstimõistmist. Uuringusse kaasatud õpilased jaotati katse- ja kontrollrühma. Katserühma õpilastele (N = 153) õpetati kolme kuu jooksul kaks korda nädalas kirjandustundide ajal kuut tekstimõistmisstrateegiat. Kontrollrühma lapsed (N = 104) jätkasid sama aja jooksul oma tavapäraste kirjandustundidega. Õpilaste eel- ja järeltesti tulemusi võrreldi kordusmõõtmiste dispersioonanalüüsiga. Ilmnes, et katserühma õpilaste tulemused olid kolmekuulise sekkumisaja järel nii sõnavara kui ka kõikide tekstimõistmise tasandite poolest oluliselt paranenud. Seevastu kontrollrühma õpilastel paranesid vaid sõnasõnalise tekstimõistmise tulemused. Lisaks võrreldi õpilaste tulemuste muutust eri profiilirühmades. Selgus, et need õpilased, kes kuulusid nõrgemate keskmiste tulemustega katserühma, parandasid oma sooritust sõnasõnalises, järeldavas ja hindavas tekstimõistmises märkimisväärselt pärast seda, kui neile oli tekstimõistmisstrateegiaid õpetatud. Kehvade tulemustega õpilaste puhul arendas tekstimõistmisstrateegiate õpetamine tuntavalt nende sõnavara ja sõnasõnalist tekstimõistmist. Kõige vähem toetas strateegiate õpetamine tugevamate keskmiste tulemustega õpilaste arengut – oluliselt paranes vaid nende sõnasõnaline tekstimõistmine. Kontrollrühmas toimus suur muutus üksnes nõrgemate keskmiste tulemustega õpilaste rühmas: nende sõnasõnaline tekstimõistmine paranes pärast kolmekuulist tekstimõistmisstrateegiate õpetamist märgatavalt.

Doktoritöö pakub olulist uut teavet selle kohta, millised tekstimõistmisülesanded taseme- ja eksamitöödes sisalduvad ning kuidas õpilaste tekstimõistmist saab arendada. Töö tugevus seisneb põhjalikus ülevaates selle kohta, milliseid tekstimõistmise tasandeid eesti keele taseme- ja eksamitööde tekstimõistmisülesanded mõõdavad ning kuivõrd on tekstimõistmise eri tasandite jaotus ülesannetes õpilaste arenguga kooskõlas. Doktoritöö raames loodud tekstimõistmisja sõnavarateste saab aluseks võtta uute üleriigiliste testide, aga ka tekstimõistmise uurimiseks vajalike testide koostamisel. Doktoritöö raames välja töötatud testide abil saavad õpetajad oma lugemistunde kavandada. Niisamuti

võimaldab doktoritöös välja töötatud sekkumisprogramm koos selle üksik-asjaliku kirjelduse ja ülesannetega õpetajatel arendada emakeeletundides õpilaste tekstimõistmist. Seevastu teadlastele võib sekkumisprogramm pakkuda lisa-ainest, mille abil uusi programme välja töötada või olemasolevaid parendada.

Doktoritöö piiranguna saab välja tuua asjaolu, et testides sisalduvad järeldava tekstimõistmise ülesanded on suhteliselt väikese sisemise usaldusväärsusega. Piiranguks võib pidada ka sekkumisprogrammi lühikest kestust. Tekstimõistmis-strateegiate pikem õpetamise aeg oleks võimaldanud õpilastel õpitud strateegiaid paremini kinnistada ja automatiseerida.

Doktoritöö tulemused rõhutavad vajadust töötada tekstimõistmise hindamiseks välja sellised testid, millega on võimalik hinnata õpilaste tekstimõistmist eri tasanditel. Niisuguste testide abil saab täpsemalt välja selgitada, missugused tekstimõistmise komponendid ja protsessid õpilastele raskusi valmistavad. Testi-tulemuste põhjal saab kavandada õppetööd nii, et see toetaks õpilase teksti-mõistmist võimalikult hästi. Doktoritöö uurimistulemused kinnitavad, et teksti-mõistmisstrateegiate õpetamine võimaldab arendada ka kõrgema tasandi teksti-mõistmist. Kui tekstimõistmisstrateegiaid koolis ei õpetata, siis võib järeldava ja hindava tekstimõistmise arendamine jääda tagaplaanile. Tekstimõistmisstra-teegiate olulisust silmas pidades tuleks nende õpetamist riiklikus õppekavas veelgi enam rõhutada ning neid peab arvestama keeletundide planeerimisel. Tähtis on suurendada õpetajate teadlikkust, kuivõrd oluline on arendada õpilaste tekstimõistmist eri tasanditel. Õpetajakoolituses tuleb pakkuda põhjalikku üle-vaadet tekstimõistmisest ning teadmisi sellest, kuidas tekstimõistmisstrateegiaid lastele eesmärgipäraselt õpetada.

PUBLICATIONS

CURRICULUM VITAE

Name: Triinu Kärbla
Date of birth: 09.06.1979
Citizenship: Estonian
Work address: Võru Kesklinna Kool, Vabaduse 12, Võru, 65609 Võru maakond
Phone: +372 5307 8910
E-mail: triinu.karbla@ut.ee

Education:
2016–…   University of Tartu, PhD studies in Education
2009–2014   University of Tartu, Master of Arts in Education (Primary School Teacher)
1999–2003   Võru County Vocational Training Centre, Business Management (first stage of higher education)
1986–1997   Parksepa Secondary School

Professional development:
2015–….   Võru Kesklinna Kool, teacher
2019–2020   University of Tartu, Faculty of Social Sciences, Institute of Education, Junior Research Fellow in Primary Education (0.15)
2016–2018   University of Tartu, Faculty of Social Sciences, Institute of Education, Junior Research Fellow in Primary Education (0.15)
2002–2015   AS Wermo, supplier

Field of research:
Students' text comprehension, its multidimensional assessment, and promoting by teaching comprehension strategies

ELULOOKIRJELDUS

Nimi: Triinu Kärbla
Sünniaeg: 9.06.1979
Kodakondsus: Eesti
Aadress: Võru Kesklinna Kool, Vabaduse 12, Võru, 65609 Võru maakond
Telefon: +372 5307 8910
E-post: triinu.karbla@ut.ee

Haridustee:
2016–…   Tartu Ülikool Haridusteadus, doktoriõpe
2009–2014   Tartu Ülikool, sotsiaal- ja haridusteaduskond, (klassiõpetaja), magister
1999–2003   Võrumaa Kutsehariduskeskus, (ärijuhtimine), rakenduskõrgharidus
1986–1997   Parksepa Keskkool

Teenistuskäik:
2015–…   Võru Kesklinna Kool, õpetaja
2019–2020   Tartu Ülikool, Sotsiaalteaduste valdkond, haridusteaduste instituut, algõpetuse nooremteadur (0,15)
2016–2018   Tartu Ülikool, Sotsiaalteaduste valdkond, haridusteaduste instituut, alghariduse nooremteadur (0,15)
2002–2015   AS Wermo, ostujuht

Teadustegevus:
Õpilaste tekstimõistmise oskus, selle mitmetasandiline hindamine ja toetamine läbi tekstimõistmise strateegiate õpetamise

Publikatsioonid:

Kärbla, T., Uibu, K. & Männamaa, M. (2020). Teaching strategies to improve students' vocabulary and text comprehension. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-020-00489-y

Kärbla, T., Uibu, K. & Männamaa, M. (2019). National Estonian-language tests: What is measured in text comprehension tasks? *New Trends and Issues Proceedings on Humanities and Social Sciences, 6*(5), pp 8–16. Available from: www.prosoc.eu

Kärbla, T., Uibu, K., & Männamaa, M. (2018). Pikilõikeline sissevaade eesti keele taseme-ja eksamitööde tekstimõistmisülesannetesse [A longitudinal introspective into the text comprehension tasks of national tests and exams of the Estonian language]. *Eesti Haridusteaduste Ajakiri. Estonian Journal of Education*, *6*(1), 157–178. https://doi.org/10.12697/eha.2018.6.1.07

Kärbla, T., Uibu, K., & Männamaa, M. (2017). Eesti keele riiklike tasemetööde tekstimõistmisülesannete analüüs [Analysis of the comprehension tasks of national standardized tests of the Estonian language]. *Eesti Rakenduslingvistika Ühingu aastaraamat, 13*, 73–87. https://doi.org/10.5128/ERYa13.05

Kärbla, T., & Uibu, K. (2016). Assessment of Students' Higher-level Text Comprehension Skills in Basic School. New Trends and Issues Proceedings on Humanities and Social Sciences, 2(5), 31–37.

DISSERTATIONES PEDAGOGICAE
UNIVERSITATIS TARTUENSIS

1. **Карлеп, Карл**. Обоснование содержания и методики обучения родному языку во вспомогательной школе. Tartu, 1993.
2. **Ots, Loone**. Mitmekultuurilise hariduse õppekomplekt eesti kirjanduse näitel. Tartu, 1999.
3. **Hiie Asser**. Varajane osaline ja täielik keeleimmersioon Eesti muukeelse hariduse mudelitena. Tartu, 2003.
4. **Piret Luik**. Õpitarkvara efektiivsed karakteristikud elektrooniliste õpikute ja drillprogrammide korral. Tartu, 2004.
5. **Merike Kull**. Perceived general and mental health, their socio-economic correlates and relationships with physical activity in fertility-aged women in Estonia. Tartu, 2006.
6. **Merle Taimalu**. Children's fears and coping strategies: a comparative perspective. Tartu, 2007.
7. **Anita Kärner**. Supervision and research training within the professional research community: Seeking new challenges of doctoral education in Estonia. Tartu, 2009.
8. **Marika Padrik**. Word-formation skill in Estonian children with specific language impairment. Tartu, 2010.
9. **Krista Uibu**. Teachers' roles, instructional approaches and teaching practices in the social-cultural context. Tartu, 2010.
10. **Anu Palu**. Algklassiõpilaste matemaatikaalased teadmised, nende areng ja sellega seonduvad tegurid. Tartu, 2010.
11. **Mairi Männamaa**. Word guessing test as a measure of verbal ability. Use of the test in different contexts and groups. Tartu, 2010.
12. **Piret Soodla**. Picture-Elicited Narratives of Estonian Children at the Kindergarten-School Transition as a Measure of Language Competence. Tartu, 2011.
13. **Heiki Krips**. Õpetajate suhtlemiskompetentsus ja suhtlemisoskused. Tartu, 2011.
14. **Pille Häidkind**. Tests for assessing the child's school readiness and general development. Trial of the tests on the samples of pre-school children and first-grade students in Estonia. Tartu, 2011.
15. **Karmen Trasberg**. Keskkooli- ja gümnaasiumiõpetajate ettevalmistus Eesti Vabariigis (1918–1940) õpetajakoolituse ajaloolise kujunemise kontekstis. Tartu, 2011, 207 lk.
16. **Marvi Remmik**. Novice University Teachers' professional development and learning as a teacher: Opportunities and Conditions at Estonian Higher Education Institutions. Tartu, 2013, 129 p.
17. **Pilve Kängsepp**. Küsimuste kasutamine kui võimalus toetada õpilaste arusaamist loetust. Tartu, 2014, 125 p.

18. **Marge Täks**. Engineering students' experiences of entrepreneurship education. A qualitative approach. Tartu, 2015, 150 p.
19. **Reelika Suviste**. Students' mathematics knowledge and skills, and its relations with teachers' teaching and classroom management practices: Comparison between Estonian- and Russian-language schools. Tartu, 2015, 147 p.
20. **Liina Lepp**. The objectives of doctoral studies and factors influencing doctoral study process from the perspectives of different parties. Tartu, 2015, 271 p.
21. **Ülle Säälik**. Reading literacy performance: Metacognitive learning strategies matter, schools have effect on student outcomes. Tartu, 2016, 119 p.
22. **Katrin Saks**. Supporting Students' Self-Regulation and Language Learning Strategies in the Blended Course of Professional English. Tartu, 2016, 216 p.
23. **Anne Okas**. Novice and experienced teachers' practical knowledge in planning, delivery and reflection phases of teaching. Tartu, 2016, 172 p.
24. **Külli Kori**. The Role of Academic, Social and Professional Integration in Predicting Student Retention in Higher Education Information Technology Studies. Tartu, 2017, 168 p.
25. **Ingrid Koni**. The perception of issues related to instructional planning among novice and experienced teachers. Tartu, 2017, 142 p.
26. **Ivar Männamaa**. Development of an educational simulation game and evaluation of its impact on acculturation attitudes. Tartu, 2017, 154 p.
27. **Egle Säre**. Developing the reasoning skills of pre-schoolers through Philosophy for Children. Tartu, 2018, 131 p.
28. **Anu Sööt**. The procedure of guided core reflection for supporting the professional development of novice dance teachers. Tartu, 2018, 135 p.
29. **Tiina Anspal**. The development of teacher identity through role and self-conception in pre-service teacher education. Tartu, 2018, 157 p.
30. **Age Salo**. The dual role of teachers: school-based teacher educators' beliefs about teaching and understandings of supervising. Tartu, 2019, 156 p.
31. **Mirjam Burget**. Making sense of responsible research and innovation in science education. Tartu, 2019, 175p.
32. **Kaire Uiboleht**. The relationship between teaching-learning environments and undergraduate students' learning in higher education: A qualitative multi-case study. Tartu, 2019, 169 p.
33. **Karin Naruskov**. The Perception of Cyberbullying among Estonian Students According to Cyberbullying Types and Criteria. Tartu, 2020, 176 p.
34. **Raili Allas**. Supporting teachers' professional development through reflection procedure. Tartu, 2020, 182 p.